DIVISION OF THE HUMANITIES AND SOCIAL SCIENCES

CALIFORNIA INSTITUTE OF TECHNOLOGY

PASADENA, CALIFORNIA 91125

INCENTIVE COMPATIBILITY IN RISK ASSESSMENT MECHANISMS

Talbot Page

SOCIAL SCIENCE WORKING PAPER 409

December 1981

Revised June 1982

ABSTRACT

This paper defines a risk assessment mechanism and compares its incentive properties with those of deterministic incentive mechanisms, particularly the Groves mechanism.

Many risk assessments involve prediction for rare or unique events; in such cases there is limited opportunity for feedback and evaluation of the assessment process. To develop a feedback mechanism, the paper requires assessments to be made for indicator events, linked to the rare or unique events of ultimate interest. Assessments are made by several assessors, or assessment techniques, acting in competition. The feedback mechanism is a transfer function based on the probability assessments of all the assessors and the outcome of the indicator event.

The incentive properties of risk assessment mechanisms are in some ways similar to those for deterministic mechanisms and in some ways quite different. The paper defines one risk assessment mechanism that looks like a Groves mechanism: it directly reveals probability and for risk neutral assessors has an unbiased or truthful dominant strategy which is discontinuous and which cannot solve the budget problem. The paper also defines a class of risk assessment mechanisms which do not look like a Groves mechanism; mechanisms in this class have unbiased dominant strategies which are continuous and which do solve the budget problem.

April 23, 1982

INCENTIVE COMPATIBILITY IN RISK ASSESSMENT*

Talbot Page

The purpose of this paper is to apply ideas from the theory of incentive compatibility for public goods demand revelation to the problem of probability revelation in risk assessment. There are several reasons for doing this: to extend the theory to a new case, which is probabilistic and where there need not be public goods; to provide a mechanism of feedback for learning and validation; and to generalize the notion of scoring rules. The link between the theory of incentive compatibility for public goods and risk assessment is achieved by a structural analogy between two models, one for public goods and the other for risk assessment. The analogy is neither a generalization nor a specialization -- different variables play parallel roles and there are some changed roles, too. Because there are both parallels and differences, part of the basic theorems of Groves and Loeb (1975) and Green and Laffont (1979) characterizing demand relevation for public goods carry over to the risk assessment model, but there are surprises as well.

Transfer functions can be viewed as the centerpiece of models of demand revelation for public goods. In this paper we take the same view toward transfer functions for risk assessment. The motivation for doing so is practical as well as conceptual. There are now a large number of risk assessments, which are increasingly used in decisions with potentially large scale consequences. In risk assessment much

of the focus has been forward looking -- developing estimates of probabilities for future events, or for the likelihood of the existence of some state which is not yet known. There has been relatively little backward looking -- looking back to evaluate previous risk assessments in the light of new information when it becomes available. The transfer function provides a means for increasing the emphasis on feedback.

In risk assessment we are interested in developing estimates of the probability of some event, which might be rare (a reactor core meltdown) or unique (chemical X is a carcinogen). There is a principal who makes use of these estimates for some decision. The principal has no direct information of his own on the probability of the event, but relies on the probability estimates provided by agents or assessors. The assessors have information on the state of nature and form inferences on the probability of the event. In general we assume that the assessors have differing amounts and qualities of information.

As part of an introduction, we can speak loosely of three goals for a principal in an assessment process: (a) to produce a "good" consensus estimate among N assessors; (b) to identify, over a limited number of assessment rounds, assessors who have better information or inference skills (or "assessment technique"); (c) and to provide incentives, relative to the principal's budget, to make it worthwhile for the agents to gather information and form probability judgments, a costly exercise. (The desirata -- "good" consensus estimator, "identifiability" of the "best" assessors, and "sharp" incentives -- can be made precise only within contexts of specified models.) As a means toward some mix of these goals, the principal agrees to reward the assessors by a transfer mechanism, which is a function of the revealed probability estimates and whether or not the predicted event occurs (in the appropriate time interval). For the first goal, a "good" consensus estimate is defined as good for some decision purpose. The idea corresponds to the public goods model, where willingnesses to pay are elicited in ways leading to Pareto optimality and the truthfulness of the revelations is "incidental." For the third goal, it is useful that the expected transfer to an assessor be a "sharp" rather than "flat" function of his revealed estimate. The idea corresponds to the desiratum of "individual rationality" in the public goods model.[1]

It is assumed that the assessors have a single goal; to do "as best they can" in response to a given transfer mechanism. The focus of the paper will be on this latter question -- how the assessors might respond to or manipulate various possible transfer mechanisms. A rationale for the focus is that no matter what mix of goals the principal might have, he will be unable to pick a transfer mechanism until he knows something of how the agents might respond to it.

The most basic result in the paper is the development of a new and more formal framework for risk assessment. More specific results are:

(1) The idea of a Groves mechanism is applied to the risk assessment model and a new mechanism is defined for revealing best, truthful probability estimates. There are both parallels and contrasts. For one thing there is no public good in the risk assessment model.

(2) In the risk assessment model, there are other direct revelation mechanisms, besides the Groves-like one, with truthful dominant

strategies for agents with linear, separable utilities (risk neutrality). This is in contrast with the public goods model, where the Groves mechanism is the only direct revelation mechanism with truthful dominant strategies for agents with linear, separable utilities (no income effects). However, finding other mechanisms with truthful dominant strategies in the risk assessment model is not a new result. These mechanisms are built on proper scoring rules, which have been known for forty years. New results are: the mechanisms built on proper scoring rules can be defined to achieve budget balance and to eliminate the requirement of risk neutrality; and the requirement of risk neutrality can also be eliminated for the Groves-like mechanism while preserving its truthful dominant strategies. But it does not appear possible to achieve budget balance for the Groves-like mechanism in the risk assessment model.

(3) Even though the Groves-like mechanism and proper scoring rules look very different, there is a close connection between them. One of the proper scoring rules, the Brier rule, is shown to be a special case of the Groves-like mechanism from the point of view of a risk neutral assessor.

(4) Under slight restrictions, a Nash equilibrium is shown to exist for the parimutuel betting rule. This widely used transfer mechanism is especially interesting from the point of view of risk assessment because not only is it the most prominant example of assessment mechanism in actual operation, it is also closely related to Bayesian inference.

(5) The expected value maximizing strategy is derived for the parimutuel mechanism. For a particular model of the assessors' information, this mechanism is shown to be more efficient than the Brier rule in identifying the best assessor over a limited number of assessment rounds. This is an interesting result because the Brier rule has been used to distinguish among assessment abilities of weather forecasters. It is interesting for another reason as well: the parimutuel mechanism has a non-truthful dominant strategy. Thus it is possible to find a manipulative mechanism superior in a least one respect to a truth revealing dominant strategy mechanism. But as a final surprise, for this same model of the assessors' information, the Groves-like mechanism, which also has a truth revealing dominant strategy, identifies the better informed assessor more efficiently than either of the other two mechanisms. Equivalence in expectation between the Brier rule and a special case of the Groves-like mechanism one does not imply equivalence in other aspects of the rules' behavior.

Concern with incentive compatibility in risk assessment goes back at least to 1950, when Brier (1950) proposed a forecasting verification system which would be immune to manipulation, or as he put it, "playing the system." Savage (1971) devoted his last paper to the problem of eliciting truthful revelation of each assessor's best judgmental probability estimate (truthful revelation rules for expected value maximizing assessors came to be known as proper scoring rules). Recently Grether has developed a procedure, not depending on risk neutrality, for truthful revelation of probability estimates.

The work just cited can be viewed in the context of 2 person games -- a single assessor versus nature, with the game established and overseen by the principal. There appears to be little previous work on the problem addressed in this paper, which can be viewed as an N + 1 person game, where N assessors may affect each other's transfer, and where there may be strategic manipulation of the mechanism in terms of one assessor against each other. Page (1977) stated this problem of manipulative competition among N risk assessors and considered a Bayesian transfer rule because it might have useful properties for the principal. However, the strategic analysis was not carried far and there appears to be little existing work on the strategic interactions among competing assessors.

On a related subject, calibration and the evaluation of judgmental probability assessment, there has been a great deal of work. Lichtenstein, Fischhoff, and Phillips (1981) provide a recent summary, and their own work on the subject needs particular mention here, as does the work by Winker and Murphy, who have contributed especially on the evaluation of probabilistic weather forecasting. As a means of identifying the "best" assessor Roberts (1965) suggested a Bayesian scoring rule, but he did not consider its strategic properties, a few of which are discussed by Winkler (1969).

I. Analogy Between Models of Incentive Compatibility for Public Goods and Risk Assessment

Definition. A risk assessment mechanism is a function $f = (t_1,\ldots t_N)$, where t_i is the transfer to i; t_i is a function of $p_1,\ldots p_N$ and X; p_i is i's revealed probability estimate; and X is the assessed event. The strategy space for each i is [0,1], and X can be either 0 or 1.[2]

The analogy between a model of incentives compatibility and a model of risk assessment is drawn in Table 1 and the subsequent paragraphs.

Table 1

	Public Goods Model	Risk Assessment Model
Utility function for the principal	$U^o(\Sigma t_i)$ $U^o_1 < 0$	$U^o(\Sigma t_i, K, X)$ $U^o_1 < 0$
Utility function for agent i	$U^i(t_i,K)=t_i+g_i(K)$	$U^i(t_i,z_i)=t_i+g_i(z_i)$ $U_2{}^i < 0$
Transfer mechanism	$t_i=f_i(w_1,\ldots,w_N)$	$t_i=f_i(p_1,\ldots,p_N,X)$
Decision function for the principal	$K=d(w_1,\ldots,w_N)$	$K=d(p_1,\ldots,p_N,I)$

As part of the analogy w_i (revealed willingness to pay in the demand revelation model) corresponds to p_i (revealed probability estimate in the risk assessment model). Underlying w_i is v_i (i's true valuation) which corresponds to $\overline{p}_i$ (i's true or "best" probability assessment). In the public goods model K is a public good for the N agents; K appears in U^i without the subscript, but not in U^o. In the risk assessment model the situation is reversed. K is a private good appearing in U^o but not in the U^i. The public goods model is an N person non-cooperative game among the N voters or agents, overseen by the principal who receives (or pays) the transfers. The risk assessment model is an N+1 person non-cooperative game, like the N+1 person game developed by Harsanyi and recently applied by Milgram and Roberts (1982). The extra player is nature, which "decides" on the existence of some (possibly unique) event X. The state of X is not known to the N assessors at the time of their revelations. The state of X is revealed after the p_i are revealed and the transfer t_i is based on the observed state of X along with the p_i.

For example, each of N forecasters makes a probability prediction of rain tomorrow. The following day the forecasters are given transfers as functions of the revealed predictions $(p_1,\ldots,p_N)$ and whether or not it rained. The dimensionality of K in the demand revelation model corresponds to the dimensionality of X in the risk assessment model, in the following way. If K is dichotomous (K = 1 means the public good is chosen, K = 0 means the status quo is preserved), only a single value for willingness to pay is elicited from each agent. Similarly if X is dichotomous ($X = 1$ means rain; $X = 0$ means no rain), only a single probability is elicited. If K is continuous in the demand revelation model, then an entire willingness-to-pay schedule is elicited. Similarly if X is continuous a probability density function is elicited.

In the demand revelation model, it is assumed that agents know their true valuations (the v_i) at no cost. In contrast, to form a probability judgment requires assessment activity. To keep matters simple we will assume that i can form at least a crude best guess $\bar{p}_i$ at no cost. But for i to refine his judgment $\bar{p}_i$ "more" assessment activity z_i is required. The more information which is gathered and processed (the higher z_i), the more costly the assessment activity to i. We will say a few words later about "better" judgments of probability of X and the costs of attaining them in Section VII.

II. Existence of a Groves-like Mechanism

The idea of the Groves-like mechanism is as follows. Each of N assessors makes his own assessment of the probability of the event X. Each i reports a probability p_i of the event without knowing the others' reported assessments. For each assessor i, the consensus of the other N-1 assessors is defined and specified q_i. Then if the event occurs, i wins if $p_i > q_i$ (his reported probability is higher than the others' consensus) and i loses if $p_i < q_i$. And if the event does not occur, i wins if $p_i < q_i$ and i loses if $p_i > q_i$. How much he wins in each case is determined by the others' consensus. The amount of a win and the resolution of ties is provided by the definition:

<u>Definition</u>. Define $q = \sum_{i=1}^{N} p_i c_i$ as the consensus of all the assessors and $q_i = \sum_{j \neq i}^{N} p_j c_j / (1 - c_i)$ as the consensus of all but i, where for all i $0 \leq c_i < 1$ and $\sum_{i=1}^{N} c_i = 1$. Then the Groves-like risk assessment mechanism $t_i = f_i(p_1, \ldots, p_N, X)$ is defined by

$$t_i = \begin{cases} 1 - q_i & \text{if } X = 1 \text{ and } p_i \geq q_i \\ q_i & \text{if } X = 0 \text{ and } p_i < q_i \\ 0 & \text{otherwise} \end{cases}$$

For short we will refer to this transfer function as mechanism f. In the symmetric case $c_i = 1/N$ and $q_i = \frac{1}{N-1} \sum_{j \neq i}^{N} p_j$.

We can think of the c_i as the credibility weights used in defining the consensus estimates. To stress the parallel the following theorem is proved for assessors with linear, separable utility (risk neutrality for the risk assessment model, no income effects for the public goods model).

Theorem 1. Mechanism f has a truth revealing dominant strategy for each risk neutral, expected value maximizing assessor. Moreover, truthful revelation is the unique dominant strategy.

Proof. We will show that truthful revelation is the unique dominant strategy for any risk neutral assessor i. For convenience, omit the subscript from p_i, q_i, t_i and $\overline{p}_i$ ($\overline{p}_i$ is i's truthful best judgment of $P(X=1)$). If $p \geq q$, i's expected transfer from his point of view is $E(t) = \overline{p} - \overline{p}q$; and if $p < q$, $E(t) = q - \overline{p}q$. For any $\overline{p}$ there are three possible cases: $q > \overline{p}$, $q < \overline{p}$, and $q = \overline{p}$. We consider the three cases and all the possible subcases in Table 2.

Table 2

	Subcase		Strategy	Expected Transfer
Case 1 $q > \overline{p}$	(1a)	$p=\overline{p}<q$	truthful revelation	$q - \overline{p}q$
	(1b)	$p<\overline{p}<q$	biased down	$q - \overline{p}q$
	(1c)	$\overline{p}<p<q$	biased up	$q - \overline{p}q$
	(1d)	$\overline{p}<q\leq p$	biased up	$\overline{p} - \overline{p}q$
Case 2 $q < \overline{p}$	(2a)	$q<\overline{p}=p$	truthful revelation	$\overline{p} - \overline{p}q$
	(2b)	$q\leq p<\overline{p}$	biased down	$\overline{p} - \overline{p}q$
	(2c)	$p<q<\overline{p}$	biased down	$q - \overline{p}q$
	(2d)	$q<\overline{p}<p$	biased up	$\overline{p} - \overline{p}q$
Case 3 $q = \overline{p}$	(3a)	$q=\overline{p}=p$	truthful revelation	$\overline{p} - \overline{p}^2$
	(3b)	$p<q=\overline{p}$	biased down	$\overline{p} - \overline{p}^2$
	(3c)	$q=\overline{p}<p$	biased up	$\overline{p} - \overline{p}^2$

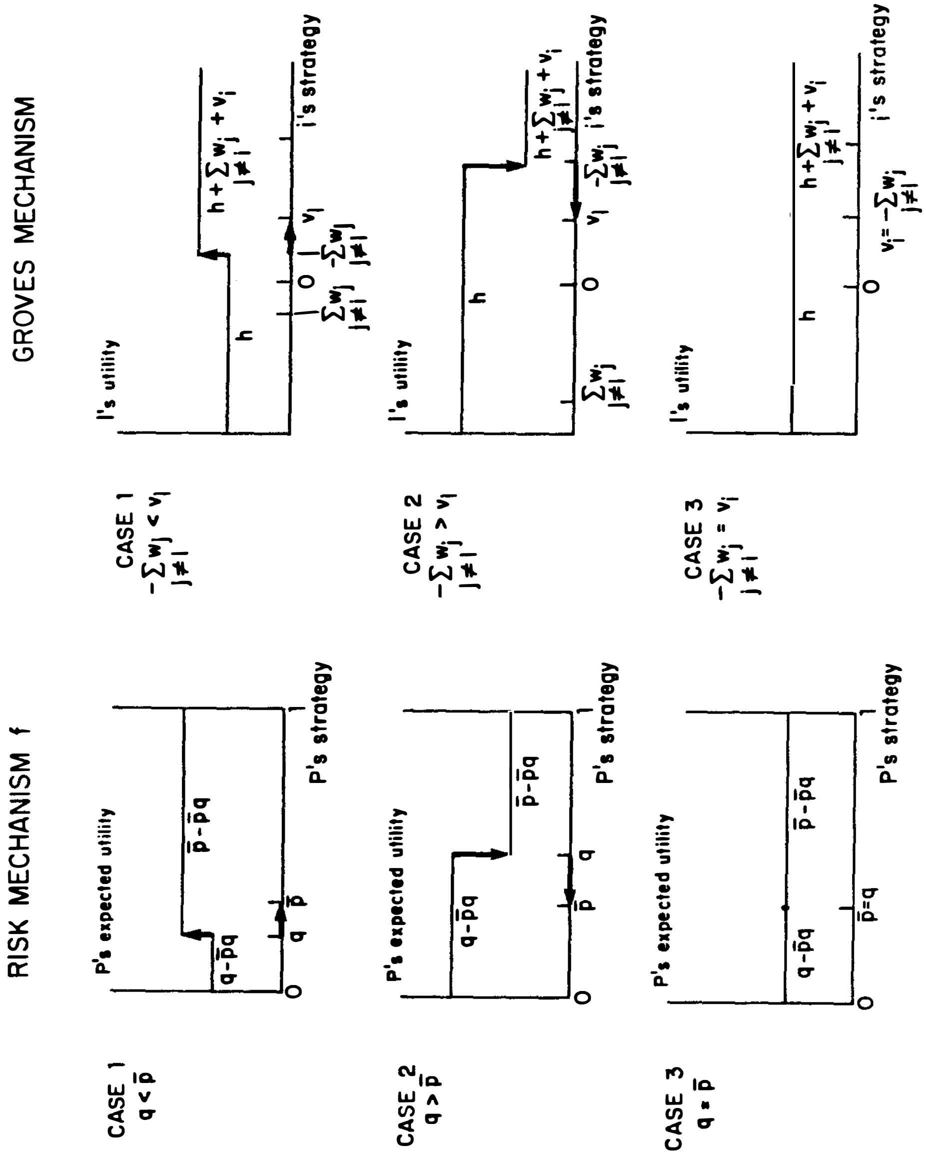
The expected transfer in subcase (1d) is smaller than for (1a), (1b), and (1c) because in case 1 $q > \overline{p}$; and the expected transfer in (2c) is smaller than in (2a), (2b) and (2d) because in case 2 $q < \overline{p}$. So in each case truthful revelation is a least as good and sometimes better than any alternative strategy. This implies that truthful revelation is a dominant strategy. From the enumeration of cases, it is clear that no other strategy is at least as good as truthful revelation in all cases. Thus truthful revelation is the unique dominant strategy.

Q.E.D.

Like the Groves mechanism the decision of who wins (who pays in the public goods model) is split from the decision of how much a winner gets (how much is paid). The decision of whether or not i wins depends on all the p_j and on X. How much depends only on p_j for $j \neq i$, and on X. Each i's expected return is discontinuous with the point of discontinuity at the point q_i. The close parallel (and difference) between the two mechanisms is shown in Figure 1.[3]

A major difference between the two models is that Theorem 1 can be strengthened by modifying mechanism f and dropping the assumption of linear, separable utility. In its place we substitute the weaker assumption of monotoxicity of preferences over lotteries.[4] This is done in Theorem 2.

Theorem 2. Define p_i and q_i as in Theorem 1, and define $t_i = f_i'(p_1, p_2, \ldots, p_N, X)$, by



$$t_i = \begin{cases} \left.\begin{array}{ll} 1 & \text{w.p. } 1 - q_i \\ 0 & \text{w.p. } q_i \end{array}\right\} & \text{if } X = 1 \text{ and } p_i \geq q_i \\ \left.\begin{array}{ll} 1 & \text{w.p. } q_i \\ 0 & \text{w.p. } 1 - q_i \end{array}\right\} & \text{if } X = 0 \text{ and } p_i < q_i \\ 0 & \text{otherwise} \end{cases}$$

Then, if i's preferences over lotteries are monotonic, mechanism $f' = (f'_1, \ldots f'_n)$ has a truthful dominant strategy, and truthful relevation is the unique dominant strategy.

Proof. Construct a table of subcases, which looks like Table 1 except for the heading on the third column. This time the entries in the third column refer to probabilities of winning a lottery. For each subcase there is a lottery of winning 1 unit with some probability s and of winning 0 with probability 1 - s. For each subcase the values for s are shown in the third column. Inspection of Table 1 shows that for any i with monotonic preferences over lotteries truthful revelation is the unique dominant strategy.

Q.E.D.

Theorem 2 turns on the fact that the transfer to i in Theorem 1 is bounded between zero and 1, a fact obvious from the definition of f. The following definitions will prove useful.

Definitions. For a given transfer mechanism $g = (t_1, \ldots, t_n)$ define $b^- = \inf.\ t_i$; $b^+ = \sup.\ t_i$; $B^- = \inf.\ \sum_{i=1}^{N} t_i$; and $B^+ = \sup.\ \sum_{i=1}^{N} t_i$, where p_i can vary over $[0,1]$ and X can either be 0 or 1.

Definition. If $B^- = B^+$ we say that the principal's budget is controlled at level B.

Budget control corresponds to the condition of "balanced budget" for the public goods model. Budget control or balance means that the principal knows the cost of the transfer mechanisms beforehand; it also facilitates comparisons among transfer mechanisms. It is easy to see that the Groves-like mechanism does not have budget control. For a counter example, note for N=2, if X=1 and $p_1 > p_2$, $\sum t_i = 1 - p_2$ which varies with p_2.

III. Mechanisms with Truthful Dominant Strategies, Budget Control, and Continuity

We know that in the public goods model the Groves mechanism is the only direct relevation mechanism with a truthful dominant strategy. The situation is quite different for the risk assessment model. Here there are an infinite number of direct revelation mechanisms with truthful dominant strategies. These other mechanisms are not "Groves-like" -- they are continuous and have budget control. These mechanisms are based on proper scoring rules.

Definition. A risk assessment mechanism is a scoring rule if it can be written in the form $t_i = g(p_i,X)$. (The transfer to each agent depends only on what he reveals and X.)

Definition. A proper scoring rule is a scoring rule $t_i = g(p_i,X)$ for which the expected transfer $E(g(p_i,X))$ is maximized at $p_i = \bar{p}_i$, for each i.

Definition. An assessment mechanism g is normalizable if the range of g is bounded. If $b^- = 0$ and $b^+ = 1$ we will say g is normalized.

Definition. An assessment mechanism g is individually rational if $b^- \geq 0$.

Because we are assuming each i can form at least a crude guess of the probability of X costlessly, i cannot lose when $b^- \geq 0$. The condition of "individual rationality" means i has an incentive to join in an assessment process (except for very pathological cases where he has no incentive to either join or not join). An advantage in working with normalized mechanisms is that the zero floor provides "individual rationality." The unit ceiling for normalized proper scoring rules will prove useful below. (Note that the Groves-like mechanism is normalized.)

Examples. (1) $t_i = g(p_i,X) = 1 - (p - X)^2$

(2) $t_i = g(p_i,X) = |p_i - X| / \sqrt{p_i^2 + (1 - p_i)^2}$

The first is the Brier scoring rule, which has been extensively used in the evaluation of weather forecasters. The second is the spherical rule. It is easy to show that both are proper and normalized. It is well known that there are an infinite number of proper scoring rules (and an infinite subclass can be normalized, maintaining properness).

Theorem 3. Let $g(p_i,X)$ be a continuous, normalized proper scoring rule. Define

$$t_i \equiv g_i'(p_1,\ldots,p_n,X) \equiv g(p_i,X) + 1 - \frac{1}{N-1}\sum_{j\neq i}^{N} g(p_j,X).$$

Then $g' = (t_1,\ldots,t_N)$ is a risk assessment mechanism with truthful dominant strateiges for risk neutral assessors. Further, g' has budget control, is continuous, and provides "individual rationality."

Proof. To show that truthful revelation is the dominant stragety for any i, note that the expected transfer to i is

$$E(t_i) = E(g(p_i,X)) + k(\bar{p},p_1,\ldots,p_{i-1},p_{i+1},\ldots p_N) \tag{3}$$

for some function k. Since k does not depend on p_i, the p_i which maximizes $E(g(p_i,X))$ also maximizes $E(t_i)$. Because g is a proper

scoring rule, we know the maximization of $E(t_i)$ is at $p_i = \bar{p}_i$. As g is continuous so is g_i'. As $g(p_j,X) \leq 1$ (all j), we know $1 \geq \frac{1}{N-1}\sum_{j\neq i}^{N} g(p_j,X)$ and since $g(p_j,X) \geq 0$, we know $g_i' \geq 0$ no matter what are the p_j and X, meeting the condition of "individual rationality." To show that g' has budget control, note that

$$\sum_{i=1}^{N} t_i = \sum_{i=1}^{N} g(p_i,X) + N - \frac{1}{N-1}\sum_{i=1}^{N}\sum_{j\neq i}^{N} g(p_j,X) = N.$$

Q.E.D.

The normalization "wraps around" partial sums in such a way that the total sum comes out constant. The idea is similar to the one used by Walker (1981) to define an incentive compatible mechanism for public goods. The normalization which defines g' is not directly useful to the Groves-like mechanism f because it would introduce all the p_i in the residual function k.

As in Theorem 2 we can discard the assumption of risk neutrality for proper scoring rules in favor of the weaker assumption of monotonicity.

Theorem 4. Let g be a normalized proper scoring rule, and define the mechanism of transfer to i, by

$$t_i = \begin{cases} 1 & \text{with probability } g(p_i,X) \\ 0 & \text{with probability } 1 - g(p_i,X) \end{cases}$$

Then if i's preferences over lotteries are monotonic, i has a truthful dominant strategy.

Proof. Because g is normalized $0 \leq g \leq 1$ and g can be defined as a probability, the proof is parallel to Theorem 2.

While Theorems 2 and 4 generalize 1 and 2, the presence or absence of risk aversion tends to be less important an issue when there are R assessment rounds and i's average transfer converges to the average of his expected transfers.[5]

IV. Link Between the Groves-like Mechanism and the Brier Rule.

So far we have said little about i's beliefs concerning the other p_j at the time i reveals his own p_i. For the Brier rule these beliefs have no impact on i's expected transfer. For mechanism f, they do. But in the special case when i has a diffuse prior over q_i (the consensus of others' revealed probabilities) the two mechanisms are very closely related. Define mechanism f' as identical to the Groves-like mechanism f except all the transfers are doubled.

Theorem 5. When i has a diffuse prior on q_i, the expected transfer to i for mechanism f' is the same as for the Brier rule, for all p_i and $\bar{p}_i$.

Proof. For a diffuse prior on q_i, i's expected transfer $E(t_i)$ is, for any particular p_i and $\bar{p}_i$

$$\int_0^{p_i} E(t_i|p_i \geq q)dq + \int_{p_i}^{1} E(t_i|p_i < q)dq.$$

By the definition of f and f' we know $E(t_i|p_i \geq q_i) = 2\bar{p}_i - 2\bar{p}_i q_i$ and $E(t_i|p_i < q_i) = 2q_i - \bar{p}_i q_i$. Making the substitutions leads to

$$E(t_i) = -p_i^2 + 2p_i\bar{p}_i + 1 - \bar{p}_i.$$

Recall that the Brier rule is defined by $g(p_i,X) = 1 - (p_i - X)^2$ for $i = 1,\ldots,N$. The expected transfer to i under this rule is

$$\bar{p}_i(1 - (p_i - 1)^2) + (1 - \bar{p}_i)(1 - p_i^2) = -p_i^2 + 2p_i\bar{p}_i + 1 - \bar{p}_i.$$

The expectations for the two mechanisms are identical.

Q.E.D.

Thus, from the point of view of an expected value maximizing assessor who has a diffuse prior on the consensus of others' judgments, the Groves-like mechanism f' and the Brier rule are equivalent. However, a diffuse prior on q_i is a rather unlikely benchmark. If i had no information on any of the other p_j, it would seem more natural to assume a diffuse prior on each p_j, which would tend toward a concentrated distribution for q_i by the central limit theorem. Moreover, it would seem plausible for i to assume that others would have some of the same information about the likelihood of X as i himself has. In this case i would adopt a prior for each p_j concentrated around $\bar{p}_i$. A tractible case of a concentrated prior is considered in the next section.

V Sharpness in Expected Transfer.

For the mechanism f' suppose i has a prior on q_i concentrated around $\bar{p}_i$, where i's subjective density function $h(q_i)$ is a triangle peaked on q_i, and defined by

$$h(q) = \begin{cases} 2q/\bar{p} & \text{if } q \leq \bar{p} \\ 2(1 - q)/(1 - \bar{p}) & \text{if } q > \bar{p} \end{cases}$$

where again the subscript i is omitted. For $p \leq \bar{p}$, i's expected transfer is

$$E(t) = \int_0^p 2\bar{p}(1 - q)(2q/\bar{p})dq + \int_p^{\bar{p}} 2(1 - \bar{p})q(2q/\bar{p})dq + \int_{\bar{p}}^1 2(1 - p)q(2(1 - q)/(1 - \bar{p}))dq$$

$$= (2/3)(1 - \bar{p}^2) + 2p^2 - (4/3)p^3/\bar{p}$$

And for $p > \bar{p}$

$$E(t) = \int_0^{\bar{p}} 2\bar{p}(1 - q)(2q/\bar{p})dq + \int_{\bar{p}}^p 2\bar{p}(1 - q)(2(1 - q)/(1 - \bar{p}))dq$$

$$+ \int_p^1 2q(1 - \bar{p})(2(1 - q)/(1 - \bar{p}))dq$$

$$= 2\bar{p}^2 - (4/3)\bar{p}^3 - 2p^2 + (4/3)p^3 + 2/3 + \frac{4\bar{p}}{1 - \bar{p}}(p - p^2 + p^3/3 - \bar{p} + p^2 - p^3/3)$$

So $$\frac{\partial^2}{\partial p^2} E(t|p \leq \bar{p}) = 4 - 8p/\bar{p}$$

and $$\frac{\partial^2}{p^2} E(t|p \geq \bar{p}) = -4(1 - p) - 8\bar{p}(1 - p)/(1 - \bar{p})$$

At the optimal $p = \bar{p}$ the left hand second derivative is -4 and the right hand second derivative is $-4 - 4\bar{p}$. But the second derivative for the Brier rule is -2 everywhere. Thus compared with the Brier rule, mechanism f' (with the triangular prior) has a sharper radius of curvature at the optimal $p = \bar{p}$. From the point of view of the assessor who maximizes his expected transfer, mechanism f' has a greater incentive than the Brier rule for an accurate assessment.

Because there is no budget control for either mechanism f' or the Brier rule, there is no direct budget comparability from the point of view of the principal. But we might expect about half the assessors for mechanism f to get nothing (those i on the "wrong" side of q_1). Further, those i on the "right" side of q_1 receive a

linear function of q_1 scaled from 0 to 1. For the Brier rule everyone receives a quadratic (concave upward) function of their own p_1, scaled from 0 to 1. Thus it is possible that for many cases mechanism f and even the twice-as-expensive f' may cost less than the Brier rule. In Section VII we return to the question of sharpness, but first we consider the problem of identifying the "best" assessor over a series of assessments.

VI. On the Definition of the Best Assessor

This section develops a notion of the best assessor. "Best" becomes definable in a context of a model of the assessor's information. If defined sensibly, the principal would like to identify the best assessor, the second best, and so on. He would also like to uncover the information structure which relates the assessors' revealed estimates to each other. In practice the principal may have to settle for much less.

In a practical problem such as risk assessment for chemical carcinogens there might be five or ten experts assessing the potential carcinogenicity of 30 or 40 chemicals. Each expert makes a probabilistic prediction that a chemical will come out positive in a "definitive" rodent bioassay, costing $750,000 and taking three years to perform. The predictive probabilities p_i might be used to help decide on the degree of precautionary control to take for each chemical during the time of testing. Or, for the part of the problem we are now interested in, they might be used, in conjunction with the outcomes of the bioassays, to identify the better predictive techniques. One assessor might rely heavily on the Ames test, another on a structure-activity model, and so on. The principal would like to identify the predictor with the best "track record," to rely more heavily on this technique or to invest more in its development. But the principal is severely constrained in his source of information. Because the rodent bioassays are so expensive only 20 or 30 are undertaken nationally in a year.

It would be desirable for the principal to estimate a parameter of bias and a parameter of central tendency for each assessor's estimate, along with estimates of the statistical dependencies among the assessors' estimators. As such parameters depend on the underlying maintained model, it would also be desirable to identify this underlying structure of information.

However, there is just not much information in the observation of 30 or 40 yes-no events, which can be accumulated in 3 or 4 years of a testing program -- not enough to support the estimation of many parameters and to identify or validate an underlying model. From a lack of information, the principal may have to settle for a crude notion of "best" and identifiability.

The key idea is this. If each assessor's "technique" (information quality and inference skill) remains constant over a series of assessments, then there is some possibility of detecting assessors with better techniques, even though each assessment is made over a unique and non-repeatable event and there are a limited number of assessments altogether. For a single assessment, the one with the best information and inference skill might easily be unlucky in a single prediction, and others with worse information and skill might be lucky. However, over a series of assessments, it may be possible to identify the best assessor with increasing probability.

Suppose that there are R assessments to be made and each event X_r is unique. Concretely, X_r is a Bernoulli random variable, with $Pr(X_r = 1) = p_r^*$. This probability can vary from round to round and can be anywhere on the closed unit interval. What remains constant is the capability of each assessor. For simplicity assume just two assessors, 1 and 2. In round r, i's information on $\overset{*}{p}_r$ is equivalent to observing a binomial random variable with parameters $(\overset{*}{p}_r, M_i)$. (The two binomial variables are drawn independently.) For this model we have a clean definition of the "best" assessor. If $M_1 > M_2$ than the first assessor has more or better information than the second, and his better capability is preserved over all R rounds. Of course i does not know $\overset{*}{p}_r$ and M_j, in fact he might not even know M_i. The principal does not know M_1 or M_2, and of course not $\overset{*}{p}_r$ either. The principal wants to identify the best assessor after R rounds. If the assessors report truthfully, this model then is also a clear definition of the "best consensus estimates" for the probability of X_r:

$$\alpha p_{1r} + (1 - \alpha)p_{2r} \qquad \text{where } \alpha = M_1/(M_1 + M_2)$$

In other words if the assessors reported truthfully and if α were known, this consensus would be a sufficient statistic for the total $M_1 + M_2$ observations from the Bernoulli process with parameter $\overset{*}{p}_r$. If the principal could estimate α accurately, and if he could elicit truthful reporting, he could do no better than use α and (1 - α) as weights, defining the observable consensus

$$q_r = \hat{\alpha}\, p_{1r} + (1 - \hat{\alpha})p_{2r}$$

An estimate of α can be obtained by least squares, by minimizing over α

$$\sum_{r=1}^{R} (x_r - \alpha p_{1r} - (1 - \alpha)p_{2r})^2$$

The minimization yields

$$\hat{\alpha} = \frac{\Sigma(x_r - p_{2r})(p_{1r} - p_{2r})}{\Sigma(p_{1r} - p_{2r})^2}$$

The principal would also have an operational definition of the best assessor: if after R rounds, $\hat{\alpha} > .5$ assessor 1 is declared best; if $\hat{\alpha} < .5$, assessor 2 is declared best; and if $\hat{\alpha} = .5$ a tie is declared.

There are two major problems with this approach. First, it is difficult to untangle how the assessors might attempt to manipulate this criterion of best. Suppose the principal uses this operational criterion of best and attempts to elicit truthful reporting, round by round, by rewarding the assessors according to repeated use of an assessment mechanism with truthful dominant strategy in a single round. If the assessor declared best is more likely to get promoted, hired as a consultant, or if the principal is likely to invest in the "best" assessor's technique, then each assessor may abandon the goal of maximizing the sum of his transfers in favor of maximizing the probability of being declared the best. If so, assessor 1 will reveal $(p_{11},\ldots,p_{1r})$ to maximize $Pr(\hat{\alpha} > .5)$ and assessor 2 will reveal $(p_{21},\ldots,p_{2r})$ to minimize this probability. Or each assessor may attempt some mix of the two goals. In either case each assessor has an incentive to manipulate this criterion of "best." And if we do not know how the assessors may manipulate the criterion we do not know how good the criterion is.

Second, the appeal of the criterion depends heavily on the structure of information underlying $\bar{p}_{1r}$ and $\bar{p}_{2r}$. For example suppose assessor 1's information on $Pr(X = 1)$ is generated by observing a binomial random variable with parameters (a_r, M_1) where $a_r = (1 + 1/M_1)\overset{*}{p}_r$ if $(1 + 1/M_1)\overset{*}{p}_r \leq 1$ and $a_r = 1$ otherwise. Then even with truthful reporting q_r is no longer an unbiased estimate of $\overset{*}{p}_r$ and the estimator $\hat{\alpha}$ is no longer a least squares estimate for $M_1/(M_1 + M_2)$. The operational definition "Assessor 1 is best if $\hat{\alpha} > .5$" loses much of its appeal.

This last problem is severe, because in actual risk assessments the principal knows little about how information is generated by the assessors. Thus he has no clearly dominant way of defining the "best" assessor. To achieve results we need to assume that the assessors' capabilities remain in some sense constant over a series of assessments, but beyond that we would like to assume as little as possible about what the principal knows of the information structure of the assessors.

A second approach suggested by Roberts (1965) illustrates another way that one assessor's capability could remain consistantly better than the others over a series of assessment rounds. In the second example there are also N assessors. The best assessor i' knows $\overset{*}{p}_r$ for each round r, and none of the others know $\overset{*}{p}_r$. Thus without specifying how much better i' is than j we have specified an information structure in which one assessor has consistently better information than the other, round by round. We have succeeded in specifying practically nothing about N - 1 assessors' structure of information at the expense of specifying a great deal about the

information of one unknown assessor. Again the principal wishes to identify the best assessor after R rounds.

The principal forms N hypotheses θ_i, where θ_i is the hypothesis that i is best. The principal also forms a prior notion of the probability that i is best, $P(\theta_i)$ where $\Sigma P(\theta_i) = 1$. As before i reveals p_{ir} as his prediction of the event $X_r = 1$. Write $s_{ir} = p_{ir}$ if the event happens $(x_t = 1)$; and $s_{ir} = 1 - p_{ir}$ if the event doesn't happen $(x_r = 0)$. Write $x = (x_1,\ldots,x_R)$, the observed record of the R events once they are known. Then if the assessors are revealing truthfully, $\prod_r s_{ir}$ equals $P(x|\phi_i)$, the likelihood of x given i is the best assessor. The posterior probability that i is the best assessor is

$$(4) \qquad P(\theta_i|x) = \frac{P(x|\theta_i)P(\theta_i)}{\sum_{j=1}^{N} P(x|\theta_j)P(\theta_j)} = \frac{P(\theta_i)\prod_r s_{ir}}{\sum_j P(\theta_j)\prod_r s_{jr}}$$

In this situation Roberts suggests that the principal might define the best assessor as the one with the highest posterior probability after the R rounds. Equivalently, the principal could compare posterior odds, pairwise among the assessors, and select the assessor with a ratio always greater or equal to 1, matched against each of the others. If the principal has no information to distinguish the assessors initially, he sets the initial priors equal $(P(\theta_i) = 1/N)$. This case reduces to comparing pairwise the likelihood ratios for assessors. In attempting to identify the best assessor Roberts recommends looking at the posterior odds as "clearly interesting," but as Roberts puts it "a complete analysis of the decision problem does not appear easy." Roberts does not consider the strategic properties of the criterion.

A third approach is for the principal to define "best" on the basis of a transfer mechanism with truthful dominant strategy for a single round. In this approach the principal chooses such a mechanism and announces that each round the assessors will be rewarded according to this mechanism, and at the end of R rounds the assessor with the highest total of transfers will be declared the "best." If R is large enough the law of large numbers takes over, and i maximizes his average transfer by maximizing his expected transfer each round. Similarly he maximizes his probability of having the highest average transfer at the end of R rounds by maximizing his expected transfer each round. By the principal choosing his criterion of the best assessor as the one with highest average transfer after R rounds the principal brings into harmony the assessor's two goals -- maximizing expected transfers round by round and maximizing the probability of being declared the best assessor. By choosing a mechanism with a truth revealing dominant strategy the principal provides an incentive for truthful revelation round by round. We return to this approach in Section IX.

A fourth approach is for the principal to exploit the properties of a strategic mechanism to reveal the best assessor. In this approach the principal chooses a strategic mechanism and announces that each round the assessors will be rewarded according to this transfer and

at the end of R rounds the assessor with the highest average transfer will be declared the best. The idea is that assessors with more information and inference skill will be able to exploit the strategic opportunities of the rule more efficiently than others and will rise to the top more quickly.

If this speculation is correct there may be a tradeoff: worse consensus estimates round by round but more efficient identification of the best assessor after R rounds. ("Worse" and "more efficient" are relative to a mechanism with truthful dominant strategy round by round.) These notions are made more precise for two models of information generation in Section IX and the Appendix. This speculation is investigated for the perimutuel mechanism, which is strategic, but first we need to know something of its strategic properties.

VII. Strategy in the Parimutual Mechanism

This section derives some of the strategic properties of the parimutuel mechanism and compares the mechanism with others having truth revealing, dominant strategies.

Definition. A parimutuel mechanism is a transfer function in the form:

$$(5) \qquad t_i = \begin{cases} \dfrac{p_i c_i}{q} & \text{if } X = 1 \text{ and not all } p_i = 0 \\ \dfrac{(1-p_i)c_i}{1-q} & \text{if } X = 0 \text{ and not all } p_i = 1 \\ c_i & \text{otherwise} \end{cases}$$

where $0 \leq c_i < 1$, all i; $\sum c_i = 1$; and $q = \sum p_i c_i$.

Clearly $\sum t_i = 1$ for all p_i and X; thus the mechanism is budget controlled. The transfer is in the form of Bayes Theorem -- if in terms of Robert's model the principal sets $c_i = P(\theta_i)$, then t_i is the principal's posterior probability that i is the "best" assessor. The mechanism is also the same as a parimutuel betting rule. In a two horse race, where q is the consensus probability of X = 1 (or $(1 - q)/q$ is the odds one sees on the totalizer

for X = 0); c_i is the function of the total betting pool put down by i; p_i is the fraction of i's wager placed on X = 1; and t_i is the fraction of the total betting pool won by i.[6]

When there are repeated applications of the parimutuel mechanism the connection between (5) and (4) is as follows. Set $c_i = P(\theta_i)$ for the first round; for round r set c_i equal to t_i of the preceding round. Then i's transfer in the final round t_{iR} is $P(\theta_i|x)$. We will call this special case of the parimutuel mechanism a Bayesian mechanism. Assessor i's credibility c_i evolves over time and his final transfer is his final credibility. For the Bayesian mechanism i is rewarded t_{iR} and he attempts to maximize $E(t_{ir})$. Unfortunately the analysis of i's strategy is complicated. His global strategy is coupled round by round, and it can be shown that maximization of expected final transfer is not in general consistent with maximizing the expected transfer round by round.[7]

The following three theorems characterize some of the properties of a single round parimutuel mechanism. Expected transfer maximization is assumed as a background condition. Proofs are in the Appendix.

Theorem 6. (Due to James Gerard and Joshua Foreman.) For the paramutuel mechanism with two assessor's, i's reaction function is

$$p_i = \frac{1 - p_j c_j (1 + \sqrt{\bar{P}_i P_j}\,)}{c_i (1 + \sqrt{\bar{P}_i P_j}\,)} \qquad \text{where } j \neq i \text{ and} \qquad \bar{P}_i = \frac{1-\bar{p}_i}{\bar{p}_i} \quad P_j = \frac{1-p_j}{p_j}$$

as long as i's strategy is not 0 or 1 and $0 < c_i < 1$. Further, if $\bar{p}_1 = \bar{p}_2$, then there is a truth revealing Nash equilibrium at $p_1 = \bar{p}_1$ and $p_2 = \bar{p}_2$.

(Calculation of reaction functions for a grid of varying c_i and $\overline{p}_i$ shows that there is usually a unique interior Nash equilibrium, but there are examples of multiple interior equilibria.)

Theorem 7. Choose any arbitrarily small h ($0 < h < .5$). When each assessor's strategy space is restricted to $[h, 1 - h]$, there exists a Nash equilibrium for the parimutuel mechanism.

(The restriction is to remove the two points of discontinuity, at $p_i = 0$, all i; and $p_i = 1$, all i.)

Theorem 8. For the parimutuel mechanism, i reports p_i greater than the consensus of others if $\overline{p}_i > q_i$; reports $p_i < q_i$ if $\overline{p}_i < q_i$; and reports $p_i = q_i$ if $\overline{p}_i = q_i$.

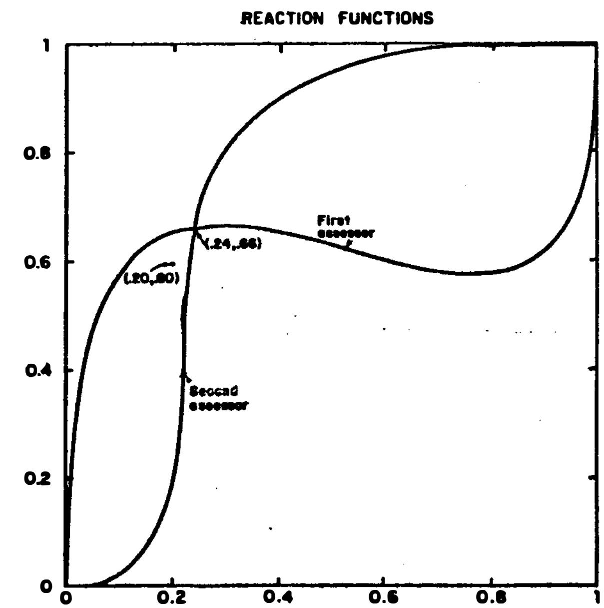
Reaction functions are plotted in Figure 2 for a case of two assessors, where $\overline{p}_1 = .2$, and $\overline{p}_2 = .6$ and $c_1 = .4$. As can be seen for this case the Nash equilibrium is not far from the point of truthful revelation, its displacement according to Theorem 8.[8] The displacement is not always small, however, as shown by the corollary to follow.

Definition. An assessor follows a zero-one knife edge strategy if he reports 0 when $\overline{p}_i < q_i$, 1 when $\overline{p}_i > q_i$, and $\overline{p}_i$ when $\overline{p}_i = q_i$.

Corollary. For the parimutuel mechanism i's strategy converges to a zero-one knife edge strategy as $c_i \to 0$.

As c_i declines, convergence to a knife edge, zero-one strategy for i is rapid. This can be seen in Figures 3 and 4 where regions of the zero-one strategy and the bluntness of the knife edge are shown for $c_i = .1$ and $c_i = .01$. In racetrack betting, N can be 10,000 or more and each c_i is correspondingly small. Thus an expected value

Figure 2



(.20, .60) is the point of best judgments $(\overline{p}_1, \overline{p}_2)$

(.24, .66) is the Nash equilibrium of revealed (p_1, p_2)

$c_1 = .4$ $c_2 = .6$

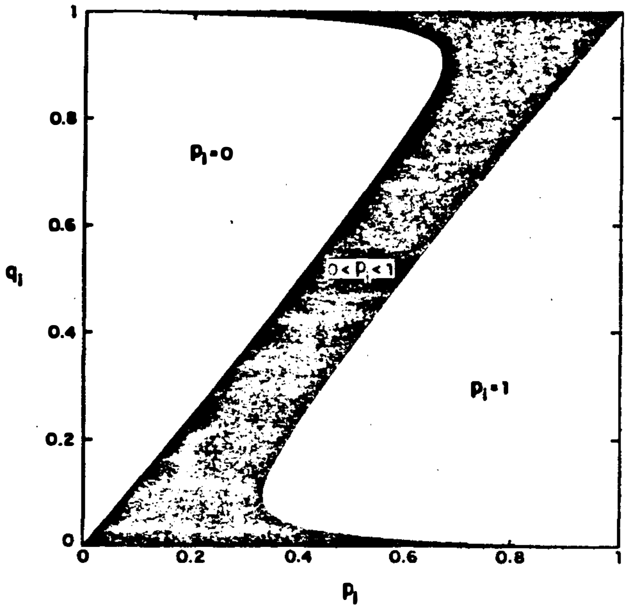


Figure 3

Regions of Zero-One Strategy

$c_i = .1$

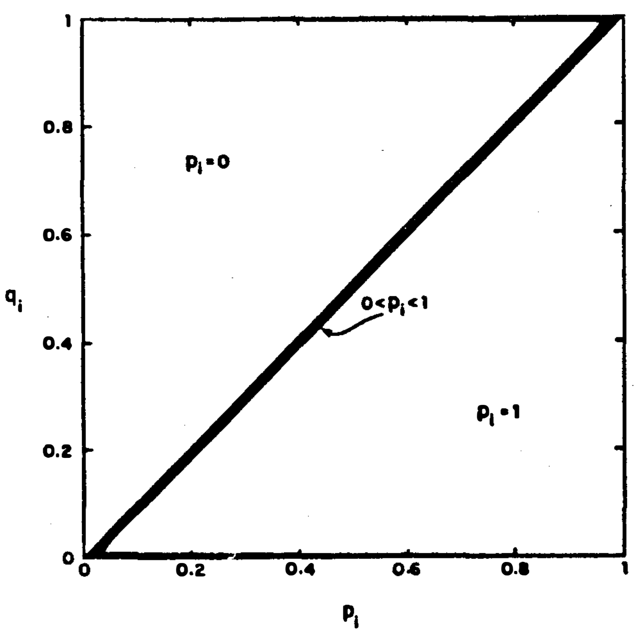


Figure 4

Regions of Zero-One Strategy

$c_i = .01$

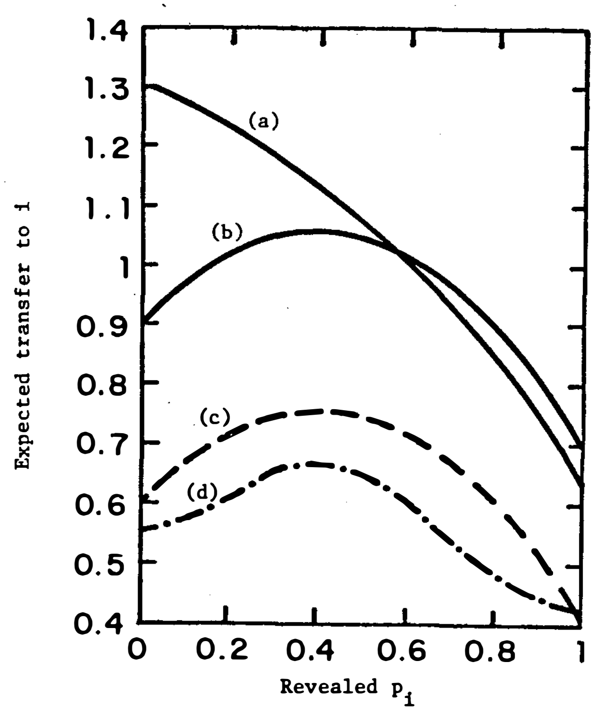
maximizing bettor at a parimutuel track has a strategy which is virtually knife edge, zero-one. And in fact parimutuel betting commonly follows this pattern (all of one's wager on one horse, for each race).

To put the expected transfer for the parimutuel mechanism on the same figure as those of mechanism f' and the Brier, scale the parimutuel mechanism by multiplying its transfers by N. Figure 5 makes the comparison when $\overline{p}_i = .4$, $c_i = .1$ and $p_j = .6$. (As noted in Figure 4 the information sets differ for the three mechanisms.) Clearly, at the point of revelation (0 for the parimutuel, and .4 for the others) the slope of the expected transfer function is steepest for the parimutuel mechanism (the others being zero). In a sense this suggests that the parimutuel mechanism provides the strongest incentives for i to sharpen his judgment of $\overline{p}_i$. Actually, however, for the parimutuel mechanism i's incentive is to sharpen his judgment of $\overline{p}_i$ "just enough" to decide which side of the knife edge to choose (if c_i is sufficiently small to elicit a zero-one strategy, as it is in this case).

Sharpness for the Brier mechanism and mechanism f' is more directly comparable. For mechanisms with truthful strategies (such as these) local convexity at $\overline{p}_i$ defines a notion of sharpness for i. As shown in Figure 5 and derived analytically in Section V, the radius of curvature is smaller for mechanism f'. This means that being a little off in one's judgment of $\overline{p}_i$ makes more difference in the mechanism f' compared with the Brier.

Figure 5

Expected Transfer as a Function of
Revealed Probability



(a) Parimutuel (budget controlled to N)
(b) Brier (adjusted for budget control at N)
(c) Brier
(d) Mechanism f'

For the three mechanisms $\overline{p}_i = .4$, $\overline{p}_j = .6$ $(j \neq i)$; $N = 10$; and $c_i = 1$.

At the time i reveals p_i, his information includes q_i for (a) by i has no direct information on p_j for (d). For (d) i has a triangular prior on q_i, peaked at $\overline{p}_i$.

To further compare the (scaled) parimutuel mechanism, with mechanism f, define c_i to be the same for both mechanisms and define c_i to be the same for both mechanisms and defined so that c_i is small for large N. Then for large N the gap between q and q_i is small. Neglecting the gap between q and q_i and the exceptional cases, we can put the two mechanisms side by side:

	Scaled parimutuel	Mechanism f	
$t_i =$	$\frac{1}{q}$	$1 - q$	if $X = 1$ and $p_i > q$
	$\frac{1}{1-q}$	q	if $X = 0$ and $p_i < q$

The two mechanisms have the same orientation, but for the same budget bound N there is a far greater opportunity for a big win in the parimutuel mechanism. For the parimutuel mechanism i has a chance at a large fraction of the principal's budget ceilign N. For mechanism f, i can never do better than 1. However, mechanism f is cheaper for the principal than the scaled parimutuel mechanism because the former never obtains its budget bound while the latter always does.

VIII. Rational Expectations

So far we have specified little about how one i might gain information about $\overline{p}_j$ through the revealed p_j. For mechanism f we have noted (see note 3) that to establish an incentive to follow a truthful dominant strategy the principal needs to keep i at least somewhat uncertain of p_j at the time of i's own revelation. For the mechanisms built on proper scoring rules there is no such strategic consideration on the part of the principal. Nonetheless, if the principal wants to identify the best assessors, he may not want some less informed i to improve his $\overline{p}_i$ by observing a more informed j's p_j which reveals information on $\overline{p}_j$.

In some ways the parimutuel mechanism is the most interesting for developing models of rational expectations. The consensus estimate q -- equivalent to the payoff odds on the totalizer of a racetrack -- functions like a price signal in a market. Individuals in a market do not see others' individual actions, but they see an aggregative summary in market prices. Similarly, in the parimutuel mechanism individuals do not see others' revealed assessments individually, but they do see and react to the aggregative q. In parimutuel betting the consensus q (or the equivalent) is prominantly displayed on a totalizer and a limited form of recontracting is allowed while q shifts in response to betting.[9] And in the same way than an individual may gain information from an aggregative price, i can modify his judgment $\overline{p}_i$ in response to the q he sees on the totalizer.

Developing a rational expectations model requires specifying the process of tatonnement or recontracting and the flows of information among assessors. While doing so would be beyond the scope of this paper, it is worth suggesting that Theorem 8 provides favorable conditions for such models. A less informed bettor j, if he knows he is less informed, is more likely to revise $\overline{p}_j$ farther toward q than a well informed assessor. (Theorem 8 holds for the full consensus q as well as the others' consensus q_i.) The more an assessor approaches q the less effect he has on it. A more informed i presumably is less influenced by q in forming and modifying his $\overline{p}_i$. And by holding to his better informed $\overline{p}_i$ he pulls q in direction of $\overline{p}_i$. (If he believes the true probability of $X = 1$ is greater than the consensus q, he bets on X, tending to increase q, and the reverse if $\overline{p}_i < q$.) While there are other factors to consider, it appears there is a tendency in racetrack betting for the uninformed to follow the lead of the informed, with the resulting q weighted toward the $\overline{p}_j$ of the informed.[10]

IX. Identifying the Best Assessor

In this section a Monte Carlo simulation is used to estimate the probability of identifying the best assessor under three mechanisms -- the Brier, the parimutuel, and the Groves-like — for a simple model of the assessors' information. The model is chosen to reflect the practical problem of risk assessment for potential carcinogens. For this problem there are a small number of assessors, a limited number of assessments, and the possibility of bias as well as variance in each assessor's best judgment. The model of the assessor's information, simple as it is, appears to preclude direct analytical treatment.

As in Section VI define the best assessor i' in terms of an underlying model of information generation, where i' has more information on $\overset{*}{p}_r$ than any other i. Write i's average transfer after R rounds $\bar{t}_i = \frac{1}{R}\sum_{r=1}^{R} t_{ir}$ where t_{ir} is the transfer to i in round r, and formalize a notion of identifiability as follows:

Definition. The best i' is identifiable in R rounds if $Pr(\bar{t}_i > \bar{t}_j,\ \text{all}\ j \neq i)$ is maximized by $i = i'$.

Definition. The probability of identifying i' (the best i) is $Pr(\bar{t}_{i'} > \bar{t}_j,\ \text{all}\ j \neq i')$.

Before we compare identifiability for the three mechanisms, we briefly ask what is the role of strategic revelation for identifiability in the parimutuel mechanism. For an analytically tractible example, we find that strategic revelation is essential to identifiability.

In the example, there are many assessors (N is large as in race track betting) and c_i is reset to 1/N each round. A poorly informed assessor j forms a judgment of $\overset{*}{p}_r$ by observing two observations on a Bernoulli process with parameter $\overset{*}{p}_r$; the best i' forms his judgment of $\overset{*}{p}_r$ by observing an infinite number from the same process (i' knows $\overset{*}{p}_r$).

For this extreme example of disparity of information among the N - 1 poorly informed j and the best informed i', we find that if the assessors, contrary to their incentives under the parimutuel mechanism, somehow reported truthfully, the best i' would not be identifiable, for any R. (Note that this version of the parimutuel mechanism is not the Bayesian mechanism, where the c_i evolve over time. However, if the assessors respond to their incentives to report strategically, i' becomes identifiable, and the probability of identifying i' goes to 1 as R goes to infinity. Details and derivations are in the Appendix.

The underlying reason for nonidentifiability in the parimutuel mechanism is that with truthful reporting q converges to $\overset{*}{p}$, and when $q = \overset{*}{p}$ the expected transfer becomes 1 no matter the reported p_i (see definition (5) and scale by N). Similarly, for mechanism f if q_i converges to $\overset{*}{p}$ there is no identifiability as $R \to \infty$ (note that if $q_i = \overset{*}{p}$, i's expected transfer is $\overset{*}{p} - \overset{*}{p}{}^2$ no matter what the reported p_i). In contrast, if $q_i = \overset{*}{p}$ then i' can still be identified for the Brier mechanism. A specific example is given in the Appendix.

However, when $q_i \neq \overset{*}{p}$ the situation is quite different. The case where we expect $q_i \neq \overset{*}{p}$ can arise as follows. A chemical is either a carcinogen (X = 1) or not (X = 0); and $\overset{*}{p}$ is also either 0 or 1. Assessors develop information on the possible carcinogenicity of the chemical but their information is imperfect and hence $0 < \overline{p}_i < 1$ (only perfect information would lead to $\overline{p}_i = 0$ or $\overline{p}_i = 1$) and thus each consensus q_i is also interior. For the simulation model, we assume there are R chemicals to assess and each assessor i has information which suggests either $X_r = 1$ or $X_r = 0$. The imperfection of i's information is characterized by false positive and false negative probabilities for i:

$$FP_i = \Pr(\text{i's information suggests } X_r = 1 | X_r = 0)$$
$$FN_i = \Pr(\text{i's information suggests } X_r = 0 | X_r = 1)$$

Suppose there is one $i = i'$ with clearly superior information: for this i' $FP_{i'} < FP_j$ and $FN_{i'} < FN_j$ for $j \neq i'$.

In this model there is an underlying prevalence rate for carcinogens and each assessor i uses this rate to calculate $\overline{p}_i$ according to the Bayes Theorem. The principal does not know the FN_i, FP_i or the prevalence rate, but declares the assessor with the highest average transfer after R rounds to be the best assessor.

Under these conditions a Monte Carlo simulation can be used to estimate the probability of identifying i'. For this model of information the qualitative results appear similar for various values of the prevalence rate and error rates, and the following numbers are illustrative.

In the testing programs of the National Cancer Institute and the National Toxicology Program, the rate of positives for chemicals in rodent bioassays is about 30 percent. (This prevalence rate is higher than what is suspected as the prevalence of carcinogens over all chemicals, because chemicals selected for testing are among the more suspicious.) Suppose for the best assessor the false positive probability is 10 percent and false negative probability is 15 percent; and for the others the false positive rate is 15 percent and false negative probability is 20 percent. By supposition there is some difference, but not a great difference, between the best assessor and the others. Finally suppose that there are thirty chemicals to assess and ten assessors. These numbers appear plausible for the problem of assessing potential carcinogens.

For the Brier mechanism the probability of correctly identifying i' as the best assessor after 30 rounds is estimated by the simulation to be .31. But if the principal chooses the parimutuel mechanism the probability of correctly identifying the best assessor after 30 rounds is .59. The difference tends to confirm our earlier intuition that a strategic mechanism may reveal the best assessor more efficiently than a mechanism with truthful dominant strategies. But as a final surprise the Groves-like mechanism f does better than either. If the principal chooses this mechanism the probability of identifying the best assessor after 30 rounds is .64. Thus, <u>for this model of the assessors' information</u>, the principal need not give up truthful reporting for greater efficiency in identifying the best assessor.

These differences in identifiability among the three mechanisms are remarkable and suggest further research. It would be interesting to investigate further the conditions of identifiability and their impact on individual incentives. It would also be interesting to investigate a multiround version of mechanism f, where after several initial rounds c_i in round r is set equal to the sum of his previous transfers divided by the sum of everyone's previous transfers. Then the c_i would evolve through time as in theBayesian mechanism. Because i has truthful dominant strategy no matter the value of q_i, changing c_i has no effect on i's strategy (unlike the Bayesian mechanism). If there is identifiability, by allowing i's credibility to evolve over time, better consensus estimates might be obtained -- because the credibilities of the better informed assessors would tend to increase, and decrease for the more poorly informed assessors.

Another line of research is suggested by noting that under mechanism f, for any fixed q_i assessor i gets the same transfer for all p_i reported greater than q_i, and the same transfer for all p_i reported less than q_i. The insensitivity of i's transfer to p_i (within the appropriate ranges) suggests an elicitation technique which might have practical value.

In many cases of actual risk assessment incentive compatibility is not the apparent issue.[11] In many cases the stated problem is not the willingness of assessors to report truthfully their best judgments, but to make them in a quantified form. Experts often express reluctance in reporting numerical probability assessments because they feel that two decimal places (or even one) overstate and misrepresent the

"accuracy" they attach to their judgments. A Bayesian would say that stating a subjective probability to two decimal places does not imply firm conviction of its "accuracy" -- nonetheless the reluctance on the part of the experts is real.

A possible resolution of this problem is to elicit judgments in the following way. A trial consensus estimate is put on a board for all to see. Each assessor is asked if he believes that the true likelihood of the event being assessed is higher, lower, or equal to that trial value. The trial value is shifted up or down until a median estimate is obtained.[12] From this benchmark, as the second step numerical estimates are elicited, to whatever degree of accuracy the assessors feel comfortable with. In this elicitation technique, the truthful dominant strategy of mechanism f remains but the incentive to follow it is weakened. The Appendix and further simulation tentatively suggest that even at this first stage, where only trichotomies are elicited, useful consensus estimates can be generated and i' can be identified under all three mechanisms. These and other speculations could be sharpened in terms of specific models of information generation and by direct experiments.

X. Summary and Conclusion

In this paper we have developed a model of risk assessment based on a structural analogy with a model of public goods demand revelation. This analogy provides a new and more formal way of viewing the problem of risk assessment. With the framework we have introduced a new mechanism with truthful dominant strategy for revelation of best judgmental probabilities. This mechanism, analogous to the Groves mechanism for public goods, is then compared with proper scoring rules (especially the Brier rule) and the parimutuel mechanism. There are several surprises in the comparisons, not the least of which is that the new mechanism appears to perform relatively well compared with other existing mechanisms.

The paper adopts an approach to probability in between those of a strict Bayesian and a strict frequentist. For a strict Bayesian every event is unique and probabilities are subjective degrees of belief which can be validated only in a limited way. Validation of personal probabilities is internal in the sense that it refers to consistency with Savage's (or someone else's) axiom system defining "rationality" and does not depend on whether or not the predicted event actually occurred. Such validation is limited because for a given event widely differing personal probabilities are consistent with the axiom system.

For a strict frequentist, the notion of probability is undefinable when we are dealing with unique events, as we often are in risk assessment. An event must be held constant and repeatable many times for its probability to be definable. In the paper, we let the

predicted events be unique and varying one from another but hold constant something else, the predictive capability of each assessor. With each assessor's "technique" held constant over a series of predictions of unique and varying events identification of assessors with better techniques may be possible. In this way an assessor's predictive capability can be validated beyond its consistency with an axiom system to reference with the outcomes of predicted events and yet this validation does not require repeatable events.

APPENDIX

Proof of Theorem 6. We assume that i observes p_j or that recontracting is possible. For convenience write p_j as q_i and c_j as $(1-c_i)$ and omit the subscript i. Let $k = q(1-c)/c$ and $m = (1-q)(1-c)/c$ and rewrite condition (5) as

$$t = \begin{cases} (1 + kp^{-1})^{-1} & \text{if } X = 1 \\ (1 + m(1-p)^{-1})^{-1} & \text{if } X = 0 \end{cases}$$

(As we are computing the reaction function for interior p and q, we do not bother with the definition of t for the two singular points $(p,q,X) = (0,0,1)$ or $(1,1,0)$.) For each fixed c,q, and $\bar{p}$ the expected transfer is a function of p; write the expected transfer $T(p)$ and

$$T(p) = \bar{p}(1 + kp^{-1})^{-1} + (1 - \bar{p})(1 + m(1 - p)^{-1})^{-1}$$

By direct calculation

$$T' = \frac{\bar{p}k}{(p+k)^2} - \frac{(1-\bar{p})m}{(1-p+m)^2} \qquad p \neq -k \text{ or } m+1$$

$$T'' = \frac{2\bar{p}k}{(p+k)^3} + \frac{2(1-\bar{p})m}{(1-p+m)^3} \qquad p \neq k \text{ or } m+1.$$

By definition $k > 0$ and $m > 0$, as by assumption $0 < c < 1$ and $0 < q < 1$. So $T''(p) < 0$ for $p \in (-k, m+1)$. Because T is everywhere concave over $(-k, m+1)$, $T' = 0$ for only one p in this interval, and this p defines a maximum. Setting $T' = 0$ yields the first order condition

$$(7) \qquad \frac{1-pc-q(1-c)}{pc+q(1-c)} = \sqrt{\bar{P}Q} \qquad \text{where} \qquad Q = \frac{1-q}{q}$$

Because the LHS is always positive for p, q and c in the open unit interval, only the positive root on the RHS applies. Solving for p,

$$(8) \qquad p = \frac{1 - q(1-c)(1+\sqrt{\bar{P}Q})}{c(1+\sqrt{\bar{P}Q})}$$

When $\bar{p}_1 = \bar{p}_2$, substitution of $p_1 = p_2 = \bar{p}_1$ into each assessor's reaction function and check of the four possible boundary cases confirms that $p_1 = \bar{p}_1$ and $p_2 = \bar{p}_2$ is a Nash equilibrium.

Q.E.D.

(I wish to thank James Gerard and Joshua Foreman for the original proof of Theorem 6.)

Proof of Theorem 7. Write $q_i = \sum_{j \neq i}^{N} p_j c_j/(1-c_i)$, the consensus of everyone but i. Substitution of q_i into (5) shows that as far as i is concerned he is playing against a single (aggregate) player q_i with credibility $(1-c_i)$. Omit the subscript i and define the RHS of condition (8) as S(q). S(q) is i's reaction function if his reaction is inside (h, h - h). Define

$$R(q) = \begin{cases} h & \text{if } S(q) \leq h \\ 1-h & \text{if } S(q) \geq 1-h \\ S(q) & \text{otherwise} \end{cases}$$

Since S(q) in continuous in q, R(q) is also continuous. If T is maximized over $p \varepsilon (-k, m + 1)$ by $p^o < h$ (by $p^o > 1 - h$), then T is maximized by h (by 1 - h) over $p \varepsilon [h, 1 - h]$ because T is everywhere concave over the bigger interval.

Thus R(q) is i's reaction function for the strategy space [h, 1 - h], including boundary reactions. Return to the original notation and write R(q) as $R_i(p_1,\ldots,p_{i-1},p_{i+1},\ldots,p_N)$ for i's reaction function. Similarly, construct reaction functions R_j for each of the other assessors and form the composite function $R = (R_1,R_2,\ldots R_N)$. Then R is also continuous and maps $[h, 1 - h]^N$ into itself. As $[h, 1 - h]^N$ is convex and compact, Brouwer's fixed point theorem applies. This means there exists at least one point $(p_1,p_2,\ldots,p_N)$ which leads to no change under application of the composite reaction function. By definition this is a Nash equilibrium.

Proof of Theorem 8. Say $\overline{P}_i < q_i$. Then $\overline{P}_i > Q_i$, and from (7), omitting the subscript i,

$$\frac{1 - pc - q(1 - c)}{pc + q(1 - c)} = \sqrt{\overline{PQ}} = \frac{1 - q}{q} - \frac{\varepsilon}{q} \quad \text{for some } \varepsilon > 0; \text{ so}$$

$$p(1 - \varepsilon) = q + \varepsilon q(1 - c)/c$$

$$p > q + \varepsilon q(1 - c)/c > q \qquad (8)$$

When $\overline{p}_1 > q_1$, $\varepsilon < 0$ and the inequalities in (8) are reversed. Then $\overline{p}_1 = q_1$, $\varepsilon = 0$.

Q.E.D.

Proof of Corollary. Observe in condition (8) that for fixed $p_i > q_i$ and hence fixed $\varepsilon > 0$, p_i goes to its upper boundary as $c_i \to 0$. Similarly if $p_i < q_i$, p_i goes to its lower boundary as $c_i \to 0$.

Q.E.D.

Nonidentifiability in the Parimutuel Mechanism

In the first model described in Section IX i' (the best i) knows $\overset{*}{p}_r$ ($\overset{*}{p}_r$ can change from round to round); each other j observes a binomial random variable with parameters ($\overset{*}{p}_r$, 2), where each of the N - 1 observations for the N - 1 j's is drawn independently. Assessor j's best judgment of $\overset{*}{p}_r$ is:

$$\bar{p}_{jr} = \begin{cases} 1 & \text{w.p. } \overset{*}{p}{}_r^2 & \text{(j observes (1,1))} \\ .5 & \text{w.p. } 2\overset{*}{p}_r(1 - \overset{*}{p}_r) & \text{(j observes (1,0) or (0,1))} \\ 0 & \text{w.p. } (1 - \overset{*}{p}_r)^2 & \text{(j observes (0,0))} \end{cases}$$

The parimutuel transfer to i in round r, t_{ir}, is defined as in (5) except all transfers are multiplied by N to achieve budget control at level N.

Suppose, contrary to their strategic incentives each assessor reports truthfully. By the Bernoulli law of large numbers, with N large, the fraction of N reporting 1 approximates (omitting subscript r) $\overset{*}{p}{}^2$ (N - 1)/N; the fraction reporting .5 approximates $2\overset{*}{p}(1 - \overset{*}{p})(N - 1)/N$; and the fraction reporting 0 approximates $(1 - \overset{*}{p})^2(N - 1)/N$. The fraction reporting $\overset{*}{p}$ is of course exactly 1/N. With large N and $C_i = 1/N$ (all i) the consensus $q = \frac{1}{N}\Sigma p_i$ converges to

$$\frac{(1)\overset{*}{p}{}^2(N - 1) + (.5)2\overset{*}{p}(1 - \overset{*}{p})(N - 1) + (0)(1 - \overset{*}{p}{}^2)(N - 1) + \overset{*}{p}}{N} = \overset{*}{p}$$

By the definition of the parimutuel mechanism (5) it is clear that when X = 0 an expected number of $(1 - \overset{*}{p})^2(N - 1)$ bad assessors

will have higher transfers than i', and when X = 1, and expected number of $(2\overset{*}{p} - \overset{*}{p}{}^2)(N - 1)$ bad assessors will have higher transfers.

However, as R increases $\bar{t}_{i'R}$ converges to $\frac{1}{R}\sum_{r=1}^{R} Et_{i'r}$.

With truthful reporting q_r converges to $\overset{*}{p}_r$, and $E(t_{i'r})$ converges to 1. (This last result follows from (5) in Section VIII (scaled by N). As the budget is controlled to N and the other j are symmetric to each other, they split the remaining budget equally in expectation, each j receiving an expected transfer of (N - 1)/(N - 1) = 1 per round. In consequence i' has the same expected transfer as each j. With this equality in expectation i' remains unidentifiable for any number of rounds, even though his information is clearly superior to the others.

Next suppose that each assessor exploits the mechanism as best he can. Each maximizes his expected transfer each round. For $0 < \overset{*}{p}_r < .5$ the corresponding Nash equilibrium is easily computed to have a consensus estimate (omitting the subscript r) $q = \overset{*}{p}(2 - \overset{*}{p})(N - 1)/N$. Because $q > \overset{*}{p}$ (for large N and $\overset{*}{p} < .5$), long shots are overbet, with a discontinuity in q, as a function of $\overset{*}{p}$, at $\overset{*}{p} = .5$. The discontinuity is an artifact of the model. If j made more than two observations on the Bernoulli process there would be more jumps but smaller ones. It can be shown that as the number of j's observations increases the gap between the Nash equilibrium q and $\overset{*}{p}$ goes to zero. Also it seems clear that when X is more than dichotomous (more than two horses in a race) the amount of overbetting on long shots decreases.

An empirical study on racetrack betting by Hoerl and Fallin (1974) indicates that longshots are in fact overbet, but the overbetting is small with q (odds on the totalizer) an excellent predictor of the order of finish.

Returning to the Nash equilibrium of the example, we find i' reports $p_{i'} = 0$, because $\overline{p}_{i'} = \overset{*}{p} < q$. The expected transfer for i' is

$$\frac{1}{1-q}(1-\overset{*}{p}) = \frac{1}{(1-\overset{*}{p}) + \frac{1}{N}\left(\frac{2\overset{*}{p} - \overset{*}{p}^2}{1-\overset{*}{p}}\right)}$$

With probability $\overset{*}{p}^2 + 2\overset{*}{p}(1-\overset{*}{p}) = 2\overset{*}{p} - \overset{*}{p}^2$, badly informed assessor j will have $\overline{p}_j > q$ and will report $p_j = 1$. When j reports 1 his expected transfer is $\overset{*}{p}/q$. With probability $(1-\overset{*}{p})^2$, $\overline{p}_j$ is less than q and j reports $p_j = 0$. When j reports 0 his expected transfer is $(1-\overset{*}{p})/(1-q)$. So j's expectation over the two possibilities is

$$(2\overset{*}{p} - \overset{*}{p}^2)\overset{*}{p}/q + (1-\overset{*}{p})^2(1-\overset{*}{p})/(1-q) = \frac{N}{N-1}\overset{*}{p} + \frac{1-\overset{*}{p}}{1 + \frac{(2\overset{*}{p} - \overset{*}{p}^2)}{N(1-\overset{*}{p})^2}}$$

For large N, $E(t_{i'r})$ converges to $1/(1-\overset{*}{p})$ and $E(t_{jr})$ converges to 1. As $\overset{*}{p} < .5$, $E(t_{i'r}) > E(t_{jr})$ for large N. Hence i' is identifiable with strategic revelation.

Having q close but not equal to $\overset{*}{p}$ has advantages for a racetrack operator who is more interested in filling the stands and betting lines than a "good consensus estimate" or "identifiability of i". (The principal is more interested in the third goal, strong individual incentives for the assessors to take part in the assessment exercise.) With q close to $\overset{*}{p}$, there is protection of the Saturday duffers who can bet randomly and have almost fair bets. (With strategic revelation, if $q = \overset{*}{p}$, random betting does just as well as any other, see (5).) But with q not equal but close to $\overset{*}{p}$, there is an edge in favor of the insiders. The duffers have the thrill of betting and almost holding their own against the insiders even with worthless information. The insiders systematically do better, as they exploit the difference between q and $\overset{*}{p}$, although they probably succumb to the house percentage eventually. And for the parimutuel mechanism there is the prospect of a big win. The parimutuel is designed, perhaps inadvertantly, for strong gambling incentives among the bettors.

Expectations for i' and j can also be calculated for the Brier and Groves-like mechanisms for this same model of information generation. Truthful reporting leads to an expected transfer to i' in the Brier mechanism of $\overset{*}{p}^2 - \overset{*}{p} + 1$ and an expected transfer to j of $(1.5)\overset{*}{p}^2 - (1.5)\overset{*}{p} + 1$. Because former is greater than the latter, there is identifiability in the Brier mechanism for this model of information. For mechanism f the expected transfer to i' and to j is the same, $\overset{*}{p} - \overset{*}{p}^2$, so there is no identifiability for this model of information. For this model of information the percent difference between the expectations of i' and j can easily be shown to be greater under strategic reporting in the parimutuel mechanism than under truthful reporting in the Brier. This suggests that the parimutuel mechanism may be more efficient in identifying the best assessor in a limited number of rounds. The conjecture is confirmed for a different model of information generation in Section IX, where the small sample properties (for small N as well as small R) and statistical dependencies among the t_i within rounds are dealt with by Monte Carlo simulation.

*This research was supported by the National Science Foundation and by the Mellon Foundation. I would like to thank Richard McKelvey, John Ferejohn, Ed green, Joshua Foreman, and Jim Gerard for many helpful comments, and especially to thank Gib Bogle for programming the Monte Carlo simulation.

Footnotes

[1] For a definition of individual rationality in public goods demand revelation see Green and Laffont (1979) p. 121.

[2] Later, for the parimutuel mechanism we will need to restrict i's strategy space. For the public goods model i's strategy space for w_i is the real line. The definition could easily be generalized for more than a dichotomous X.

[3] As another parallel note that if i knew for sure that $q_i > \overline{p}_i$, revelation biased down would be just as good as truthful revelation. For truthful revelation to have an expected value strictly greater than that for any other strategy, i must attach some probability weight to both $q > \overline{p}$ and $q < \overline{p}$, weight to both cases 1 and 2. In the same way even though truthful revelation is the unique dominant strategy for the Groves mechanism, overstatement of willingness to pay is just as good if i is certain that $\sum_{j \neq i}^{N} w_j > v_i$ and understatement if $\sum_{j \neq i}^{N} w_j < v_i$. For truthful revelation to be strictly better each individual must attach some probability weight to his pivot being both above and below his true valuation.

[4] Monotonicity is defined as follows. Let A be the lottery where i wins a with probability p and wins b with probability $1 - p$; and let A' be the lottery where i wins a w.p. p' and b w.p. $1 - p'$. Then if i prefers a to b, i prefers lottery A to A' $\Longleftrightarrow$ $p > p'$.

[5] For an application of Theorem 4 let g be the Brier mechanism. Then i wins 1 w.p. $-p_i^2 + 2p_i\overline{p}_i + 1 - \overline{p}_i$ and 0 w.p. $p_i^2 - wp_i\overline{p}_i + \overline{p}_i$. This mechanism is similar to, but sharper than the one proposed by Grether. In Grether's mechanism, after p_i is elicited two random variables Y and Z are drawn, each uniform over [0,1]. If $Y < p_i$, i wins 1 if $X = 1$; if $Y \geq p_i$ i wins 1 if $Z > Y$; otherwise i wins nothing. Thus i wins 1 w.p. $(-p_i^2 + 2p_i\overline{p}_i + 1)/2$ and 0 otherwise. Taking as a measure of sharpness the second derivative w.r. + p_i, the Brier mechanism is twice as sharp as Grether's. If the stakes were doubled in Grether's mechanism there would be equal sharpness in the expected transfer. But then the expected transfer in Grether's mechanism would be larger than for Brier's, and thus the principal would have to spend a higher expected budget to achieve the same sharpness.

[6] Joshua Foreman pointed out to me the close connection between Bayes rule and the parimutuel mechanism. In Page (1977) the transfer mechanism (5) was called a Bayesian game. In application of the parimutuel betting rule the betting pool is first reduced by a fraction equal to the "house percentage," before it is divided among the winners according to (5).

[7] Winkler (1969) has shown, however, that for $N = 2$ when i is rewarded $\log(t_{iR}/t_{jR})$ is has a truth revealing, expected value maximizing strategy and the mechanism boils down to a logarithmic proper scoring rule. This reward structure lacks individual rationality, as $b^- = -\infty$.

[8] For the parimutuel mechanism with two assessors in a Nash equilibrium, if it is interior, the consensus odds ratio $(1 - q)/q$ is the geometric mean of i's revealed odds $(1 - p_i)/p_i$ and j's true odds $(1 - \bar{p}_j)/\bar{p}_j$; it is also the geometric mean of j's revealed odds and i's true odds. Moreover, the revealed consensus lies between $\bar{p}_1$ and $\bar{p}_2$. This result follows immediately from the first order condition (7) of the Appendix.

[9] It is possible to define a parimutuel mechanism in which i is ignorant of q at the time he reveals p_i. In such a case i would base his strategy on his Bayesian priors of the other p_j. However, existing parimutuel mechanisms reveal q during the betting period.

[10] For a formal model of polls and voting behavior where the uninformed follow the lead of the informed see McKelvey and Ordeshook (draft).

[11] It is widely believed that the presense or absence of monetary transfers strongly affects behavior in experimental settings. However, in two papers dealing with judgmental probabilities Grether (1979, 1980) found the role of monetary incentives "surprisingly weak."

[12] An alternative to the median is as follows. The trial value is shifted until the number reporting "higher" equals the trial value. Then this benchmark value is the same as what would be obtained as a Nash equilibrium for the parimutuel mechanism with small c_i and knife edge, zero-one strategies. In the simulation the median benchmark yielded slightly higher probabilities of identification compared with this alternative. Gib Bogle suggested the median benchmark.

References

Brier, G. (1950). "Verification of Forecasts Expressed in Terms of Probability." Monthly Weather Review. Vol 78, No. 1, Jan. 1950.

Green, J. and J.J. Laffont (1979) Incentives in Public Decision-Making. North-Holland Pub. Co., New York.

Grether, David M. (1980). "Bayes Rule as a Descriptive Model: The Representativeness Heuristic." Quarterly Journal of Economics, Vol. 95, pp. 537-557.

Grether, David M., and Charles R. Plott (1979). "Economic Theory of Choice and the Preference Reversal Phenomenon." The American Economic Review, vol. 69, no. 4, pp. 623-638.

Groves, T. and M. Loeb (1975). "Incentives and Public Inputs," J. of Public Economics, 4.

Hoerl, A. and H. Fallin (1974), "Reliability of Subjective Evaluations in a High Incentive Situation." J.R. Statist. Soc. A. Vol 137, Part 2.

Lichtenstein, S., B. Fischhoff, and L. Phillips (1981). "Calibration of Probabilities: The State of the Art to 1980," in Judgment under Uncertainty: Heuristics and Biases, D. Kahneman and A. Tversky (eds.), Cambridge Univ. Press, New York.

McKelvey, R. and P. Ordeshook (draft). "Information in Two Candidate Elections: Theory and Experiments," Caltech mimeo.

Milgrom, P. and J. Roberts (1982). "Limit Pricing and Entry under Incomplete Information: An Equilibrium Analysis." Econometrica, Vol 50, No 2, March 1982.

Page, T. (1977)."Keeping Score: An Actuarial Approach to Zero-Infinity Dilemmas," in Energy Risk Management, G.T. Goodman and W.D. Rowe (eds). Academic Press, New York.

Roberts, H. (1965). "Probabilistic Production." American Statistical Association Journal, March 1965.

Savage, L.J. (1971). "Elicitation of Personal Probabilities and Expectations." J. of the American Statistical Association, Vol 66, No 336.

Walker, M. (1981). "A Simple Incentive Compatible Scheme for Attaining Lindahl Allocations." Econometrica, Vol 49, No 1, Jan. 1981.

Winkler, R.L. (1969). "Scoring Rules and the Evaluation of Probability Assessors." J. of the American Statistical Association, Sept. 1969.