ECONOMIC RESEARCH REPORTS

WHAT MOTIVES SHOULD GUIDE REFEREES? ON THE DESIGN OF MECHANISMS TO ELICIT OPINIONS

by **Jacob Glazer**
and
Ariel Rubinstein

RR # 96-01 **January 1996**

C.V. STARR CENTER
FOR APPLIED ECONOMICS

NEW YORK UNIVERSITY
FACULTY OF ARTS AND SCIENCE
DEPARTMENT OF ECONOMICS
WASHINGTON SQUARE
NEW YORK, NY 10003-6687

What Motives Should Guide Referees?
On the Design of Mechanisms to Elicit Opinions

Jacob Glazer
The Faculty of Management, Tel Aviv University, Tel Aviv, Israel

and

Ariel Rubinstein
School of Economics, Tel Aviv University, Tel Aviv, Israel

January 1996

We wish to thank Dilip Abreu, Kyle Bagwell, Matt Jackson, Albert Ma, Tom Palfrey and Mike Riordan for comments on an earlier version of the paper. Financial support from of the C.V. Starr Center for Applied Economics is gratefully acknowledged.

Abstract

The refereeing process can be conceived of as a mechanism for deciding whether or not to accept a paper based on information gathered from a number of referees, each of whom receives a noisy signal regarding the appropriateness of publication. The public target is to make the best possible decision on the basis of all the information held by the referees.

We compare two worlds. In one, all referees are driven only by the public motive--to accept appropriate papers. In the second, each referee is also driven by a private motive--to have his recommendation accepted. It is shown that in the first world, every mechanism will have a "bad" equilibrium, that is, one which does not achieve the public target. For the second world, we construct a mechanism whose unique equilibrium outcome does achieve the public target.

According to the mechanism we construct, one of the referees is assigned a special role, that of determining the subgroup whose members' votes, when combined with his own vote, will determine the decision by employing the majority rule. He makes his selection of the subgroup simultaneously with all other referees casting their votes whereas his own vote is cast only after hearing the outcome of the subgroup voting.

J.E.L. Classification Numbers: C72
Key Words: public motive, personal motive, implementation, eliciting opinions.

1. Introduction

The fate of a paper has to be determined. It can be either accepted or rejected. Each member from among a group of referees has an opinion regarding the acceptability of the paper. The desirable outcome, is the view held by the majority of referees. However, any referee, if asked to make a recommendation, may not provide a sincere one. This paper is concerned with the design of mechanisms for decision making based on elicited opinions.

The refereeing example is our leading metaphor. However, the basic ingredients of our analysis exist in most situations in which a decision has to be made on the basis of information elicited from a set of experts. Other examples include:

(i) A decision is to be made whether or not to operate on a patient on the basis of consultations with several physicians.

(ii) An investigator must determine whether or not a certain event has occurred, based on the evidence provided by a group of witnesses.

(iii) A military commander has to decide whether to attack at night or at dawn, on the basis of recommendations made by his aides.

In such scenarios, the agents may have different opinions due to the random elements which affect their judgments. The existence of such randomness is precisely the principle rationale for why such decisions are often made on the basis of more than a single opinion.

In such situations, each expert, is torn between two types of motives: first, the public motive--the desire that the decision taken is the "right" one, second, the private motive--the wish that his recommendation be accepted.

Societies differ in their attitudes toward the private motive. In some societies, private motives are considered as more legitimate than in others. The desirability of the private motives can be questioned on different grounds, some of them ethical. In this paper, we wish to address the issue from the point of view of the functionality of the private motive as a means to achieve the target of making the best decision on the basis of all the opinions offered. We refer to this goal as the public target (PT).

To gain some intuition about the difficulties in implementing the PT, consider the mechanism where three referees are asked to make simultaneous recommendations and the alternative which gets most votes is executed. If all referees care only about the public motive, the above mechanism has the desired equilibrium in which all referees make sincere recommendations. However, other equilibria also exist, for example, the one where all referees recommend acceptance regardless of their sincere opinions. If each referee is also driven by the private motive, the desire that his recommendation be accepted, this "bad" equilibrium becomes even more stable, since a deviation increases the chance that he will provide the minority opinion. Note also, that this "bad" equilibrium will be strengthened even further if the strategy to always recommend acceptance is less costly to the referee than the sincere recommendation strategy, which requires the referee to actually read the paper.

We find that if all referees are driven only by the public motive, every mechanism, will have a "bad" equilibrium. In this equilibrium, the probability that the right decision is made is not higher than the probability obtained if only one referee is asked for his opinion. On the other hand, for the world in which both motives are active in all referees' minds, a mechanism exists such that regardless of the referees' tradeoff between the two motives, the unique equilibrium outcome elicits sincere opinions from all referees. Thus, the message of this paper is that the two motives are requisite to achieve the public target.

Our analysis ignores the existence of other motives whose emergence depends on additional information being discovered at the end of the process. For example, it might be the case that after the decision regarding an emergency operation on the patient is made (i.e., performed or not), some additional information is discovered which clearly identifies the right course. In such a case, an additional motive may emerge, such as the desire of each physician to be proven objectively right. Our analysis is restricted to the case in which the "truth" never becomes apparent: the acceptability of a paper remains unresolved, the true condition of the patient is not discovered, the fact whether the event occurred or not remains concealed, and the correct hour to attack is never clarified.

2. The Model

An action 0 or 1 has to be taken. The desirable action depends on a random variable ω, the state, which receives a value of 0 or 1 with equal probability. The desired action in state ω is ω. There is a set of agents, $N=\{1,..,n\}$ (n is odd and $n > 2$). Agent i receives a signal x_i, which at the state ω gets the value ω with probability $1 > p > 0$ and the value $-\omega$ with probability 1-p (we use the convention that $-1=0$ and $-0=1$). The signals are conditionally independent. Note that we have assumed symmetry in the sense that all agents are identical, both states are equally likely and the probability of an agent getting the "right" signal ($x_i=\omega$) is independent of the state.

The number of 0s and 1s observed by the agents is the best information that can be collected in this situation. Note that in this model, no useful information is obtained if, for example, 10 signals are observed, 5 of which are 0s and 5 of which are 1s. In this case, the ex-post beliefs about the state remain identical to the ex-ante beliefs. This will not be the case under some other informational structures, where such an outcome may signal the diminished importance of the decision.

Denote by V(K) the highest probability that the desirable action will be taken if a decision is made on the basis of the realization of K signals only. That is, for any given K agents,
V(K)=prob{strict majority of the agents get the right signal}+
1/2prob{exactly one-half of the agents get the right signal}.
As is often the case with value of information functions (see Radner and Stiglitz (1984)), the function V is not concave. Actually, in our model, the function V is only weakly increasing: $V(2\ell)=V(2\ell-1)$ and $V(2\ell+1) > V(2\ell)$.

We refer to the operation of collecting information from the agents, calculating the consequence and executing it as a mechanism. We model a mechanism as a finite extensive game with imperfect information (but no imperfect recall), with the n agents being the players, no chance players and with consequences which are identified as either 0 or 1. Note that this definition excludes mechanisms with chance moves.

Three examples of mechanisms are:
The direct simultaneous mechanism: All agents simultaneously make a recommendation, 0 or 1, and the majority determines the consequence.

The direct sequential mechanism: The agents move in a predetermined order. Each agent moves only once and makes his recommendation public; the majority determines the consequence.

The leader mechanism: In the first stage, agents 1,2,...,n-1 (the lower-level agents) simultaneously make a recommendation transferred to agent n (the "leader"), who makes the final recommendation which determines the consequence.

A mechanism together with the random elements define a Bayesian game form. Executing an n-tuple of strategies in a mechanism yields a lottery with the consequences 0 or 1.

The public target (PT) is to construct a mechanism which maximizes π_1, the probability that the desirable action is taken (the consequence ω at state ω). This definition assumes that there is no greater loss in taking one of the two possible mistakes (taking the action $—\omega$ at state ω).

The agents in our model can be driven by two motives. The public motive is to maximize π_1. The private motive is to have his recommendation accepted. In order to precisely define the private motive we will have to add to the description of a mechanism a profile of sets of histories $(R_i)_{i\in N}$ so that R_i is interpreted as the set of histories in which agent i makes a recommendation. We require that for every $h\in R_i$, the set P(h) contains i, that $A_i(h)=\{0,1\}$, and that there is no terminal history h which has two subhistories in R_i. The action of agent i at history $h\in R_i$ is interpreted as a recommendation. Whenever we discuss the private motive, we will refer to the comparison between the agent's action at the history h and the resulting consequence. The probability that agent i's recommendation will coincide with the consequence of the mechanism is denoted by $\pi_{2,i}$. Agent i's private motive is to maximize $\pi_{2,i}$.

The preferences of each agent i are assumed to depend only on π_1 and $\pi_{2,i}$. When we say that all agents are driven only by the public motive, we mean that they wish only to increase π_1. When we say that they are driven by both the private and the public motives we mean that each agent i has some preferences according to which he wishes to increase both π_1 and $\pi_{2,i}$.

The concept of equilibrium we adopt is that of sequential equilibrium in pure strategies (for simultaneous mechanisms this coincides with Bayesian-Nash equilibrium). We say that a mechanism implements the PT if in every sequential equilibrium of the game, $\pi_1 = V(n)$. That is, the consequence of any sequential equilibrium, for every profile of signals, is identical with the signal observed by the majority of agents.

3. The Impossibility of Implementation When All Agents are Driven by the Public Motive Only

In this section, we will show that if all agents are driven by the public motive only, no mechanism implements the PT. That is, the game obtained by any mechanism coupled with the agents' objective of increasing π_1 only, has a sequential equilibrium with π_1 strictly less than V(n).

In order to achieve a better understanding of the designer's difficulties in implementing the PT, we will now consider the three mechanisms described in the previous section and see what prevents "truth-telling" from being the only equilibrium. We say that an agent uses the "T" strategy if, whenever he makes a recommendation, it is identical to the signal he received. "NT" is the strategy in which an agent who received the signal x recommends -x and "c" (c=0,1) is the strategy in which an agent announces c independently of the signal he received.

The direct simultaneous mechanism: In this mechanism all agents playing "T" is an equilibrium. However, the two equilibria offered below do not yield the PT:

(1) All agents play "c" (a single deviation of agent i will not change π_1).

(2) Agents 1 and 2 play "0" and "1", respectively, while all other agents play "T".

One may argue that the equilibrium in which all agents play "T" is the most reasonable one since telling the truth appears to be a natural focal mode of behavior. However, the notion of implementation which we use does not relate any focality to truth-telling. Note that although we do not include the cost of implementing a strategy in the model, we can conceive of some costs, associated with the strategies "T" or "NT", which can be avoided by executing "0" or "1". These costs make the equilibrium in which all agents choose "c" quite stable: Executing the strategy "T" will not alter the mechanism's outcome (will not increase π_1) but will strictly increase the agent's costs.

Note also that in this game, the strategy "T" is not a dominant strategy (not even weakly) when n > 3. To comprehend this note that if agents 1 and 2 play "0", and agents 3 and 4 play "T", then "1" is a better strategy for agent 5 than "T". The events in which the strategy "1" and "T" lead to different outcomes occur when agents 3 and 4 get the signal 1 and agent 5 gets the signal 0. The strategy "1" is better than "T" in the event $\{\omega=1$ and $(x_3,x_4,x_5)=(1,1,0)\}$ and is worse in the less likely event $\{\omega=0$ and $(x_3,x_4,x_5)=(1,1,0)\}$.

The direct sequential mechanism: This mechanism also does not implement the PT. All agents playing "T" is an equilibrium, however the following are two other equilibria:

(1) Agent 1 plays "T" and all other agents repeat his recommendation with beliefs that assign no significance to any out-of-equilibrium moves. This is a sequential equilibrium with $\pi_1=V(1)$.

(2) Agent 1 plays "NT", agents 2,...,n-1 play "T", and agent n announces the opposite of what agent 1 has announced. This is a sequential equilibrium strategy profile with $\pi_1=V(n-2)$. Agent 1 cannot profitably deviate (as agent n neutralizes his vote in any case). Agent n cannot profitably deviate, since if he conforms to the equilibrium, $\pi_1=V(n-2)$, and if, instead, he plays "T" then π_1 will be even smaller. Note that this equilibrium does not have any out-of-equilibrium histories and thus cannot be excluded by any of the standard sequential equilibrium refinements.

The leader mechanism: Once again there is an equilibrium with $\pi_1=V(n)$, however the following is a sequential equilibrium with $\pi_1=V(1)$: agents 1,2,,..,n-1 play "0"; agent n, the leader, always announces his signal independently of the recommendations he receives from the lower-level agents and assigns no significance to deviations.

The following proposition not only shows that there is no mechanism which implements the PT but, also, that every mechanism has a "bad" equilibrium with π_1 no larger than what would be obtained if a single agent was nominated to make a decision based only on his signal.

Proposition 1: If all agents are only interested in increasing π_1, then every mechanism will have a sequential equilibrium with $\pi_1 \leq V(1)$. Thus, there is no mechanism which implements the PT when all agents are driven only by the public motive.

We provide a proof of this proposition in Appendix 1. Here, we wish to provide an intuition for the main idea of the proof. Consider first a one-stage, simultaneous-move mechanism. For such a mechanism, we will construct a sequential equilibrium with $\pi_1 \leq V(1)$. If the outcome of the mechanism is constant, then the behavior of the agents is immaterial and $\pi_1 = V(0)$. Otherwise, there is an agent i and a profile of actions for the other agents $(a_j)_{j \neq i}$ so that the consequence of the mechanism is sensitive to agent i's action: that is, there are two actions, b_0 and b_1, for agent i which yield the consequences 0 and 1, respectively. Assign to any agent j $\neq$ i to play the action a_j independently of the signal he received. Assign to player i to play the action b_x if he received the signal x. Following this profile of strategies yields $\pi_1 = V(1)$ and any deviation is unprofitable since it makes the outcome of the mechanism dependent on at most two signals, and $V(2) = V(1)$.

Now consider a two-stage mechanism where, at each stage, each agent makes a move. We first construct the strategies for the second stage. For every profile of actions taken at the first stage, for which the consequence is not already determined, assign strategies in a manner similar to the one we did for the one-stage mechanism. We proceed by constructing the strategies for the first stage. If the outcome of the mechanism is always determined in the first stage, then the two-stage mechanism is essentially one-stage, and we can adapt our construction of the sequential equilibrium for the one-stage mechanism. Otherwise, assign to each agent i at the first stage to play an action a_i^* independent of his signal, where (a_i^*) is a profile of actions which does not determine the consequence of the mechanism. Coupled with beliefs that do not assign any significance to deviations in the first stage, we obtain a sequential equilibrium with $\pi_1 = V(1)$.

4. The Main Proposition: Implementation is Possible When All Agents Have Both Motives

We now move from the world in which all agents are driven only by the public motive to the world in which they are driven by both the public and the private motives. We show that, unlike in the previous case, implementation of the PT is possible here. Our main result is the construction of a mechanism which implements the PT for any profile of preferences as long as all agents are driven by both the public and private motives.

The mechanism we offer is as follows: One of the agents, say agent 1, is assigned the special status of "controller". In the first stage, each agent, excluding the controller, secretly makes a recommendation while the controller simultaneously determines a set of agents S whose votes will be counted. The set S must be of even size and should not include the controller. In the second stage, the controller learns the results of the votes casts by the members of S and only then adds his vote. The majority of votes in $S \cup \{1\}$ determines the outcome.

The controller in this mechanism has a double role. First, he has the power to exclude from the voting group those agents whose votes has a negative effect on π_1. This decision of the controller expresses his assessment of the "quality" of each agent's recommendation. Second, the controller contributes his own view whenever his vote is pivotal. Note, that each agent (but not the controller) makes a recommendation in the first stage even if his vote is not counted. All agents, including those whose votes are not counted, are driven by the private motive and aim to make their recommendations equal to the consequence chosen by the mechanism. According to the mechanism's design, if the controller's vote is pivotal, his recommendation will determine the outcome; if it is not, he can join the majority without affecting the outcome. Thus, the mechanism is designed so that the private motive of the controller does not prevent him from optimizing the PT.

We will prove that the only equilibrium possible is the one where, in the first stage, all agents other than the controller play "T" and they are all included in S while in the second stage, the controller joins the majority in S unless he is pivotal, in which case he plays "T". Following are the main arguments of the proof showing that no other equilibria are possible:

(1) The controller's decision regarding the inclusion of an agent in S is the result of two considerations: first, he may also obtain information from agents who play "NT" but, second, an NT-agent's vote can negatively affect the outcome. We will show that the latter is a stronger consideration. Therefore, agents who play "NT" will be excluded from S.

(2) Since the mechanism enables the controller to maximize the public motive without worrying about the private motive, he selects the set S so as to be the "most informative". Thus, the set S consists of all agents who play "T" and possibly some agents who play "0" or "1" (as long as the difference between the numbers of "0"s and "1"s is not greater than 1).

(3) There is no equilibrium in which some of the agents in the set S choose a pooling strategy ("c") as there will always be an agent who plays a pooling strategy that will increase $\pi_{2,i}$ without decreasing π_1 by switching to "T".

(4) There is no equilibrium with some agents excluded from the set S. If agent i is excluded from S, then by (2) he does not play "T", but since he does not affect the consequence he will deviate profitably to playing "T" so as to maximize $\pi_{2,i}$.

Proposition 2: The following mechanism implements the PT for any profile of preferences that satisfies the condition that each agent i's preferences increase in both π_1 and $\pi_{2,i}$:

Stage 1: Simultaneously, each agent, except agent 1, makes a recommendation, 0 or 1, while agent 1 announces an even-numbered set of agents, S, which does not include himself.

Stage 2: Agent 1 is informed about the total number of members of S who voted 1 and makes his own recommendation, 0 or 1.

The majority of votes among $S \cup \{1\}$ determines the consequence.

Proof: See Appendix 2.

For the mechanism to work, it is important that the controller only learns the result of the votes in S, not who voted what. In order to see why, assume that there are three agents who participate in our mechanism with the modification that agent 1 receives the additional information of who voted what. The following is a sequential equilibrium: agent 2 plays "0", agent 3 plays "T" and agent 1 chooses $S=\{2,3\}$ in the first stage; in the second, agent 1 plays "T" in case agents 2 and 3 voted 0 and 1, respectively, and he plays "0" in case agents 2 and 3 voted 1 and 0, respectively. This strategy profile is supported by out of equilibrium beliefs that a vote 1 by agent 2 means that he received the signal 0. This is not an equilibrium in our proposed mechanism since in the second stage agent 1 cannot distinguish between the two profiles of votes (1,0) and (0,1).

5. Comments

(1) The Case Where One Agent is Driven Only by the Public Motive

If agent 1, say, is a "special" agent in the sense that he is driven by the public motive only, whereas all other agents are driven by both the public and the private motives, then the following simultaneous mechanism implements the PT:

Agent 1: announces a pair, an even numbered set of agents not including himself, S, and a recommendation, 0 or 1.

Agent i $\neq$ 1: announces a recommendation, 0 or 1.

The consequence of the mechanism is determined by the majority of votes in $S \cup \{1\}$.

The mechanism above will not necessarily implement the PT if agent 1 is also driven by the private motive. In such a case, he may be interested in decreasing the size of the set S in order to increase the probability that his recommendation is accepted even if such a move decreases π_1.

(2) The Case Where all Agents are Driven Only by the Private Motive:

Of course, implementation of the PT is impossible in a world where all agents are driven only by the private motive, that is, when each agent i is interested only in increasing $\pi_{2,i}$. In fact, implementation is impossible whenever all agents are motivated by any sort of considerations which are independent of the state of nature. The reason is that in such a world, whatever is the mechanism, if $\sigma=(\sigma_{i,x})$ is a sequential equilibrium strategy profile ($\sigma_{i,x}$ is i's strategy given i's receipt signal x), then the strategy profile σ' where $\sigma'_{i,x}=\sigma_{i,-x}$ (each agent who receives signal x plays as if he had received the signal -x) is also sequential equilibrium strategy profile. However, if all signals are 1, the consequence under σ is 1 whereas it is 0 under σ'.

(3) Other Motives: Aiming to be Right

Implicit in the construction of our model is the assumption that the true state of the world never becomes known. In some scenarios, ω eventually gets revealed. In such a case, agents may have another objective--to make a recommendation that will turn out to be the "right" one. Denote by $\pi_{3,i}$ the probability that agent i's recommendation is ω.

If maximizing $\pi_{3,i}$ is a major objective for the agents, new strategic phenomena may arise. For example, in the direct sequential mechanism, an agent may make inferences about ω not only on the basis of his own signal but also on the basis of recommendations made by agents who moved before him. If each agent i cares only about increasing $\pi_{3,i}$, then there is no equilibrium in which all agents play "T". The sequential mechanism with such preferences is equivalent to a game studied by Bikhchandani, Hirshleifer and Welch (1992), (see also Benarjee (1992) and Chamley and Gale (1994)) in which the equilibria exhibited a "herd" type of behavior, that is, agents preferred to rely on information revealed by the recommendations of other agents rather than expressing views based on their own signals.

(4) The Symmetry Assumption

Symmetry plays an important role in our analysis: the two states are equally likely, the loss of taking the action 1 when the state is 0 is equal to the loss of taking the action 0 when the state is 1, the signal random variable is identical to all agents and the probability that the signal is ω given the state ω is p, independent of ω.

Further research is needed to examine how the departure from the symmetry assumption changes the results. At this moment, we would like to make do with commenting on the case where one assymetry element is introduced: the probability of the state 0, p_0, is strictly larger than the probability of the state 1, p_1. The following mechanism implements the PT for the case of 2 agents if both agents are driven by the public motive only.

	0	1	v
0	0	0	1
1	0	1	0
v	1	0	0

Note that in this case, the function V is strictly increasing but is also non-concave. For example, V(2) > V(1) (in case of conflicting signals, it is strictly better to choose 0), and V(3) > V(2). However, note that if the difference p_0-p_1 is small then V(2)-V(1) is small. As shown in the above example, when V is monotonic, implementation may be possible.

However, if the agents also take into account the extra cost of playing "T" (relative to "c"), then there is a range of p_0 and p_1 for which implementation of the PT will again be impossible.

6. The Relationship to the Implementation Theory

We believe that our paper also contributes to the implementation literature. The construction of mechanisms which overcome the problem of "bad" equilibria is one of the main issues discussed in implementation theory (for an up-to-date survey on implementation with perfect information see Moore (1992) and on implementation with imperfect information see Palfrey (1992)). Our paper investigates sequential equilibrium implementation in a certain environment with imperfect information. Here are several comments on the relationship between our paper and the implementation literature.

(a) A general point of criticism

The structure of the standard implementation problem is one where the designer is given a set of consequences which he can use in the construction of the mechanism. The agents' preferences relate only to the given set of consequences but are neutral to events occurring during the play of the mechanism.

For example, consider the case where there is a seller and a buyer who evaluate an item with reservation values s and b, respectively. Assume that the designer wishes to maximize the seller's profits, that is, he wants to implement the transfer of the good from the seller to the buyer for the price b as long as $b > s$. The standard implementation literature views the procedure of the seller making a "take it or leave it" offer as a solution to the designer's problem. However, this view ignores the emotions aroused when playing this kind of mechanism. The "take it or leave it" mechanism may trigger a new motive that may affect the buyer's behavior--he may consider the offer of a price which leaves him with less than, say, 10% of the surplus, "insulting". Thus, although he may still prefer to get 10% of the surplus to 0%, he will nevertheless refuse such an offer.

One can view the moves in a mechanism as abstract messages. However, the attractiveness of a mechanism is judged by its verbal interpretation. Such interpretations are associated in real life situations with additional motives which should not be ignored.

In the context of our paper, even if referees are originally concerned only about the public target, once asked to make a move, interpreted as a recommendation, they may also be driven by the private motive. Interestingly, in our problem, the introduction of the private motive does not disturb but rather help to design a mechanism that implements the PT.

(b) Virtual implementation

Virtual Bayesian Nash implementation of the PT may be possible even when all agents are driven by the public motive only. Abreu and Matsushima (1992) suggest a direct simultaneous mechanism in which the outcome is determined with probability 1-ε by the majority of announcements and with probability ε/n by agent i's recommendation (i=1,...,n). The proof that this mechanism virtually implements the PT follows from the claims made in Appendix 3. (These claims can be easily applied here to show that "NT" is dominated by "T" and, if more agents choose "c" than "-c", then one of those who chooses "c" can profitably deviate to "T"). Note that this mechanism allows the possibility that while n-1 agents observe and report the signal 0, the outcome is 1. Note also that the above mechanism uses random devices to determine the consequence. We rule out the use of such randomizations, thus, our analysis is particularly appropriate to those common situations in which the use of randomizations is unsound.

(c) Using an informer

In the standard implementation literature, the tools given to the designer consist of a set of procedures that he can impose upon the agents as well as a set of consequences that he can enforce; most important, however, the set of agents is also provided (for exceptions see Palfrey (1992) and Baliga (1994)). Notwithstanding, there are contexts in which it seems proper to consider a change in the set of agents by the addition of an extra agent, which we call a designer's agent. As described by Palfrey (1992), the role of this agent is "...to eliminate unwanted equilibria because, while he does not know the types of his opponents, he can perfectly predict their strategies, as always assumed in equilibrium analysis." In an earlier version of this paper, we analyzed in detail the use of a designer's agent in the context of our model. In particular, we proved that Proposition 1 is still valid when the use of such an agent is allowed. As for Proposition 2, the role of the controller in the first stage of the mechanism is similar to the role of a designer's agent.

(d) Bibliographic comments

A model related to ours is discussed by Palfrey and Srivastava (1989). In their Example 2, the uncertain element is a vector $(s_1,..,s_n)$, where s_i is i's signal (0 or 1). The probability that agent i's signal is 0 is q. The planner wishes to implement the rule that the action to be taken is the one that receives the majority of signals. Each agent is interested that that action be identical to the signal he receives. Thus, in their example an agent cannot learn about other agents' signals from his own signal nor does he change his mind about the preferred action given the information he receives from other agents. This model is an example of the limits of Bayesian implementation: When q is very high, it is impossible to implement the majority rule with simultaneous games. The sequential mechanism does work here, although it does not implement the PT in our model. The difference is that in Palfrey and Srivastava's example, the implementation difficulty stems from the risk of being stuck in a "pooling equilibrium." Inherent in our model is the additional potential difficulty of agents learning about the state of nature from other agents' actions. (Another related example is Example 3 in Palfrey and Srivatava (1989), in which signals are uncorrelated and all agents prefer that the action taken will be the one that received the majority of signals.)

Proposition 1 is related to results presented in Jackson (1991), which provided both a necessary condition and a sufficient condition for Bayesian implementation using simultaneous mechanisms. The PT in our model does not satisfy Bayesian Monotonicity, which is a necessary condition for such implementation. Our results do not follow from his since we refer to all mechanisms, and not solely to simultaneous mechanisms.

7. References

Abreu, D. and H.Matsushima (1992), "Virtual Implementation in Iteratively Undominated Strategies: Complete Information", Econometrica, 60, 993-1008.

Baliga, S. (1994) "The Not-So-Secret-Agent": Professional Monitors, Hierarchies and Implementation", mimeo.

Banerjee, A. (1992), "A Simple Model of Herd Behavior", The Quarterly Journal of Economics, 107, 797-817.

Bikhchandani, S., D. Hirshleifer and I.Welch (1992), "A Theory of Fads, Fashion, Custom and Cultural Change as Informational Cascades", Journal of Political Economy, 100, 992-1026.

Chamley, C. and D.Gale (1994), "Information Revelation and Strategic Delay in a Model of Investment", Econometrica, 62, 1065-1085.

Jackson, M. (1991), "Bayesian Implementation", Econometrica, 59, 461-478.

Moore,J. (1992), "Implementation, Contracts, and Renegotiation in Environments with Complete Information", in Advances in Economic Theory Sixth World Congress Volume I, J.J.Laffont (editor), Cambridge University Press, 182-282.

Osborne, M and A. Rubinstein, (1995), A Course in Game Theory, MIT Press.

Palfrey, T. (1992), "Implementation in Bayesian Equilibrium: the Multiple Equilibrium Problem in Mechanism Design", in Advances in Economic Theory Sixth World Congress Volume I, J.J.Laffont (editor), Cambridge University Press, 283-323.

Palfrey ,T. and S.Srivastava, (1989), "Mechanism Design with Incomplete Information: A Solution to the Implementation Problem", Journal of Political Economy, 97, 668-691.

Radner, R. and J.Stiglitz (1984), "A Nonconcavity in the Value of Information", in

Bayesian Models in Economic Theory, edited by M.Boyer and R.E.Kihlstrom, Elsevier Science Publishers, Chapter 3.

Appendix 1

Proof of Proposition 1: We here provide a proof for the case where the mechanism is one with perfect information and possibly simultaneous moves (see Osborne and Rubinstein (1994), page 102, for a definition). The proof here does not cover the possibility of imperfect information, however, our definition of a game form with perfect information allows for several agents to move simultaneously. A history in such a game is an element of the type $(a^1,..,a^K)$ where a^k is a profile of actions taken simultaneously by the agents in a set of agents denoted by $P(a^1,..,a^{k-1})$.

For any given mechanism, we construct a sequential equilibrium with $\pi_1 \leq V(1)$. For any non-terminal history h, denote by d(h), the maximal L, so that, $(h,a^1,..,a^L)$ is also a history. Let $(h^t)_{t=1,...,T}$ be an ordering of the histories in the mechanism so that $d(h^t) \leq d(h^{t+1})$ for all t.

The equilibrium strategies are constructed inductively. At the t'th stage of the construction, we deal with the history h^t (and some of its subhistories). There are two possibilities: If the strategies at history h^t have been determined in earlier stages, move to the next stage; if not, two possible cases arise:

Case 1: There are two action profiles, a and b, in A(h) and an agent $i^* \in P(h)$ such that:

(i) $a_i = b_i$ for all $i \neq i^*$ and

(ii) If the agents follow the strategies as previously defined, the outcomes which follow histories (h,a) and (h,b) are 0 and 1, respectively.

In such a case, do the following two things:

(I) For every $i \in P(h)-\{i^*\}$, assign the action a_i to history h, independently of the signal i observes; for agent i^*, assign the action a_{i^*} if his signal is 0 and the action b_{i^*} if his signal is 1.

(II) If h' is a proper subhistory of h and the strategy profile for h' was not defined earlier assign to any $i \in P(h')$ the action a_i where (h',a) is a subhistory of h as well (that is, the agents in P(h') move towards h).

Case 2: If for every a and b in A(h) the outcome of the game is the same if the agents follow the strategies after (h,a) and (h,b), pick an arbitrary $a \in A(h)$ and assign the action a_i to each $i \in P(h)$ independently of his signal.

To demonstrate the construction of the strategies, consider the direct sequential mechanism with the agents 1, 2 and 3 moving in that order. Consider the ordering of the histories (1,1), (1,0), (0,1), (0,0), (1), (0) and ϕ. After history (1,1) (that is, after agents 1 and 2, have both announced 1), agent 3 cannot affect the consequence. After the history (1,0), assign the strategy "T" to agent 3, and move backwards to histories (1) and ϕ (the initial history) and assign the strategies "0" and "1" to agents 2 and 1, respectively. After history (0,1) (which is by now an out-of-equilibrium history), assign the strategy "T" to agent 3, and assign the strategy "1" to agent 2 after the history (0). To summarize, the strategies which we have constructed are such that agent 3 plays "T" at the histories in which he is pivotal, agent 2 announces the opposite of agent 1's announcement, and agent 1 plays the strategy "1".

Returning to the construction of the sequential equilibrium for the general case, we still have to specify beliefs. The construction takes place by updating the beliefs according to the strategies. Moreover, whenever an out-of-equilibrium event occurs, the agents continue to hold their initial beliefs.

We now show that we have indeed constructed a sequential equilibrium. Note, first, that for every history h, there is at most one agent whose equilibrium behavior in the game following h depends on his own signal. Furthermore, if the consequence of the mechanism is not already determined at history h, there is one agent who determines the outcome according to his signal. If an agent can at all affect the consequence by a move at history h, it must be that the beliefs of the agents at this history continue to be the initial beliefs. Therefore, the agent, by a unilateral deviation, cannot increase π_1 beyond V(2)=V(1).

The extension of the proof for the case of imperfect information requires a somewhat more delicate construction to respond to the requirement that the same action is assigned to all histories in the same information set. □

Appendix 2

Proof of Proposition 2: The following is an equilibrium with π_1=V(N): In the first stage, agent 1 chooses S=N-{1} and all agents except agent 1 play "T". In the second stage, if more agents recommended x than -x, agent 1 votes x; if there is a tie in the votes of S, he plays "T". We will show that this is the only sequential equilibrium.

Note first that in equilibrium, the following must hold: $\pi_{2,1}=1$ and $\pi_1 \geq V(1)$.

Consider an equilibrium in which agent 1 chooses the set S and the members of S play $(s_i)_{i\in S}$. Denote by S_c, S_T, and S_{NT}, the sets of agents in S which choose "c", "T" and "NT" respectively. Clearly, $|S_T| \geq |S_{NT}|$; otherwise, π_1 is less than V(1).

We will show that agent 1's optimization implies that no agent in S plays "NT" and that $k=|\,|S_0|-|S_1|\,|\leq 1$.

Agent 1 is informed only about the number of members of S who voted 1. Let S_Δ be a subset of S_T so that $|S_\Delta| = |S_T| - |S_{NT}|$. Let δ be the difference between the number of 1s and 0s in the votes of S_Δ, and let δ' be the difference between the 1s and 0s in the vote of $S_T \cup S_{NT} - S_\Delta$. Consider an auxiliary problem in which agent 1 can determine the consequence (rather than just adding his vote), based on his signal, δ and δ'. Note that this information is finer than what agent 1 actually obtains in our mechanism.

The variable δ' is uninformative since for any α, the probability that $\delta'=\alpha$ given $\omega=0$, is equal to the probability that $\delta'=\alpha$ given $\omega=1$. Thus, in the auxiliary problem, agent 1 bases his decision solely on δ.

Let us distinguish between two cases.

(1) $|S_\Delta|$ is <u>even</u>. The best course of action for agent 1 in the auxiliary problem is to determine the consequence according to his signal if $\delta=0$, to choose 1 if $\delta \geq 2$ and to choose 0 if $\delta \leq -2$. In our mechanism no matter what agent 1 does in the second stage, he cannot increase π_1 above what he would achieve in the auxiliary problem. But if it is not true that $S = S_\Delta$ and $k=0$, agent 1 can profitably deviate by excluding from S all those who are not members of S_Δ.

(2) $|S_\Delta|$ is <u>odd</u>. The best course of action for agent 1 in the auxiliary problem is to choose 1 if $\delta > 1$, to choose 0 if $\delta < -1$ and, in the case that $\delta=1$ or $\delta=-1$, to determine the consequence either according to his signal or according to the majority vote in S_Δ. In

our mechanism, agent 1 can achieve the solution value of the auxiliary problem by including all members in S_Δ, as well as some who play "0" or "1", as long as $k \leq 1$. The value of π_1 will be lower in any other case.

Thus, we are left with two possibilities:

1) All members of S adopt the strategy "T" and $k=0$. Agent 1, in case of a tie among members of S, plays "T". The outcome of the vote among $S \cup \{1\}$ is identical to the outcome of the direct simultaneous mechanism with the set of voters $S \cup \{1\}$. From Claim 2 in Appendix 3, it follows that an agent $i \in S$ who plays "c" will not affect π_1 by switching to "T" but will increase $\pi_{2,i}$. Also, all agents outside S play "T" (they cannot affect π_1 and hence maximize $\pi_{2,i}$). Thus, it must be that $S=N-\{1\}$ and that all agents in S play "T".

2) No agent in S plays "NT" and $|k|=1$. Assume, without loss of generality, that $|S_0| = |S_1|+1$. It follows that $|S_\Delta|$ is odd. No such equilibrium is possible since, we will show, $i \in S_0$ can profitably deviate to "T". Note, that such a deviation cannot lead to an out-of-equilibrium event. If agent 1 plays "T" in case of a tie, then the same argument applied in possibility 1 above, applies here. If agent 1 votes 1 in case of a tie, then $\pi_{2,i} < 1/2$ and $\pi_1=V(|S_\Delta|-1)$. By switching to strategy "T", agent i does not affect π_1 but strictly increases $\pi_{2,i}$.

Appendix 3

In this appendix we present two useful probabilistic facts.

Let S be a subset of agents and $\{s_i\}_{i\in S}$ be a profile of strategies, each of which belongs to {"T","NT","0","1"}. Denote by $\pi_1(\{s_i\}_{i\in S})$ the probability that the majority of the recommendations in S will coincide with the state of nature given that the agents play the profile of strategies $\{s_i\}_{i\in S}$. Denote by N_x the number of agents who choose strategy x. Denote by M_y the number of agents in S who report y.

Claim 1: $\pi_1(\{s_i\}_{i\in S})$ is strictly increasing when s_i="NT" is replaced with s_i="T".

Proof: By changing the strategy from "NT" to "T", agent i affects the consequence in two events:

E_1: The set of all instances where the state is ω, $x_i=-\omega$, $M_\omega-M_{-\omega}=1$. In this case, switching from "NT" to "T" decreases π_1.

E_2: The set of all instances where the state is ω, $x_i=\omega$, $M_{-\omega}-M_{\omega}=1$. In this case, switching from "NT" to "T" increases π_1.

Note that in any instance belonging to one of these two events, the number of agents other than agent i who report 0 equals the number of agents who report 1. The vector $(\omega,x_1,..,x_i=-\omega,..,x_n)$ belongs to E_1 iff $(\omega,x_1,..,x_i=\omega,..,x_n)$ belongs to E_2. The number of mistakes (wrong signals) in the instance $(\omega,x_1,..,x_i=-\omega,..,x_n)$ is larger by one than the number of mistakes in the instance $(\omega,x_1,..,x_i=\omega,..,x_n)$ due to an additional mistake, the one in the signal of agent i. Therefore, the probability of E_1 is smaller than that of E_2. Thus ,switching from "NT" to "T" increases π_1.

Claim 2: Assume none of the strategies is "NT", $N_0 \geq N_1$, and $s_i=$"0". A switch of agent i to "T" increases π_1 if $N_0 > N_1$ and does not change π_1 if $N_0=N_1$.

Proof: A change of agent i's strategy from "0" to "T" changes the consequence in the following two events.

E_1: $\omega=0$, $x_i=1$ and $M_0-M_1=1$.

E_2: $\omega=1$, $x_i=1$ and $M_0-M_1=1$.

In terms of increasing π_1, the strategy "T" is better than "0" if and only if E_2 is more likely than E_1, which is true if and only if $(N_T-N_0+1+N_1)/2 < 1+(N_T+N_0-1-N_1)/2$ (the number of mistakes among {j | j=i or j uses "T"}), which is equivalent to $N_1 < N_0$. Similarly, "T" and "0" yield the same π_1 if and only if $N_1=N_0$.