JOURNAL OF ECONOMIC THEORY **79**, 157–173 (1998)
ARTICLE NO. ET972385

Motives and Implementation: On the Design of Mechanisms to Elicit Opinions

Jacob Glazer

The Faculty of Management, Tel Aviv University, Tel-Aviv, Israel 69978
E-mail: glazer@post.tau.ac.il

and

Ariel Rubinstein

The School of Economics, Tel-Aviv University, Tel-Aviv, Israel 69978; Department of Economics, Princeton University, Princeton, New Jersey 08544
E-mail: rariel@ccsg.tau.ac.il

Received April 18, 1997; revised December 4, 1997

A number of experts receive noisy signals regarding a desirable public decision. The public target is to make the best possible decision on the basis of all the information held by the experts. We compare two "cultures." In one, all experts are driven only by the public motive to increase the probability that the desirable action will be taken. In the second, each expert is also driven by a *private motive*: to have his recommendation accepted. We show that in the first culture, every mechanism will have an equilibrium which does not achieve the public target, whereas the second culture gives rise to a mechanism whose unique equilibrium outcome does achieve the public target. *Journal of Economic Literature* Classification Numbers: C72, D71. © 1998 Academic Press

1. INTRODUCTION

Motives are basic building blocks of decision makers' preferences. For example, a parent's preferences over his child's school may combine educational, religious, and social motives. A consumer's preferences over bundles of food may combine the motives of taste, health and visual appearance. A voter's ranking of different candidates for a political position may be formed by motives such as the candidates' attitudes towards issues of security, foreign affairs, welfare, or their private life.

Within a society we often observe uniformity in the motives that drive its members, although the weights assigned to the different motives may not be uniform. We refer to the set of motives that drive the members of a certain society as a culture. In some cultures, for example, the private life

0022-0531/98 $25.00

of a candidate is a nonissue for the voters, whereas in others it is important. In some cultures, voters are expected to care only about the "well-being of the nation" whereas in others it is considered admissible for them to be driven by personal interests too.

One way to compare between cultures is at the normative level, asking whether the motives in one culture are more moral than those in the other. This paper takes the different approach of comparing cultures on the basis of the implementability of social targets. Given a certain social target and a set of different cultures, we ask, for each culture, whether there is a mechanism that yields the social target, in any society in which all individuals are guided only by the motives of this culture.

In our model, the decision on whether to take a certain public action is made on the basis of the recommendations of a group of experts, each of whom holds some information about the social desirability of the action. We have in mind situations such as: a group of referees determines whether a paper is accepted or rejected, each referee having an opinion regarding the acceptability of the paper; a decision is to be made whether or not to operate on a patient on the basis of consultations with several physicians; an investigator must determine whether or not a certain event has occurred, based on the evidence provided by a group of witnesses. In such scenarios, different agents may have different opinions, due to the random elements which affect their judgments, and this randomness is the principal rationale for making such decisions on the basis of more than a single opinion.

The *public target* (PT) is to take the best action, given the aggregation of all sincere opinions. To gain intuition about the difficulties in implementing the PT, consider the mechanism with three experts asked to make simultaneous recommendations and the alternative that gets most votes is executed. If all the experts only care about the PT, this mechanism has the desired equilibrium of all experts making sincere recommendations. However, other equilibria also exist. For example, the one where all the experts recommend the same action, regardless of their sincere opinions. This "bad" equilibrium will be strengthened if each expert is also driven by the desire that his recommendation be accepted, since a deviation increases the chance that his will be the minority opinion. This equilibrium will be strengthened even further if the strategy to always recommend the same action is less costly for an expert than the sincere recommendation strategy (which, for example, requires a referee to actually read the paper).

The main point of this paper is the comparison between two cultures: One in which the experts are driven only by the *public motive* (i.e., they want the action to be taken only if it should be taken according to the social objectives), and the other, in which each expert is also driven by a *private motive* (i.e., he wants the public action to coincide with his recommendation). We find that in the culture where all experts are driven only by the public motive the social target cannot be implemented: Every mechanism also has a bad equilibrium in which the probability that the right decision will be made is not higher than the probability obtaining if only one expert was asked for his opinion. On the other hand, in the culture in which both motives exist, the social target is implementable: There is a mechanism that yields only the desirable outcome, regardless of the experts' trade-off between the public and the private motives.

The introduction of private motives is a departure from the standard implementation literature and can also be viewed as a criticism of that literature. In the standard implementation problem, the designer is endowed with a set of consequences which he can use in the construction of the mechanism. The definition of a consequence does not contain details about the events that take place during the play of the mechanism and the agents' preferences are defined only over these consequences. This assumption is very restrictive as it assumes that preferences are not sensitive to events that take place during the play of the mechanism. In the context of our paper, for example, even if an expert is initially concerned only about the public target, if asked to make a move, interpreted as a recommendation, he may also wish that his recommendation be accepted. The implementation literature ignores the possibility that such a motive will enter into an expert's considerations and treats the expert's moves in the play of the mechanism as meaningless messages.

Ignoring mechanism-related motives may yield misleading results. For example, consider the case where a seller and a buyer evaluate an item with reservation values s and b, respectively. The designer wishes 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 suggests the seller make a "take it or leave it offer" as a solution to this problem. However, this "solution" ignores the emotions aroused when playing this mechanism. A buyer may consider the offer of a price which leaves him with less than, say, 1% of the surplus insulting. Although he may prefer to get 1% of the surplus to rejecting the transaction if offered by "nature," he will nevertheless prefer to reject an offer of 1% made by the seller. A "defense" which may be made by the implementation literature is that the moves in a mechanism are abstract messages. However, the attractiveness of a mechanism should be judged, in our view, by its interpretation. The "take it or leave it" mechanism is attractive because the first move is interpreted as a price offer and not as a meaningless message. However, in real life, an attractive interpretation may also be associated with additional motives which cannot be ignored.

Interestingly, in our problem, the introduction of the private motive does not hamper but rather helps to implement the PT. This, however, does not diminish the significance of our general point: individuals are not indifferent to the content of the mechanism, as assumed by the standard implementation literature.

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 desirable 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 $1 > p > 1/2$ and the value $-\omega$ with probability $1 - p$ (we use the convention that $-1 = 0$ and $-0 = 1$). The signals are conditionally independent.

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 certain other informational structures, where such an outcome may signal the diminishing importance of the decision.

Let $V(K)$ be the probability that the desirable action is taken if, for every realization of K signals, the action taken coincides with the majority of the K signals. That is, for any given K agents,

$$V(K) = \text{prob}\{\text{strict majority of the } K \text{ agents get the right signal}\}$$
$$+ 1/2 \text{ prob}\{\text{exactly one-half of the } K \text{ agents get the right signal}\}.$$

Note that $V(K)$ is the highest probability that the desirable action is taken, over all mechanisms that employ K signals.

An important property of the function V is that for every k, $V(2k) = V(2k - 1)$ and $V(2k + 1) > V(2k)$. The fact that V is only weakly increasing is a special case of the observation made by Radner and Stiglitz [8] that value of information functions are often not concave, that is, the marginal value of a signal is not decreasing in the number of signals. The equality $V(2k) = V(2k - 1)$ follows from our symmetry assumptions but it holds under less restrictive conditions (see Section 5 for a detailed discussion of this issue).

We define a *mechanism* as the operation of collecting information from the agents, calculating the consequence, and executing it. We model a mechanism as a finite extensive game form with imperfect information (but no imperfect recall), with the n agents being the players, no chance players and with consequences being either 0 or 1.

The following are examples of mechanisms.

The direct simultaneous mechanisms. All agents simultaneously make a recommendation, 0 or 1, and the majority determines the consequence.

The direct sequential mechanism. The agents move sequentially 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 $1, 2, ..., n-1$ simultaneously make recommendations, 0 or 1, which are transferred to agent n (the "leader"), who makes the final recommendation that 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 maximize π_1, the probability that the desirable action will be taken (the consequence ω at state ω). This definition assumes that the loss entailed in making the mistake of taking the action 0 at state 1 is the same as the loss entailed in making the mistake of taking the action 1 at state 0.

Each agent i can be driven by at most two motives, public and private. The *public motive*, which coincides with the PT, is to maximize π_1. The *private motive* is to maximize $\pi_{2,i}$, the probability that his recommendation coincides with the consequence of the mechanism. In order to precisely define the private motive we 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 $h \in R_i$, player i has to choose between two actions named 0 and 1, and that there is no terminal history h which has two subhistories in R_i. Whenever we discuss the private motive of agent i, we will refer to the comparison between the consequence of the played path and the agent's action at that (at most one) subhistory along this path which is in R_i.

When we say that agent i is driven only by the public motive, we mean that he wishes only to increase π_1. When we say that he is driven by both the private and the public motives we mean that he has certain preferences strictly increasing in *both* π_1 and $\pi_{2,i}$.

Our analysis ignores the existence of other private motives. For example, after the decision regarding an operation on a patient is made, some additional information may be obtained identifying the right action ex post. Then, a new motive may emerge: the desire of each physician to be proven ex post right. We do not study cultures with this motive and our analysis best fits situations in which the "truth" never becomes apparent.

The concept of equilibrium we adopt is sequential equilibrium in pure strategies (for simultaneous mechanisms this coincides with the Bayesian–Nash equilibrium). Given a profile of preferences over the public and the private motives, we say that a mechanism *implements the PT* if in every sequential equilibrium of the game, $\pi_1 = V(n)$.

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, for any mechanism, the game obtained by the mechanism coupled with the agents' objective of increasing π_1 only, has a sequential equilibrium with $\pi_1 < V(n)$.

In order to achieve a better understanding of the 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 has received. "NT' is the strategy whereby an agent who has received the signal x recommends $-x$ and "c" ($c = 0, 1$) is the strategy whereby an agent announces c independently of the signal he has 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" (since $n \geqslant 3$ a single deviation of agent i will not change π_1); and

(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 is a natural focal mode of behavior. However, the notion of implementation which we use does not relate any focal status to truth-telling. Note that although we do not include the cost of implementing a strategy in the model, one 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 increase π_1 but will impose some costs on the agent.

Note also that in this game, the strategy "T" is *not* a dominant strategy (not even weakly) when $n > 3$. For example, for $n = 5$, 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." These strategies lead to different outcomes only when agents 3 and 4 get the signal 1 and agent 5 gets the signal 0. The strategy

"1" is better for agent 5 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 too 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 match 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 then $\pi_1 = V(n-2)$, and if he plays "T" instead, 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 agents and assigns no significance to deviations.

In all of the above mechanisms there is an equilibrium which is optimal in the sense that it maximizes π_1 over all strategy profiles. This equilibrium strategy profile will emerge if each agent follows a general principle which calls him to follow his share in a profile which is both Pareto optimal and a Nash equilibrium, if such a profile exists. One may argue that this belies the importance of the problem we are considering. We disagree.

First, on the basis of casual empirical observation we note that groups of experts are often "stuck" in bad equilibria. The reader will probably have little difficulty recalling cases in which he participated in a collective decision process and had a thought of the type: "There is no reason for me to seriously consider not supporting α, since everybody else is going to support α anyway." Second, note that even though an agent in this section is driven only by the public motive, we think about him as having another motive in the background: to reduce the complexity of executing his strategy. If agents put a relatively "small" weight on the complexity motive, truth telling remains a unique Pareto-optimal behavior which is a Nash equilibrium; however it is less obvious that an agent will indeed invoke this principle, since the complexity motive plays against it.

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 the probability that would obtain, were a single agent nominated to make a decision based only on the signal he receives.

PROPOSITION 1. *If all agents are only interested in increasing* π_1, *then every mechanism will have a sequential equilibrium with* $\pi_1 \leqslant V(1)$.

For the proof see Appendix 1. Here, we provide an intuition for the main idea of the proof. Consider first a one-stage, simultaneous-move mechanism. We will construct a sequential equilibrium with $\pi_1 \leqslant V(1)$. 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 any agent $j \neq i$ to play the action a_j independently of the signal he has received. Assign agent i to play the action b_x if he has received the signal x. This profile of strategies yields $\pi_1 = V(1)$, and any deviation is unprofitable since it makes the outcome of the mechanism depend on at most two signals, but $V(2) = V(1)$.

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 yet determined, assign strategies in a manner similar to the one we used 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 the sequential equilibrium constructed for the one-stage mechanism above. Otherwise, assign each agent i to play an action a_i^* at the first stage independently of his signal, where (a_i^*) is a profile of actions which does not determine the consequence of the mechanism. Coupling with beliefs that do not assign any significance to deviations in the first stage, we obtain a sequential equilibrium with $\pi_1 = V(1)$.

Virtual Implementation

Virtual Bayesian Nash implementation of the PT may be possible. Abreu and Matsushima [1] suggest a direct simultaneous mechanism according to which the outcome is determined with probability $1 - \varepsilon$ by the majority of announcements and with probability ε/n by agent i's recommendation ($i = 1, ..., n$). This mechanism requires the use of random devices and it allows the unsound possibility that while $n-1$ agents observe and report the signal 0, the outcome is 1.

Related Literature

Up to this point our analysis is a standard investigation of a problem of sequential equilibrium implementation with imperfect information (see Moore [4] and Palfrey [6])). A related model is Example 2 in Palfrey and Srivastava [7] which differs from ours as each agent prefers that the social action will coincide with the signal he has received. Both models demonstrate the limits of Bayesian implementation. Proposition 1 is related to results presented in Jackson [3] 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. Proposition 1 does not follow from Jackson's results since we also refer to extensive and not solely to simultaneous mechanisms.

4. IMPLEMENTATION IS POSSIBLE WHEN ALL AGENTS ARE DRIVEN BY BOTH MOTIVES

We now move from the culture in which all agents are driven only by the public motive to the culture in which they are driven by both the public and the private motives. We show that here implementation of the PT is possible.

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 even numbered (and may be empty) and it should not include the controller. In the second stage, the controller learns the *result* of the votes cast by the members of S and only then adds his vote. The majority of the votes in $S \cup \{1\}$ determines the outcome.

Following are three points to note about this mechanism.

(1) The controller has a double role. First, he has the power to discard the votes of those agents who play a strategy that negatively affect π_1. Second, he contributes his own view whenever his vote is pivotal.

(2) Each agent (except the controller) makes a recommendation in the first stage even if his vote is not counted. An agent whose vote is not counted is driven only by the private motive and hence will vote honestly

if he believes that the outcome of the mechanism will be positively correlated with the signal he receives.

(3) Whenever the controller is pivotal, his recommendation will be the outcome, when he is not pivotal, he does not reduce π_1 by joining the majority. Thus, the mechanism is such that the private motive of the controller never conflicts with his public motive.

We will prove that this mechanism implements the PT for every profile of preferences in which the agents are driven by both the public and the private motives (independently of the weights they assign to the two motives as long as both weights are positive). For every game induced by the mechanism and a profile of such preferences, the only equilibrium is one where, in the first stage, all agents other than the controller play "T" and they are all included in S, and in the second stage, the controller joins the majority in S unless he is pivotal, in which case he plays "T."

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.*

The detailed proof is given in Appendix 2. Following are its main arguments, showing that no other equilibria are possible:

(1) The controller's decision whether to include in S an agent who plays "NT" is the result of two considerations: the information he obtains from such an agent, and the fact this agent's vote negatively affects the outcome. We will show that the latter is a stronger consideration and, therefore, agents who play "NT" are excluded from S.

(2) Since the mechanism enables the controller to maximize his 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" (the difference between the numbers of "0"s and "1"s cannot exceed 1).

(3) There is no equilibrium in which some of the agents in S choose a pooling strategy ("c"), since one of them increases $\pi_{2,i}$ without decreasing π_1 by switching to "T."

(4) There is no equilibrium with $S \neq N - \{1\}$. If agent i is excluded from S, then by (2) he does not play "T," but since he does not affect the consequence and since in equilibrium $\pi_1 > 1/2$, he can profitably deviate to "T" and increase $\pi_{2,i}$.

For the mechanism to work, it is important that the controller only learns the result of the votes in S, not how everyone voted. 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 how everyone voted. The following is a sequential equilibrium with $\pi_1 < V(3)$: In the first stage, agent 1 chooses $S = \{2, 3\}$, agent 2 plays "0," agent 3 plays "T." In the second stage, 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).

Note that the role of the controller in the first stage of the mechanism is somewhat similar to the role of the "stool-pigeon" in Palfrey [6] and Baliga [2]. The stool-pigeon is an agent appended to the mechanism whose role, as described by Palfrey [6], 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 a previous version of the present paper we showed that Proposition 1 is still valid when the use of a stool-pigeon is allowed. The mechanism of Proposition 2 "works" because *all* agents are also driven by the private motive.

Comment: The Culture with Only the Private Motive

Implementation of the PT is impossible in the culture in which 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 of the PT is impossible in any culture in which all motives are independent of the state. The reason is that in such a culture. whatever the mechanism, if $\sigma = (\sigma_{i,x})$ is a sequential equilibrium strategy profile ($\sigma_{i,x}$ is i's strategy given that he observes the signal x), then the strategy profile σ' where $\sigma'_{i,x} = \sigma_{i,-x}$ (each agent who receives the signal x plays as if he had received the signal $-x$) is also a sequential equilibrium strategy profile. Thus, the outcome of σ when all agents receive the signals 1, is the same as the outcome of σ' when all agents receive the signal 0, and thus one of them does not yield the PT.

5. A DISCUSSION OF THE SYMMETRY ASSUMPTION

One may suspect that symmetry plays a crucial role in obtaining Proposition 1, which is the springboard of our analysis. Indeed symmetry conditions are imposed overall; the two states are equally likely, the loss from taking the action 1 when the state is 0 is equal to the loss from taking the action 0 when the state is 1, the signal random variable is the same for all agents and the probability that the signal is correct, given the state, is independent of the state.

Furthermore, at least one "deviation" from the symmetry assumptions indeed invalidates Proposition 1. Assume that the probability of state 0 is "slightly" larger than the probability of state 1 (so that given only one signal it is best to follow it). In this case $V(2) > V(1) > V(0)$. It is easy to verify that the following simultaneous mechanism implements the PT for the case where there are two agents driven by the public motive only:

	a	b	c
a	0	0	1
b	0	1	0
c	1	0	0

The example above demonstrates that the key element in the proof of the nonimplementability of the PT in the culture with the public motive only, is that $V(2) = V(1)$, an equality which follows from the symmetry assumptions. Thus, one may suspect that we are dealing with a "razor-edged" case. We have three responses.

(1) Deviations from the symmetry assumptions will not necessarily make the PT implementable when all agents are driven by the public motive only. Here are two examples, presented for simplicity for the case where $n = 3$.

(a) Assume that β, the probability of state 1, is such that only if all three agents receive the signal 0, does it become more likely that the state is indeed 0 (i.e., $[p/(1-p)]^3 > \beta/(1-\beta) > [p/(1-p)]^2$). Then, $V(3) > V(2) = V(1)$ and the PT is not implementable.

(b) Assume that the signals observed by the three agents are not equally informative. Denote by p_i the probability that agent i at state ω gets the signal ω. Assume that $p_1 > p_2 = p_3 > 1/2$. Assume that it is optimal not to follow agent 1's signal only if the signal observed by both agents 2 and 3 is the opposite of the one observed by agent 1. Then, it is easy to see that any mechanism has an equilibrium with $\pi_1 = p_1 < V(3)$.

In fact, it can be shown that in every situation where there is a number $k < n$ for which $V(k) = V(k+1)$ but $V(k) < V(n)$, the PT is not implementable when agents are driven by the public motive only.

(2) The main ideas of the paper are also relevant in the asymmetric cases in which V is strictly increasing. Note that in the background of our model one may conceive an additional cost imposed on an agent who executes a strategy which requires him to actually observe the signal before making a recommendation. Denote this cost by γ. Let m^* be the solution of $\max_{m \leqslant n} V(m) - m\gamma$. In other words, m^* is the "socially optimal" number of active agents. Even when it is strictly increasing, the function V is typically not concave. Hence, it is possible that there is an $m \leqslant m^*$ so that $V(m) - V(m-1) < \gamma$. In such a case, the PT is not implementable when agents are driven by the public motive only. The key point is that if $m-1$ agents operate to increase π_1, the marginal contribution of the mth agent is less than his cost.

(3) Finally, we do not agree with the claim that symmetric cases are "zero probability events." The importance of a symmetry condition cannot be judged according to the measure of the number 0.5 in the unit interval. Symmetric models have special status as they fit situations in which all individuals cognitively ignore asymmetries.

APPENDIX 1

Proof of Proposition 1. We provide a proof for the case where the mechanism is one with perfect information and possibly simultaneous moves (see Osborne and Rubinstein [5, p. 102] for a definition). Though the proof here does not cover the possibility of imperfect information, 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 $\pi_1 \leqslant V(1)$. For any nonterminal 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 $d(h^t) \leqslant d(h^{t+1})$ for all t.

The equilibrium strategies are constructed inductively. At the tth stage of the construction, we deal with the history $h^t = h$ (and some of its sub-histories). There are two possibilities. If the strategies at history h 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 $a_i = b_i$ for all $i \neq i^*$ and, 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^* (or b_i^*) if his signal is 0 (or 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.

Beliefs are updated according to the strategies. 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, that for every history h, there is at most one agent whose equilibrium behavior in the game following h depends on his own signal. If the outcome of the subgame starting at h depends on the moves of one of the players, then all players at h still hold their initial beliefs and 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 $\pi_1 = V(N)$. 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 $x_{2,1} = 1$ and $\pi_1 \geqslant V(1)$ must hold.

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 choosing "c," "T," and "NT,"respectively. Clearly, $|S_T| \geqslant |S_{NT}|$; otherwise, $\pi_1 < V(1)$.

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

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 number of 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 *even*. 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 $\delta \geqslant 2$ and to choose 0 if $\delta \leqslant -2$. 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 *odd*. 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 $k \leqslant 1$. 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 $S \cup \{1\}$. It follows from Appendix 3 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, by Appendix 3, $i \in S_0$ 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 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

Claim. Let S be a subset of agents and $\{s_i\}_{i \in S}$ a profile of strategies, each of which belongs to {"T," "0, ", "1"}. Denote by $\pi_1(\{s_i\}_{i \in S})$ the probability that the majority of the recommendations in S will coincide with ω 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.

If $N_0 \geqslant N_1$ and $s_i =$ "0," then a switch of agent i to "T" increases π_1 if $N_0 \geqslant N_1$ and does not change π_1 if $N_0 = N_1$.

Proof. Denote by M_y the number of agents in S who report y. A change of agent i's strategy from "O" to "T" changes the consequence in the following two events.

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

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

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 $\{j \mid j = i \text{ or } j \text{ uses "T"}\}$), which is equivalent to $N_1 < N_0$. Similarly, "T" and "0" yield the same π_1 if and only if $N_1 = N_0$. ▮

ACKNOWLEDGMENTS

We thank Dilip Abreu, Kyle Bagwell, Matt Jackson, Albert Ma, Tom Palfrey, and Mike Riordan for comments on an earlier version of the paper. We also thank the Associate Editor of this journal for his encouragment. J. Glazer acknowledges financial support from the Institute for Business Research in Israel. A. Rubinstein acknowledges partial financial support from the United States–Israel Binational Science Foundation, Grant Number 1011-341.

REFERENCES

1. D. Abreu and H. Matsushima, Virtual implementation in iteratively undominated strategies: Complete information, *Econometrica* **60** (1992), 993–1008.
2. S. Baliga, "The Not-So-Secret-Agent: Implementation and Professional Monitors," Mimeo, Kellogg Graduate School of Management (M.E.D.S.), Northwestern University, 1996.
3. M. Jackson, Bayesian implementation, *Econometrica* **59** (1991), 461–478.
4. J. Moore, Implementation, contracts, and renegotiation in environments with complete information, *in* "Advances in Economic Theory: Sixth World Congress" (J. J. Laffont, Ed.), Vol. I, pp. 182–282, Cambridge Univ. Press, Cambridge, UK, 1992.

5. M. Osborne and A. Rubinstein, "A Course in Game Theory," MIT Press, Cambridge, MA, 1994.
6. T. Palfrey, Implementation in Bayesian equilibrium: The multiple equilibrium problem in mechanism design, *in* "Advances in Economic Theory: Sixth World Congress" (J. J. Laffont, Ed.), Vol. I, pp. 283–323, Cambridge Univ. Press, Cambridge, UK, 1992.
7. T. Palfrey and S. Srivastava, Mechanism design with incomplete information: A solution to the implementation problem, *J. Polit. Econ.* **97** (1989), 668–691.
8. R. Radner and J. Stiglitz, A nonconcavity in the value of information, *in* "Bayesian Models in Economic Theory" (M. Boyer and R. E. Kihlstrom, Eds.), Chapter 3, Elsevier, Amsterdam/New York, 1984.