Sociedad Española de Estadística
e Investigación Operativa

Test

Volume 11, Number 1. June 2002

Incentive contracts and strictly proper scoring rules

Robert T. Clemen
Fuqua School of Business
Duke University, USA.

Sociedad de Estadística e Investigación Operativa
Test (2002) *Vol.* **11**, *No.* 1, pp. 167–189

Sociedad de Estadística e Investigación Operativa
Test (2002) *Vol.* **11**, *No.* 1, pp. 167–189

Incentive contracts and strictly proper scoring rules

Robert T. Clemen*
Fuqua School of Business
Duke University, USA

Abstract

When a decision maker (DM) contracts with an expert to provide information, the nature of the contract can create incentives for the expert, and it is up to the DM to ensure that the contract provides incentives that align the expert's and DM's interests. In this paper, scoring rules (and related functions) are viewed as such contracts and are reinterpreted in terms of agency theory and the theory of revelation games from economics. Although scoring rules have typically been discussed in the literature as devices for eliciting and evaluating subjective probabilities, this study relies on the fact that strictly proper scoring rules reward greater expertise as well as honest revelation. We describe conditions under which a DM can use a strictly proper scoring rule as a contract to give an expert an incentive to gather an amount of information that is optimal from the DM's perspective. The conditions we consider focus on the expert's cost structure, and we find that the DM must have substantial knowledge of that cost structure in order to design a specific contract that provides the correct incentives. The model and analysis suggest arguments for hiring and maintaining experts in-house rather than using outside consultants.

Key Words: Scoring Rules, agency theory, revelation games, incentives, expert information

AMS subject classification: 62C99

1 Introduction

Probability forecasts have become relatively common as a way for experts to express their uncertainty about uncertain future events or variables. For example, meteorologists have routinely provided probabilistic forecasts in the United States since the late 1960s (Murphy and Winkler (1984)). Risk analyses of complex systems often require probabilistic assessments of risks by experts, and guidelines for making and using such assessments are available (e.g., Morgan and Henrion (1990), Cooke (1991), Meyer and Booker (1991)).

---

*Correspondence to: Robert T. Clemen, Fuqua School of Business, Duke University Durham NC 27708, USA. Email:clemen@mail.duke.edu

Received: December 1998; Accepted: November 2001

Scoring rules (de Finetti (1962), Winkler (1967), Savage (1971)) provide a way to evaluate probabilistic forecasts by calculating a reward based on the forecast and the actual outcome of the event or random variable. The broad class of strictly proper scoring rules has been characterized by Shuford et al. (1966), Savage (1971), and Schervish (1989). Strictly proper scoring rules play special roles in both ex-post and ex-ante senses. Used in an ex-post sense, strictly proper scoring rules can be used for forecast evaluation. Properties of strictly proper scoring rules used this way have been studied extensively. For a review, see Winkler (1996).

In an ex-ante sense, strictly proper scoring rules can be used as tools for eliciting subjective probabilities from knowledgeable experts, and their properties in this context have been studied by Stael Von Holstein (1970) and Winkler (1977, 1986) among others. In this paper we also adopt an ex-ante perspective, viewing a scoring rule as a contract between a decision maker (DM), who is interested in an unknown event or random variable Θ, and an expert, who has access to information about Θ. When considered as such a contract, a strictly proper scoring rule has two useful properties. First, it induces truthful revelation of the expert's information. Second, a strictly proper scoring rule has the property that a "more knowledgeable" expert must have a higher expected score (DeGroot and Fienberg (1982, 1983); Winkler (1977, 1986)). By more knowledgeable, we mean someone who is more sure of an event or who has a narrower distribution (in the sense of Rothschild and Stiglitz (1970)) for some random variable. More knowledge may be associated with greater expertise. It could also mean that the expert has spent more effort to reduce her uncertainty in a variety of ways, such as learning about the situation, collecting information, performing appropriate analyses, and so on. In general, the expert has an incentive to achieve a particular level of precision in her probabilistic beliefs. One way to formalize this, which we will do for the purposes of this article, is to model the expert's efforts as equivalent to gathering sample information from a process related to the variable of interest to the DM.

The observation that strictly proper scoring rules reward greater expertise with a higher expected score implies that the expert would prefer to collect more information if doing so were economically feasible. If the expert knows the scoring rule that will be used, she will collect information until her marginal cost equals the marginal increase in expected score. Thus, the expert's choice as to how much information to gather can be thought of as a preposterior choice of an optimal sample size, with the scoring rule playing the part of the expert's payoff function. The DM who consults the expert presumably has a decision to make that will be based in part on the information from the expert. The DM seeks the best trade-off between his own payoff and the expected cost of information; he wants the marginal increase in expected score (his cost) to equal the marginal increase in expected payoff from the extra sample observation. Again, the DM's decision is a preposterior choice of an optimal sample size, complicated by the fact that he has to design a scoring rule (the contract) that will make the expert perform as desired. In this case, the DM wants the expert to take a sample of a particular size and then report the results honestly.

Based on the above description, the DM is not using a scoring rule only to say, "Tell me what you know." A more complete paraphrase would be, "Go learn about the situation, collect information if needed, perform appropriate analyses, and then tell me what you learned." Thus, the scoring rule is like a contract and provides the expert with an incentive structure on which she bases her data-gathering and revelation decisions. When the expert's actions cannot be monitored, then the contract must provide the appropriate incentives so that the expert does as the DM desires.

This problem embodies two issues that are familiar in informational economics. First is the principal-agent problem: The principal in general cannot observe the agent's level of effort, and hence the agent may be tempted to cheat, or to provide less than the level of effort that the principal prefers. This is known as the moral-hazard problem, and the principal's problem is to construct a contract that ensures that the agent will provide the level of effort that the principal prefers. In our situation, the DM is the principal, and the expert is the agent; as long as the DM cannot monitor the expert's activities, the expert may exert more or less effort in data collection than the DM desires. If she chooses to spend less effort, the DM has less information than he wants. If she expends extra effort, then the DM may end up paying more for the additional information than it is worth. For more details about agency theory, see Holmstrom (1979), Jensen and Meckling (1976), and Ross (1973).

The second problem is called the incentive-compatibility problem. Consider a situation in which some or all players must reveal privately held information. Such a game is called a "revelation game," and the incentive-compatibility problem is that a player may prefer to reveal his or her information dishonestly. In the context of this paper, once the expert collects her information, she may have an incentive to misrepresent that information depending on the nature of the scoring rule. It is well known that strictly proper scoring rules are incentive-compatible; that is, they provide a positive incentive for the expert to be honest (Savage (1971)). It is not obvious, however, that the DM's optimal choice is to choose a strictly proper scoring rule. Fortunately, Myerson's (Myerson (1977)) revelation principle provides the proof that the DM can indeed limit his search to proper scoring rules and still be able to make an optimal choice.

We will use the term "contract" synonomously with "scoring rule", and we use these terms somewhat more broadly than usual; the term "scoring rule" as used in the Bayesian literature usually refers to a function $S(p, \theta)$, where θ is the realization of a random variable Θ, and p is a probability measure on Θ. In this paper, we may use the term occasionally to refer to functions $S(g[p], \theta)$, where g is some function of p (e.g., expectation).

This paper's contribution is to reinterpret scoring rules, drawing on agency theory and the theory of revelation games. We will see that the DM's problem in designing a contract (scoring rule) for the expert in general requires him to solve the moral-hazard problem. The incentive-compatibility problem is solved by the revelation principle, allowing the DM to consider only proper scoring rules. In analyzing the problem, we find that under some conditions the DM can use strictly proper scoring rules to resolve the agency issue. In particular, we consider conditions on the DM's knowledge about the expert's cost structure for obtaining information. Even in the relatively simple world of risk neutrality, the DM must know a considerable amount about that cost structure in order to devise an appropriate contract. If these conditions are met, finding an optimal contract involves no more than calculating appropriate scaling constants for the scoring rule. Cervera (1996) briefly discusses scoring rules from an agency-theory perspective. He provides a general formulation that is consistent with the model presented here, but he does not identify optimality conditions.

We proceed as follows. The next section formulates the DM's problem in the agency-theory framework under the assumption that the expert's cost structure is known. The analysis leads to examples using specific scoring rules in Section 3. In Section 4, we consider the DM's problem when we relax the assumption that the expert's cost structure is known. Section 5 summarizes the results, considers possible extensions to the model, and discusses implications for managers.

2 The DM's Problem

Consider the following formal description of the situation: A DM must take action $a \in \mathbf{A}$. The DM's payoff depends on a and the realization (θ) of uncertain random variable Θ, and we denote the DM's payoff function by $V(a, \theta)$. The DM has prior density and distribution functions $f'(\theta)$ and $F'(\theta)$, respectively.

In order to learn about Θ, the DM contracts with an expert to collect a sample of independent observations x_i from a stochastic process and to report what she learns (which we will make precise momentarily). For now, we assume that the DM and the expert agree on the prior distribution for θ and the likelihood function for the data. The DM and the expert must agree on the sample size $n \in \{1, \ldots, N\}$ where $N \in \mathbb{N}$ is fixed in advance. The expert's cost for collecting n observations is C_n, which is known to the DM. Let the vector (y, n) be a minimal sufficient statistic for observations $(x_1, ..., x_n)$, where $y \in \mathbf{Y}$. Let $f(\theta|y, n)$ denote the expert's posterior density for θ, given y and n. The expert reports $f(\theta|y_r, n)$ to the DM. Knowing n, and given the assumptions above, the DM can deduce y_r. We assume that the expert has the option to report her posterior density dishonestly. Thus, $y_r \in \mathbf{Y}$, but y_r is not necessarily equal to y.

We will consider contracts of the form $S(y_r, \theta) \in \mathbf{S}$, depending on both what the expert reports and on the realization θ. After hearing from the expert (and thus knowing y_r and n), the DM chooses action $a^*(y_r, n) \in \mathbf{A}$. After θ becomes known, the DM receives $V(a^*, \theta)$ and pays the expert according to the contract. For now, both DM and expert are assumed to be risk-neutral within this framework. Thus, V and S are measured in monetary units.

The assumptions above are all fairly strong, especially the assumptions that the DM and expert agree on prior and likelhood functions and that they are both risk neutral. These assumptions allow the development and analysis of a model from which our main results are derived. In the subsequent discussion, we examine the implications of relaxing these assumptions.

As described in the introduction, the DM's task can be informally stated at this point as choosing a contract that provides incentives for the expert to collect a sample of the optimal size and to reveal honestly the resulting information to the DM. "Optimal sample size" is in relation to the DM's decision problem; Raiffa and Schlaifer (1961) show that the DM selects an optimal sample size by balancing the incremental preposterior expected payoff due to additional information against the incremental expected cost of the sample (i.e., the expected cost to be paid via the contract).

Formally, the DM's problem is to specify both the scoring rule or contract S and the sample size n. This problem can be formulated as a constrained optimization problem:

$$\max_{n,S} \int_{\mathbf{Y}} \int_{\Theta} \left[V(a^*, \theta) - S(y_r, \theta)\right] dF(\theta|y_r, n) dF(y_r|n), \tag{2.1}$$

subject to:

$$\int_{\mathbf{Y}} \int_{\Theta} S(y_r, \theta) dF(\theta|y, n) dF(y|n) - C_n \geq 0 \tag{2.2}$$

$$n \in \arg\max \left[\int_{\mathbf{Y}} \int_{\Theta} S(y_r, \theta) dF(\theta|y, n) dF(y|n) - C_n \right]. \tag{2.3}$$

The objective function (2.1) can be interpreted as the DM's preposterior (Raiffa and Schlaifer (1961)) expected payoff that results from learning y_r from the expert, taking action a^*, and finally experiencing a realization θ. At this point, the DM receives $V(a^*, \theta)$, and the expert is paid $S(y_r, \theta)$. The expectation must be taken before knowing the expert's report y_r, and hence the use of the predictive distribution $F(y_r|n)$. Because the prior and likelihood are common knowledge, the DM can calculate $F(y_r|n)$ for any contract. Note that $F(y_r|n)$ in the objective function and $F(y|n)$ are different distributions, although the former can in principle be derived from the latter, given the relationship between the two for any specified contract. The DM optimizes (2.1) over both n and S because both affect his payoff.

Constraint (2.2) represents a rationality constraint for the expert, who will decline an assignment not expected to at least break even (in the sense of making an economic, but not excess, return). This constraint need only hold for the optimum n. Further, (2.2) must hold with equality for the optimal contract S by the following simple argument: Any S that satisfies (2.3) and satisfies (2.2) as a strict inequality can be modified by subtraction of a constant so that (2.2) holds with equality; (2.3) is not affected by the subtraction of a constant from S and hence would still hold. If (2.2) holds with equality, the expert is just indifferent between taking the job and not. Note that $F(y|n)$ is the predictive distribution of y given n. Under our assumptions, the DM and expert agree on this predictive distribution, and the DM can thus include it in his optimization problem.

Constraint (2.3) provides the incentive for the expert to choose sample size n because it is a maximum for her own expected profit. Following the discussion in the introduction, this is the constraint that resolves the moral-hazard problem. In our model, incorporating this constraint means that the expert's expected payoff net of sampling costs attains a maximum for the same n that maximizes the DM's objective function. As in (2.2), note that the DM and expert agree on $F(y|n)$.

With the problem formulated as optimizing (2.1) subject to (2.2) and (2.3), the incentive-compatibility question remains. That is, as it stands, an optimal contract could be one that gives the expert an incentive to be dishonest in revealing her information, in which case y_r would not be equal to y. A question is whether the DM should consider only incentive-compatible contracts (or proper scoring rules)? The answer to this question is yes, and the justification derives from Myerson's (Myerson (1977)) revelation principle. The DM's optimization problem can be viewed as a revelation game; that is, the DM presents the expert with a scoring rule, and the expert then reveals what she knows according to that rule. The optimal (n, S) pair is a Nash solution to this game: Given that the DM chooses this pair, the expert's best response is to collect n observations and report according to the rule. Likewise, given that the expert will behave optimally given contract S, the DM's best choice is n, S. Myerson (1977) (see also Myerson (1991) and Kreps (1990)) showed that all Nash equilibria in a revelation game can be replicated by an incentive-compatible mechanism. This is known as the revelation principle.

For our purposes, the revelation principle means that we can restrict our search to incentive-compatible contracts without eliminating any equivalent solutions to the DM's problem. To implement the revelation principle formally, we add the constraint:

$$\int_{\Theta} S(y,\theta)dF(\theta|y,n)dF(y|n) \geq \int_{\Theta} S(y_r,\theta)dF(\theta|y,n)dF(y|n), \forall y_r \in Y \tag{2.4}$$

which requires that the expert's expected score for reporting honestly ($y_r = y$) be at least as great as the expected score for reporting dishonestly. Thus, in what follows, we will consider the DM's problem as optimizing (2.1) subject to (2.2), (2.3), and (2.4). In the subsequent analysis and examples, we further limit our consideration to the class of strictly proper scoring rules.

Without the incentive-compatibility constraint (2.4), the DM's problem as formulated above has proven to be tractable as a standard agency problem (Conroy and Hughes (1987)). The addition of the incentive-compatibility constraint changes the problem somewhat. To proceed, we reformulate the DM's constrained optimization problem as suggested by Grossman and Hart (1983) by decomposing the problem in two stages:

Stage I. In the first stage, the DM finds an incentive-compatible contract (a strictly proper scoring rule) that induces the expert to choose a particular sample size $n \in \{1, ..., N\}$, and he does this for every n. Thus, he solves (2.5) below N times:

$$\max_{S(y_r,\theta|n)} \int_{\mathbf{Y}} \int_{\Theta} [V(a^*, \theta) - S(y_r, \theta)] dF(\theta|y_r, n) dF(y_r|n), \tag{2.5}$$

subject to:

$$\int_{\mathbf{Y}} \int_{\Theta} S(y_r, \theta) dF(\theta|y, n) dF(y|n) - C_n \geq 0 \tag{2.6}$$

$$n \in \arg\max \left[\int_{\mathbf{Y}} \int_{\Theta} S(y_r, \theta) dF(\theta|y, n) dF(y|n) - C_n \right] \tag{2.7}$$

$$\int_{\theta} S(y, \theta) dF(\theta|y, n) \geq \int_{\theta} S(y_r, \theta) dF(\theta|y, n), \ \forall y, y_r \in \mathbf{Y}. \tag{2.8}$$

The outcome of stage I is a set $\Sigma = \{S_1, ..., S_N\}$ of strictly proper scoring rules where S_n is an abbreviation for $S_n(y_r, \theta)$.

Stage II. In the second stage, the DM uses Σ in the following simple optimization problem:

$$\max_{S_n \epsilon \Sigma} \int_{\mathbf{Y}} \int_{\theta} [V(a^*, \theta) - S_n(y_r, \theta)] dF(\theta|y_r, n) dF(y_r|n). \tag{2.9}$$

Finding the optimum S_n in (2.9) amounts to choosing the sample size n^* that maximizes the DM's expected payoff net of the expected score paid to the expert. Grossman and Hart show that solving (2.5) and (2.9) is equivalent to solving (2.1).

The main concern of this section is whether the DM will be able to devise a strictly proper scoring rule that is a solution to (2.9). To do this, he must be able to find a strictly proper scoring rule that solves (2.5) for every n.

In order to proceed further, we assume that sampling costs increase with n at a weakly increasing rate:

$$0 < C_j - C_{j-1} \ \leq \ C_{j+1} - C_j, \forall j \in \{2, \ldots, N-1\}. \tag{2.10}$$

Under this assumption, the expert must expend additional effort for each additional independent observation, and there can be no returns to scale. A linear cost structure (constant marginal cost per observation) weakly satisfies (2.10). Clemen and Winkler (1985) describe a scenario of dependent information sources in which the cost of incremental equivalent independent observations could increase at an increasing rate.

We call $E_n(S) \equiv \int_{\mathbf{Y}} \int_{\theta} S(y, \theta) dF(\theta|y, n) dF(y|n)$ the preposterior expected score for S, given that n observations will be taken. Note that $E_{n-1}(S) \leq E_n(S)$, which follows directly from the LaValle's (LaValle (1968)) results regarding information value, especially that incremental information cannot have negative incremental expected value. In our case, S is the expert's payoff function, and the sample information is information about the random state variable θ. Thus, $E_n(S)$ must be nondecreasing in n.

For Proposition 2.1 below, we further require that $E_n(S)$ is strictly increasing at a strictly decreasing rate. Formally, suppose that a strictly proper scoring rule S exists such that:

$$0 < E_{n+1}(S) - E_n(S) < E_n(S) - E_{n-1}(S), \forall n \in \{2, \ldots, N-1\} \tag{2.11}$$

With (2.10) and (2.11), we have:

Proposition 2.1. *If (2.10) holds, and if a strictly proper scoring rule exists that satisfies (2.11), then the DM can solve (2.5) for every n.*

Proof. For ease of notation, let $E_n(S) \equiv E_n$. Choose b_n so that:

$$b_n(E_n - E_{n-1}) \quad > \quad C_n - C_{n-1} \tag{2.12}$$

$$b_n(E_{n+1} - E_n) \quad < \quad C_{n+1} - C_n. \tag{2.13}$$

Note that such a b_n always exists: Let

$$\begin{aligned} b^- \equiv (C_n - C_{n-1})/(E_n - E_{n-1}) &> 0, \text{ and} \\ b^+ \equiv (C_{n+1} - C_n)/(E_{n+1} - E_n) &> 0. \end{aligned}$$

Then $b^-(E_n - E_{n-1}) = C_n - C_{n-1} < C_{n+1} - C_n = b^+(E_{n+1} - E_n)$, which implies that $b^+ > b^-$. Any b_n such that $b^- < b_n < b^+$ will satisfy conditions (2.12) and (2.13).

We have to show that $b_n E_k - C_k < b_n E_n - C_n$, $\forall k \neq n$. Suppose $k < n$. Write

$$b_n(E_n - E_k) = b_n(E_n - E_{n-1} + E_{n-1} - E_{n-2} + \cdots + E_{k+1} - E_k)$$

and

$$C_n - C_k = C_n - C_{n-1} + C_{n-1} - C_{n-2} + \cdots + C_{k+1} - C_k.$$

Applying (2.12) and (2.13), we obtain directly that $b_n E_k - C_k < b_n E_n - C_n$. For $k > n$ a similar argument applies. Finally, choose a_n so that $a_n + b_n E_n = C_n$. Then $S_n = a_n + b_n S(y_r, \theta)$ is a strictly proper scoring rule that satisfies (2.6), (2.7), and (2.8). Thus there exists a solution to (2.5) $\forall n \in \{1, ..., N\}$. □

This proposition says simply that if the DM can find one strictly proper scoring rule that satisfies (2.11), and if the expert's cost structure satisfies (2.10), then the DM can scale that scoring rule in order to obtain a set of strictly proper scoring rules, each one of which induces the expert to choose a specific sample size n. Thus, the outcome in Stage II of the problem is guaranteed to be an n for which the scoring rule is strictly proper. Futhermore, the proof shows a technique for finding the appropriate scaling constants.

3 Examples

In this section we study two examples in which we show how the DM can devise strictly proper scoring rules in order to induce the expert to choose a specific n. To simplify the discussion, we abstract away from the DM's action: If the DM can create a strictly-proper scoring rule for each n, thus solving (2.5), he can in turn choose a member of that set that is optimum in the context of his decision, thereby solving (2.9).

Example 3.1 (Squared Error Penalty). Suppose the expert is required to provide her posterior mean μ_r of a distribution for a continuous random variable θ. In this case, then, we let $y_r = \mu_r$. The expert can be rewarded with a contract that incorporates a penalty proportional to the squared error:

$$S(\mu_r, \theta) = \ d_1 - d_2(\theta - \mu_r)^2,$$

where d_1 and d_2 are positive and μ_r is the reported mean of the expert's distribution. As indicated in the introduction, this is an example of a contract that is not a "scoring rule" in the sense that the term is typically used. It does, however, have similar properties. For example, suppose the expert has a prior density for θ with mean μ' and variance $Var(\theta)$. Then it is straightforward to show that $S(\mu_r, \theta)$ provides a positive incentive for the expert to report the mean of this density honestly, i.e., $\mu_r = \mu'$, in which case her expected score $E(S) = d_1 - d_2 Var(\theta)$. Hence we will call $S(\mu_r, \theta)$ "strictly proper."

Suppose now that the expert samples from a normal process with mean θ and known variance σ^2 and that $Var(\theta) = \sigma^2/t$. After observing n observations, her posterior distribution for θ is normal with mean $\mu = (t\mu' + \sum_{i=1}^{n} x_i)/(t+n)$ and variance $\sigma^2/(t+n)$. Thus, $E_n(S) = d_1 - d_2\sigma^2/(t+n)$, and it is straightforward to show that (2.11) is satisfied. Suppose also that the expert's sampling cost is linear in n such that $C_n = cn$, satisfying (2.10) weakly. Now, consider contracts S_j, $j \in \{1, ..., N\}$, with $d_1 = 2c(j + t/2)$ and $d_2 = c(j+t)^2/\sigma^2$ so that

$$S_j(\mu_r, \theta) = 2c(j + t/2) - \frac{c(j+t)^2}{\sigma^2}(\theta - \mu_r)^2. \tag{3.1}$$

For each j, $E_n(S_j) - C_n$ is maximized by setting $j = n$, for which $E_n(S_n) - cn = 0$. Thus, using equation (3.1), the DM is able to find N contracts to use in solving problem (2.5). These contracts make up the set Σ, and they are of the form $S_n(\mu_r, \theta) = 2c(n+t/2) - c(n+t)^2(\theta - \mu_r)^2/\sigma^2$ for $n = 1, ..., N$. The last step is for the DM to use Σ along with the characteristics of his decision situation, as embodied by $V(a^*, \theta)$, to solve (2.9) in which he chooses the optimal n^* and hence the optimal contract to offer the expert.

To take this example further, assume that the DM's decision problem is to choose an estimate a for θ and that $V(a, \theta) = u_1 - u_2(\theta - a)^2$ with u_1, $u_2 > 0$. The DM's expected payoff, assuming the expert optimally collects

n observations, is

$$EU_n = u_1 - u_2E(\theta - a)^2 - d_1 + d_2\sigma^2/(t+n) = u_1 - d_1 + (d_2 - u_2)\sigma^2/(t+n).$$

The second equality follows by noting that the DM's optimal a^* is to estimate θ with μ_r, the expert's honest report of the posterior mean, and that $E(\theta - \mu_r)^2$ is the posterior variance $\sigma^2/(t+n)$. Substituting for d_1 and d_2,

$$EU_n = u_1 - 2c(n + t/2) + [c(n+t)^2/\sigma^2 - u_2]\sigma^2/(t+n).$$

To find the optimal n^*, take first differences:

$$\Delta EU_n = EU_{n+1} - EU_n = u_2\sigma^2/[(t+n)(t+n+1)] - c,$$

which decreases in n. The optimal n^* is the least n such that $u_2\sigma^2/[(t+n)(t+n+1)] < c$, thus solving problem (2.9). Note that the DM could have enough prior information (t large enough) so that gathering more information would not be worthwhile. This would be the case if $\Delta EU_0 \leq 0$ or $u_2\sigma^2/[t(t+1)] \leq c$. In this case, $n^* = 0$, there is no contract with the expert, and the DM's $EU_0 = u_1 - u_2\sigma^2/t$. (Thanks to Dennis Lindley for suggesting this example and showing that the same results, including derivation of d_1 and d_2, can be obtained by a conventional optimization approach.) □

Example 3.2 (Binary Scoring Rules). Consider a situation in which the DM is interested in an upcoming event G and asks the expert to report the probability p that G will occur. In this case, the expert can be rewarded according to a binary scoring rule where she receives $S_1(p)$ if G occurs and $S_2(p)$ otherwise. From Savage (1971), strictly proper binary scoring rules can be generated by taking any function $J(p)$ that is convex and differentiable for $0 \leq p \leq 1$ and setting $S_1(p) = J(p) + (1-p)J'(p)$ and $S_2(p) = J(p) - pJ'(p)$. The expected score $E[S(p)] = J(p)$.

Suppose the expert can learn about the event by binary outcomes 1 (success) and 0 (failure). This experiment is related to G (and the complement G^c) by the likelihood function:

$$P(1 \mid G) = \alpha; \quad P(0 \mid G) = \beta;$$

$$P(1 \mid G^c) = \gamma; \quad P(0 \mid G^c) = \delta;$$

where $\alpha+\beta=\gamma+\delta=1$, and α,β,γ, and δ are all positive. If the expert's prior probability of the event is 0.5, then after observing the outcome from one experiment her probability will be:

$$P(G \mid n=1) = \begin{cases} \frac{\alpha}{(\alpha+\gamma)} \text{ with preposterior probability } P(1)=(\alpha+\gamma)/2 \\ \frac{\beta}{(\beta+\delta)} \text{ with preposterior probability } P(0)=(\beta+\delta)/2 \end{cases}.$$

Likewise, for any n identical independent experiments,

$$P(G \mid k \text{ successes}, \ n \text{ trials}) = P_{n,k} = \alpha^k\beta^{n-k}/(\alpha^k\beta^{n-k}+\gamma^k\delta^{n-k}).$$

Furthermore, prior to observing the n observations, we can calculate the predictive probability of getting k successes in n trials:

$$P(k \text{ successes}|n) = 0.5\binom{n}{k}\left(\alpha^k\beta^{n-k}+\gamma^k\delta^{n-k}\right).$$

Now consider the expert's preposterior expected score as she decides how many experiments to conduct. Her preposterior expected score $E_n[J(p)]$ for conducting n experiments can be written:

$$E_n[J(p)|n] = 0.5\sum_{k=0}^{n}\binom{n}{k}\left(\alpha^k\beta^{n-k}+\gamma^k\delta^{n-k}\right)J(P_{n,k}).$$

Taking $\alpha=\delta=0.75$ and $\beta=\gamma=0.25$, Table 1 shows how $E_n[J(p)]$ changes as the expert considers different sample sizes. $E_n[J(p)]$ has been calculated in Table 1 for the three common binary scoring rules shown below :

	$S_1(p)$	$S_2(p)$	$E[S(p)] = J(p)$
Quadratic	$-(1-p)^2$	$-p^2$	$-p(1-p)$
Logarithmic	$\log(p)$	$\log(1-p)$	$p\log(p)+(1-p)\log(1-p)$
Spherical	$p[J(p)]^{-1}$	$(1-p)[J(p)]^{-1}$	$p^2+(1-p)^2]^{1/2}$

Note that in all three cases the expected score increases at a decreasing rate, thus satisfying (2.11).

As before, assume that $C_n = cn$. To create a quadratic scoring rule that will induce the expert to choose sample size 5, for example, choose b_5 so that

$$b_5(E_5-E_4) = b_5(.0178) \geq c$$

n	Quadratic	Δ	Logarithmic	Δ	Spherical	Δ
0	-0.2500		-0.6931		0.7071	
1	-0.1875	0.0625	-0.5623	0.1308	0.7906	0.0835
2	-0.1500	0.0375	-0.4631	0.0992	0.8311	0.0406
3	-0.1205	0.0295	-0.3837	0.0794	0.8669	0.0357
4	-0.0988	0.0218	-0.3197	0.0640	0.8901	0.0232
5	-0.0810	0.0178	-0.2672	0.0525	0.9110	0.0210
6	-0.0672	0.0138	-0.2241	0.0431	0.9256	0.0146
7	-0.0557	0.0115	-0.1884	0.0358	0.9390	0.0134
8	-0.0466	0.0091	-0.1587	0.0297	0.9487	0.0096

Table 1: Preposterior expected scores $E_n[J(p)]$ for binary scoring rules. In the table, we use the constants $\alpha = \delta = 0.75$ and $\beta = \gamma = 0.25$. The column headed by Δ is $E_n[J(p)] - E_{n-1}[J(p)]$.

and

$$b_5(E_6 - E_5) = b_5(.0138) \leq c.$$

These conditions are satisfied when $b_5 = 65c$. Now choose a_5 so that $a_5 + 65c(-0.0810) = 5c$, which implies that $a_5 = 10.265c$. Thus, the strictly proper scoring rule that we seek is:

$$S_1(p) = 10.265c - 65c(1-p)^2$$

$$S_2(p) = 10.265c - 65cp^2$$

The following table shows the preposterior expected score $E_n[J(p)]$ using this scoring rule for n = 3, 4, 5, and 6:

$E_3[J(p)]$	$E_4[J(p)]$	$E_5[J(p)]$	$E_6[J(p)]$
$2.43c < C_3$	$3.84c < C_4$	$5.00c = C_5$	$5.90c < C_6$

□

4 Asymmetric Information

The analysis and examples above require that the DM know the expert's cost structure C. In this section, we investigate the importance of this assumption. We would like to find conditions under which the DM can guarantee that he will be able to create a contract that induces the expert to take a sample of a size that is optimal for both the expert and DM, and that the expert will report the resulting information honestly. Further, if the DM does not have this guarantee, can we say anything about the decision that he must make?

Suppose that the DM knows only that each element of C lies in a closed interval: $C_i^- \leq C_i \leq C_i^+$, $\forall\ i \in \{1, \cdots, N\}$. Now consider the general nature of the DM's problem. The task is to solve some version of (2.5) and (2.9) that deals with unknown cost C, and to offer the resulting contract to the expert in an effort to obtain her services. There are two observations to make. First, if the expert accepts the contract, the DM automatically knows that he will be paying the expert at least as much as her reservation wage, because the expert would reject the contract if her expected payoff were less. Second, if the expert accepts the contract, this only means that $E_n(S_n) - C_n \geq 0$ for some n, and in fact, the expert will choose sample size that maximizes this difference. Thus, the DM is unsure of the sample size actually chosen, but knows that he is paying too much.

Under some circumstances, the DM can get around the first problem. Suppose that he chooses contract S_j so that

$$E_j(S_j) - C_j^+ \geq 0, \text{ and} \tag{4.1}$$

$$E_k(S_j) - C_k^- \leq 0,\ k \neq j. \tag{4.2}$$

These conditions say that the only way the expert can obtain a positive expected profit is by choosing sample size j, for which he is guaranteed a positive profit. From this, we have:

Proposition 4.1. *If there exists a set of strictly proper scoring rules such that S_j satisfies (4.1) and (4.2) for every j, then $C_j^+ \leq C_{j+1}^-$, for every $j \in \{1, ..., N-1\}$.*

Proof. The proof follows directly from the monotonicity of $E_j(S_i)$ with respect to j and the conditions (4.1) and (4.2). □

This proposition says that if the DM wants to use a strictly proper scoring rule to induce the expert to take the optimal sample size, the DM must know that each C_j lies in an interval that does not overlap with the interval for any other C_i. The condition is necessary but not sufficient; even if the DM knows this much about the cost structure, he may still be unable to find a complete set of scoring rules, each of which satisfies (4.1) and (4.2).

A sufficient condition can be found, however. Replace condition (2.10) with:

$$C_j^+ \leq C_{j+1}^- \text{ and } C_j^+ - C_{j-1}^- \leq C_{j+1}^- - C_j^+, \ \forall j \in \{2, \ldots, N-1\}. \quad (4.3)$$

This is a strong condition specifying that the intervals are spaced in such a way as to guarantee that the costs increase at a weakly increasing rate. With (4.3), we have:

Proposition 4.2. *If (4.3) holds, and if a strictly proper scoring rule exists that satisfies (2.11), then the DM can solve (2.5) for every n.*

Proof. Choose b_n so that:

$$b_n(E_n - E_{n-1}) > C_n^+ - C_{n-1}^- \quad (4.4)$$

$$b_n(E_{n+1} - E_n) < C_{n+1}^- - C_n^+. \quad (4.5)$$

With (4.4) and (4.5) replacing (2.12) and (2.13), the proof follows the same lines as the proof of Proposition 2.1, ultimately finding constants a_n and b_n so that $E_n \geq C_n^+$ and $E_k < C_k^-$, $\forall\, k \neq n$. □

Example 4.1 (Binary Scoring Rules Revisited). Example 3.2 above demonstrated the technique for scaling a quadratic scoring rule so that it would induce the expert to take a sample of a specified size. Suppose now that the DM has only the following information about costs: $1 \leq C_1 \leq 2$, $2 \leq C_2 \leq 3, 5 \leq C_3 \leq 6, 9 \leq C_4 \leq 10$. A quadratic scoring rule that induces sample size 3 is: $S_1(p) = 22.513 - 137(1-p)^2$ and $S_2(p) = 22.513 - 137p^2$, leading to preposterior expected scores $E_n(S)$ as follows:

$E_n(S)$	$E_n(S)$	$E_n(S)$	$E_n(S)$
$-3.17 < C_1^-$	$1.96 < C_2^-$	$6.00 = C_3^+$	$8.98 < C_4^-$

Note, though, that this example does not satisfy (4.3) because $C_4^- - C_3^+ = 3 < C_3^+ - C_2^- = 4$. Thus, we have demonstrated that condition (4.3) is not a necessary condition. □

Up to this point, the focus of the paper has been to look for conditions that allow the DM to design a contract that induces the expert to obtain a sample of a specific (i.e., optimal) size. When the DM's knowledge of the expert's cost structure is such that he cannot guarantee a particular sample size, the basic problem changes. Now n is no longer a decision variable in the problem; the only decision is the scoring rule. Without any analysis, we can make some observations about the problem. First, any scoring rule S that the DM uses will have preposterior expected score $E_n(S)$ for each n. If the DM has a probability distribution for C, he can use this distribution to infer a probability distribution for the expert's choice of n, conditional on the expert accepting the contract. Using this distribution for n, the DM can compute a predictive distribution for y_r. That predictive distribution will be a mixture of predictive distributions for y_r conditional on n, with the mixing weights being the DM's probabilities for the n's. Equipped with this information, the DM can perform his preposterior analysis, conditional on the expert accepting the contract.

For any scoring rule, the DM can also calculate the probability that $E_n(S) < C_n$ for every n. This is probability π_S that the expert will refuse the contract altogether and the DM will have to take an immediate terminal action. Assuming that in this case the DM must make his decision based only on his prior information, the expected payoff of the optimal terminal action a^0 must be included in the payoff function. Because the DM no longer controls n, constraint (2.3) is no longer in the problem. The asymmetry of the information about C renders (2.2) irrelevant. Note, though, that the cost of not contracting and settling for a^0, which (2.2) prevented, is now endogenous to the problem through π_S. By the revelation principle, we can still restrict the DM to proper scoring rules. Thus, the DM's problem becomes:

$$\max_{S(y_r,\theta)} \left[\pi_S \int_\Theta V(a^0,\theta)dF(\theta) \right.$$
$$\left. +(1-\pi_S)\int_{\mathbf{Y}}\int_\Theta [V(a^*,\theta) - S(y_r,\theta)]dF(\theta|y_r)dF_m(y_r)\right] \tag{4.6}$$

subject to

$$\int_\Theta S(y,\theta)dF(\theta|y) \geq \int_\Theta S(y_r,\theta)dF(\theta|y), \forall y_r \in \mathbf{Y},$$

where F_m denotes the mixed predictive distribution. This problem is solvable, but further analysis yields little in the way of insights without substantial additional modeling assumptions.

5 Conclusion

In this paper, we have reinterpreted scoring rules using tools and concepts from agency theory and revelation games in economics. From this perspective, the DM's problem of contracting with an expert can be viewed as having to devise a contract or scoring rule that provides positive incentives for the expert to obtain an optimal sample size and then reveal the resulting information honestly. An important observation is that the use of strictly proper scoring rules (familiar from the statistics literature) is indeed quite appropriate; no improper scoring rule could make the DM better off. We further showed conditions under which the DM can design a scoring rule that provides the appropriate incentives to the expert, and those conditions centered on knowledge of the expert's cost structure. Proposition 1 shows that, under certain conditions, strictly proper scoring rules can be scaled to accomplish incentives as the DM desires. Although this is a relatively simple approach, it has important implications for managers in general, even though the specific conditions of the propositions may not apply. In particular, by scaling a strictly proper scoring rule, a manager can manipulate an expert's incentive to exert more or less effort, or to obtain more or less precise information.

Throughout, we have assumed that the expert and the DM agree on prior information. If the DM has some uncertainty about the expert's prior information, he may wish to model that uncertainty (see Kadane and Larkey (1982)). In this case, the DM can know only a noisy version of the expert's preposterior expected score for any scoring rule. This amounts to the same kind of informational asymmetry about the expert's expected payoff that we studied under uncertainty about C.

There is another useful perspective for considering the situation in which the expert has prior information and the DM knows that he has this information but not what it is. Augment C with a 0th element $C_0 = 0$ to represent the expert's incremental cost of acquiring the information that she already has. Thus $n = 1$ now means that the expert takes a sample of size 1 and incorporates the data with her prior information. The problem remains essentially as before, except that now one of the expert's alternatives is to acquire no new information, but simply to reveal her current information and earn her reservation wage. The DM still has the problem of assessing his beliefs about the expert's prior, and thus the discussion of the previous paragraph applies.

The model also assumes that the DM and the expert agree on the likelihood function. This is a strong assumption that, in the context of the model and taken along with agreement on prior information, allows the DM to adopt the expert's posterior distribution as his own without knowing the actual sample size taken. Whether it is reasonable to endow a DM with such knowledge is debatable; indeed, reasons for hiring an expert in the first place may include knowlege about available data and the nature of those data. One can imagine alternate models in which the DM is able to adopt the expert's posterior distribution as his own without this assumption; for example, the DM might believe that the expert's information set subsumes all of his own information.

Our model has assumed implicitly that the contract made between the DM and the expert is all that motivates the expert. Kadane and Winkler (1988) discuss the importance of this condition, showing that the presence of other incentives can, in principle at least, lead the expert to distort her information. In the context of our model, if the expert contracts with multiple DMs via different incentive schemes, it is no longer clear that a strictly proper scoring rule will provide a positive incentive to tell the truth.

Yet another assumption of the model is that θ is revealed before the DM pays the expert. In practice this rarely happens explicitly. The parallel, however, might be that the DM must decide in the future whether to hire the expert again; if the expert has not been accurate in the past, the DM may be less inclined to hire her.

Finally, we invoked the assumption of risk neutrality for both expert and DM. If the expert is not risk-neutral, but the DM knows the expert's utility function, then the contract or scoring rule can be designed to conform to that utility function. In particular, if the expert has utility function U_E, then the contract can be designed as $U_E^{-1}(S)$. If U_E is unknown, then the DM can assess his uncertainty about U_E and incorporate this uncertainty into the problem as described above in the case of uncertain information.

The problem is more complex if the DM is not risk neutral. Risk neutrality has played an important role in our model, albeit implicitly, because it allows the objective function in (2.1) to be separated into the expected value of the optimal action minus the expected value of the contract:

$$\int_{\mathbf{Y}} \int_{\Theta} [V(a^*, \theta) - S(y_r, \theta)] dF(\theta|y_r, n) dF(y_r|n)$$
$$= \int_{\mathbf{Y}} \int_{\Theta} V(a^*, \theta) dF(\theta|y_r, n) dF(y_r|n) - \int_{\mathbf{Y}} \int_{\Theta} S(y_r, \theta) dF(\theta|y_r, n) dF(y_r|n).$$

This allowed an important simplification in the analysis: We were able to focus only on the expected value of S rather than the form of S per se. Second, the economic interpretation of the objective function is clear: The DM seeks a contract that equates the marginal increase in expected payoff equal to the marginal increase in the expected cost of the contract. If the DM has utility function U_D, then (2.1) becomes

$$\max_{n,S} \int_{\mathbf{Y}} \int_{\Theta} U_D[V(a^*, \theta) - S(y_r, \theta)] dF(\theta|y_r, n) dF(y_r|n).$$

Optimizing this expression over n and S, subject to the constraints, becomes a problem in optimal control to choose functional S, which will involve more than choosing an S with an appropriate preposterior expected value. As mentioned in the introduction, Cervera (1996) discusses this problem briefly, but does not provide optimality conditions. We leave the analysis of this problem for future research.

The arguments in the previous paragraphs can be classified into matters of information and control on the part of the DM. On the information side, the DM needs to know about the expert's cost structure (and possibly utility function) and must agree with the expert regarding prior information and the likelihood function. On the control side, the DM needs the power to specify the scoring rule and also to control the nature and number of incentives to which the expert is subject. All of these requirements suggest reasons why the DM might prefer to hire and maintain experts in-house rather than hire outside consultants: Knowing the cost structure as well as agreeing on prior and likelihood may be more easily accomplished if the expert is an employee in the DM's firm. Likewise, controlling the specific scoring rule and limiting the expert's exposure to other incentives can be more easily accomplished if the expert is an employee. The DM's economic problem becomes that of deciding whether the cost of maintaining the expert in-house (as opposed to hiring an outside expert) is justified by the ability to exercise finer control of the expert's incentives. We leave the analysis of this problem to future research.

Acknowledgements

Thanks to Bob Winkler, Dennis Lindley, and to an anonymous associate editor for helpful comments on earlier drafts of this paper. As usual, all responsibility for errors reside with the author. This research has been supported by the National Science Foundation under grant SES 00-84383.

References

CERVERA, J. L. (1996). Discussion of scoring rules and the evaluation of probabilities. *Test*, 5:27–31.

CLEMEN, R. T. and WINKLER, R. L. (1985). Limits for the precision and value of information from dependent sources. *Operations Research*, 33:427–442.

CONROY, R. M. and HUGHES, J. S. (1987). Delegated information gathering decisions. *Accounting Review*, 62:50–66.

COOKE, R. M. (1991). *Experts in Uncertainty: Expert Opinion and Subjective Probability in Science*. Oxford University Press, New York.

DE FINETTI, B. (1962). Does it make sense to speak of good probability appraisers? In *The Scientist Speculates: An Anthology of Partly Baked Ideas*, pp. 357–364. Basic Books, New York.

DEGROOT, M. H. and FIENBERG, S. E. (1982). Assessing probability assessors: Calibration and refinement. In S. Gupta and e. J.O. Berger, eds., *Statistical Decision Theory an Related Topics III*, vol. 1, pp. 291–314. Academic Press, New York.

DEGROOT, M. H. and FIENBERG, S. E. (1983). The comparison and evaluation of forecasters. *The Statistician*, 32:12–22.

GROSSMAN, S. J. and HART, O. (1983). An analysis of the principal-agent problem. *Econometrica*, 51:7–46.

Holmstrom, B. (1979). Moral hazard and observability. *Bell Journal of Economics*, pp. 74–91.

Jensen, M. C. and Meckling, W. H. (1976). Theory of the firm: Managerial behavior, agency costs and ownership structure. *Journal of Financial Economics*, 3:305–360.

Kadane, J. B. and Larkey, P. D. (1982). Subjective probability and the theory of games. *Management Science*, 28:113–120.

Kadane, J. B. and Winkler, R. L. (1988). Separating probability elicitation from utilities. *Journal of the American Statistical Association*, 83:357–363.

Kreps, D. (1990). *A Course in Microeconomic Theory*. Princeton University Press, Princeton.

LaValle, I. H. (1968). On cash equivalents and information evaluation in decisions under uncertainty. *Journal of the American Statistical Association*, 63:252–290.

Meyer, M. A. and Booker, J. M. (1991). *Eliciting and Analyzing Expert Judgment: A Practical Guide*. Academic Press, London.

Morgan, M. G. and Henrion, M. (1990). *Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis*. Cambridge University Press, Cambridge.

Murphy, A. H. and Winkler, R. L. (1984). Probability forecasting in meteorology. *Journal of the American Statistical Association*, 79:489–500.

Myerson, R. B. (1977). Incentive compatibility and the bargaining problem. *Econometrica*, 47:61–73.

Myerson, R. B. (1991). *Game Theory: Analysis of Conflict*. Harvard University Press, Cambridge.

Raiffa, H. and Schlaifer, R. (1961). *Applied Statistical Decision Theory*. Graduate School of Business Administration, Harvard University, Cambridge.

Ross, S. (1973). The economic theory of agency: The principal's problem. *American Economic Review*, 63:134–139.

Rothschild, M. and Stiglitz, J. E. (1970). Increasing risk: I. A definition. *Journal of Economic Theory*, 2:225–243.

Savage, L. J. (1971). The elicitation of personal probabilities and expectations. *Journal of the American Statistical Association*, 66:783–801.

Schervish, M. J. (1989). A general method for comparing probability assessors. *Annals of Statistics*, 17:1856–1879.

Shuford, E. H., Albert, A., and Massengill, H. E. (1966). Admissible probability measurement procedures. *Psychometrica*, 31:125–145.

Stael Von Holstein, C. A. S. (1970). *Assessment and Evaluation of Subjective Probability Distributions*. The Economic Research Institute at the Stockholm School of Economics, Stockholm.

Winkler, R. L. (1967). The quantification of judgment: Some methodological suggestions. *Journal of the American Statistical Association*, 62:1105–1120.

Winkler, R. L. (1977). Rewarding expertise in probability assessment. In H. Jungermann and G. de Zeeuw, eds., *Decision-Making and Change in Human Affairs*, pp. 127–140. D. Reidel, Dordrecht.

Winkler, R. L. (1986). On "good probability appraisers". In P. K. Goel and A. Zellner, eds., *Bayesian Inference and Decision Techniques: Essays in Honor of Bruno de Finetti*, pp. 265–278. North-Holland, Amsterdam.

Winkler, R. L. (1996). Scoring rules and the evaluation of probabilities. *Test*, 5:1–26.