Mechanism Design for the Truthful Elicitation of Costly Probabilistic Estimates in Distributed Information Systems

Athanasios Papakonstantinou, Alex Rogers, Enrico H. Gerding, Nicholas R. Jennings*

Schoool of Electronics and Computer Science
University of Southampton
Southampton, SO17 1BJ, UK

Abstract

This paper reports on the design of a novel two-stage mechanism, based on *strictly proper scoring rules*, that allows a centre to acquire a costly forecast of a future event (such as a meteorological phenomenon) or a probabilistic estimate of a specific parameter (such as the quality of an expected service), with a specified minimum precision, from one or more agents. In the first stage, the centre elicits the agents' true costs and identifies the agent that can provide an estimate of the specified precision at the lowest cost. Then, in the second stage, the centre uses an appropriately scaled strictly proper scoring rule to incentivise this agent to generate the estimate with the required precision, and to truthfully report it. In particular, this is the first mechanism that can be applied to settings in which the centre has no knowledge about the actual costs involved in the generation an agents' estimates and *also* has no external means of evaluating the quality and accuracy of the estimates it receives. En route to this mechanism, we first consider a setting in which any single agent can provide an estimate of the required precision, and the centre can evaluate this estimate by comparing it with the outcome which is observed at a later stage. This mechanism is then extended, so that it can be applied in a setting where the agents' different capabilities are reflected in the maximum precision of the estimates that they can provide, potentially requiring the centre to select multiple agents and combine their individual results in order to obtain an estimate of the required precision. For all three mechanisms (the original and the two extensions), we prove their economic properties (i.e. incentive compatibility and individual rationality) and then perform a number of numerical simulations. For the single agent mechanism we compare the quadratic, spherical and logarithmic scoring rules with a parametric family of scoring rules. We show that although the logarithmic scoring rule minimises both the mean and variance of the centre's total payments, using this rule means that an agent may face an unbounded penalty if it provides an estimate of extremely poor quality. We show that this is not the case for the parametric family, and thus, we suggest that the parametric scoring rule is the best candidate in our setting. Furthermore, we show that the 'multiple agent' extension describes a family of possible approaches to select agents in the first stage of our mechanism, and we show empirically and prove analytically that there is one approach that dominates all others. Finally, we compare our mechanism to the peer prediction mechanism introduced by Miller et al. (2007b) and show that the centre's total expected payment is the same in both mechanisms (and is equal to total expected payment in the case that the estimates can be compared to the actual outcome), while the variance in these payments is significantly reduced within our mechanism.

Key words: Multiagent Systems, Scoring Rules, Auction Theory, Mechanism Design

*Tel: +44 (0) 23 8059 7681, Fax: +44 (0) 23 8059 2865
Email address: `nrj@ecs.soton.ac.uk` (Nicholas R. Jennings)

1. Introduction

Real-time information about the state of the world is increasingly being made available through distributed on-line systems that are owned by different stakeholders and accessed by multiple users. In such systems, it is important to develop processes that can evaluate the information provided to users and provide some guarantees to its quality. This is particularly so in cases where the information in question is an imprecise probabilistic estimate or forecast whose generation involves some cost. Examples include forecasts of future events such as weather conditions (Xue et al., 2004), where the costs are those of running a large scale weather prediction model, or probabilistic estimates of the quality of service within a reputation system (Jøsang et al., 2007) where such costs represent the computational task of accessing and evaluating previous interactions records. In such settings, it is reasonable to assume that the providers of such information are rational self-interested agents, and as such, may have an incentive to misreport their estimates, or to allocate less costly resources to their generation, if they can increase their own utility by doing so (e.g. by being rewarded for a more precise estimate than is actually provided or by claiming to expend more resources than was actually done)[1]. Thus, an information buyer must present the providers with a payment scheme that incentives the agents to commit resources to generating their estimates, and to truthfully report them.

A number of researchers have proposed the use of *strictly proper scoring rules* to address these challenges (Matheson and Winkler, 1976; Savage, 1971). Mechanisms using these rules reward accurate estimates or forecasts by making a payment to agents based on the difference between an event's predicted and actual outcome (observed at some later stage). Such mechanisms have been shown to incentivise agents to truthfully report their estimates in order to maximise their expected payment (Selten, 1998). This principle can be demonstrated through a meteorological scenario, which uses a *logarithmic scoring rule*. Specifically, we consider that a risk-neutral agent is asked to provide a probabilistic prediction of whether it will rain or not the following day. The agent's true estimate of the probability of rain tomorrow is denoted by p, and the prediction that it actually reports to the centre is denoted by $\widehat{p}$.

We first consider the perfectly plausible sounding rule that the agent should be rewarded in proportion to how confidently it predicted the actual outcome. That is:

$$S(\widehat{p}|x = \text{rain}) = \widehat{p} \quad \text{and} \quad S(\widehat{p}|x = \text{no rain}) = 1 - \widehat{p} \tag{1}$$

where x is the actual outcome verified the next day. In this case, the agent's expected utility is given by:

$$U(p,\widehat{p}) = p\widehat{p} + (1-p)(1-\widehat{p}) \tag{2}$$

Now, for any particular true belief, p, the agent will seek to report a value of $\widehat{p}$ which will maximise its expected utility. In this case, we note that $\partial U(p,\widehat{p})/\partial\widehat{p} = 2p - 1$, which is independent of $\widehat{p}$. Thus, we must consider the boundary conditions and find that if $p < 1/2$, then the agent maximises its expected reward by reporting $\widehat{p} = 0$, and if $p > 1/2$, then the agent maximises its expected reward by reporting $\widehat{p} = 1$. Clearly, under this scoring rule the agent will misreport its true beliefs, and thus, the centre will not receive a true estimate of the probability of rain tomorrow.

---

[1] Note that problems of this type are often categorised as *principal-agent problems* (Grossman and Hart, 1983; Rogerson, 1985) since there is an asymmetry of information between the contractor and contractee.

In contrast, consider the case when the logarithmic scoring rule is used such that the agent is now rewarded in proportion to the logarithm of the probability with which it predicted the actual outcome; that is:

$$S(\widehat{p}|x = \text{rain}) = \ln\widehat{p} \quad \text{and} \quad S(\widehat{p}|x = \text{no rain}) = \ln(1 - \widehat{p}) \qquad (3)$$

In this case, the agent's expected utility is given by:

$$U(p,\widehat{p}) = p\ln\widehat{p} + (1-p)\ln(1-\widehat{p}) \qquad (4)$$

and its derivative is given by:

$$\frac{\partial U(p,\widehat{p})}{\partial\widehat{p}} = \frac{p - \widehat{p}}{\widehat{p}(1-\widehat{p})} \qquad (5)$$

Now, solving $\partial U(p,\widehat{p})/\partial\widehat{p} = 0$ gives $\widehat{p} = p$, and thus, the agent will truthfully report its true belief regarding the probability of the outcome.

Due to the attractive property outlined above, strictly proper scoring rules have recently been used in computer science to promote the honest exchange of beliefs between agents (Zohar and Rosenschein, 2008), and within reputation systems to promote truthful reporting of feedback regarding the quality of a service experienced (Jurca and Faltings, 2005, 2006, 2007). Furthermore, Miller et al. (2007b,a) have exploited the fact that any affine transform of a strictly proper scoring rule is also a strictly proper scoring rule, and have shown how an appropriately scaled strictly proper scoring rule can be used induce agents to commit costly resources to generate their estimates[2].

While these approaches are effective in the specific cases that they consider, they all rely on the fact that the cost of the agent providing the estimate or forecast is known by the centre. This is not the case in our scenario where these costs represent private information known only to each individual agent (since they are dependent on the specific computational resources available to the agent). In this paper, we use techniques from *mechanism design* (Mas-Colell et al., 1995), and specifically auction theory (Krishna, 2002), to address this challenge. In particular, through the use of an auction protocol that uses *strictly proper scoring rules* to determine the payments to the agents, we incentivise the agents to truthfully reveal their costs to the centre and to generate and truthfully report an estimate at a required precision. In more detail, we introduce a novel two-stage mechanism. In the first stage, the centre elicits the agents' true costs and identifies the agent that can provide an estimate of the specified precision at the lowest cost. Then, in the second stage, the centre uses an appropriately scaled strictly proper scoring rule to incentivise this agent to generate the estimate with the required precision, and to truthfully report it.

We then go on to extend this mechanism in two ways. First, we relax the assumption that the selected agent can always provide an estimate as precise as the centre requires. Although this assumption is made in all the aforementioned work in this area, we believe it is unrealistic as often agents may have to deal with restrictions such as the lack of previous records for the reputation systems or the physical limits of making probabilistic predictions of real-world events, which subsequently enforce limitations on the precisions of the estimates they can provide. Hence, we extend the mechanism to consider the case where multiple suppliers can provide

---

[2]We shall describe their approach in detail in Section 3 since our results build upon their setting.

estimates, but due to their limited precisions, the centre may have to combine several of them in order to obtain the desired degree of accuracy. In doing so, we provide a non trivial extension of the initial mechanism in which a centre in the first stage asks N agents to report their costs and then pre-selects M of them, while in the second stage it asks those pre-selected agents to reveal their maximum precision and then to generate an estimate at that precision until it achieves its required precision.

Second, we relax the assumption that the centre has knowledge of the actual outcome of the event some time in the future after it receives the agents' estimates (in order that it can calculate the payments to the agents). Whilst this assumption is common in the strictly proper scoring rule literature, it is restrictive since in practice the centre may not always be able to observe the outcome. This may happen when reputation models are unable to monitor the constant changes in dynamic systems, such as markets, and hence the quality of a provided service cannot be verified, or when the agents' estimates relate to physical measurement from sensors deployed in hostile environment (such as floods (Zhou and Roure, 2007), glaciers (Hart and Martinez, 2006) or volcanoes (Werner-Allen et al., 2005)), where it is impossible to ascertain the 'ground truth' through external means. Mechanisms that operate under this regime are termed *self-verifying* by Goel et al. (2009)[3], and thus, we provide a third mechanism in which the centre uses the pre-selected agents' fused reported estimates, instead of the outcome, when calculating the payments. Now, Miller et al. (2007b) address this issue by evaluating an agent directly against each one of the other agents in turn, and hence, calculate the average payment due to each agent. However, we show that under their mechanism the payment that any agent receives is highly dependent on the accuracy of the other agents' reports. This results in an increase in both the variance in the payments received by each agent, and the variance in the total payment made by the centre. Thus, both the agents and the centre are more uncertain about the payments that they expect to receive and make. In contrast, our approach, which uses the fused estimates of all other agents, results in much lower variance in these payments.

In summing up, in this paper we contribute to the state of the art in the following ways:

- We introduce the first mechanism that elicits both effort and honest reporting of a single agent's estimate, in a setting where the centre has no information about the agents' costs involved in the generation of that estimate. We empirically evaluate our mechanism by comparing the standard quadratic, spherical and logarithmic scoring rules with a parametric family of scoring rules, and show that for certain values of the parameter, the resulting payment is similar to the logarithmic (optimal scoring rule) but also has finite lower bounds (as opposed to the logarithmic rule, which is potentially unbounded).

- In extending our initial mechanism, we present the first class of mechanisms that elicit estimates from multiple agents in a setting where the centre may have to combine several estimates of low precision due to agents' restrictions in the quality of the estimates

---

[3]Note that Goel et al. (2009) present a self-verifying mechanism which incentivises agents to truthfully reveal their subjective expectations regarding a physical parameter by scoring each individual agent's report against those of the other agents. Our approach is similar in that we score each agent against the fused reports of the other agents. However, Goel et al.'s mechanism operates in a very different setting in which the agents do not have to strategise over the precision of the measurement that they make, there is a common prior known to all agents, and there is no notion of a required minimum precision or a cost. Rather, the goal is to find the budget-balanced payments to be exchanged between the agents in order to ensure that they all truthful report.

they provide. We empirically compare several approaches to perform the pre-selection, hence the class of mechanisms, and identify one that minimises the centre's expected total payment.

- In extending the above mechanism we introduce a novel mechanism, in which the centre does not rely on knowledge of the realised outcome when calculating payments to the agents reporting their estimates. Furthermore, we modify strictly proper scoring rules accordingly so they can motivate agents to truthfully report their estimates, under the knowledge that they will be evaluated based on the other agents' reports.

- We compare the two extensions with the peer prediction mechanism (Miller et al., 2007b) and identify the differences between fusion and peer prediction. We show that the agents derive the same payment in all three mechanisms, hence the centre derives no additional penalty as a result of the lack of knowledge. However, we also note that the fusion mechanism results in payments of significantly smaller variance than the peer prediction, therefore it has more reliable and robust payments.

- We show that all three mechanisms are incentive compatible, in both costs and estimates revealed, and individually rational.

The rest of this paper is organised as follows: In Section 2 we present our problem formalisation, and in Section 3 we provide some necessary background on scoring rules. Based on this, in Section 4, we describe and analyse the single agent two-stage mechanism with unknown costs. In Section 5 we extend our mechanism so multiple agents can provide estimates of limited precisions, while in Section 6 we further extend our mechanism so the centre does not have to rely on knowledge of the actual outcome when calculating payments. Finally, we discuss related work in Section 7 and we conclude and discuss future work in Section 8.

2. The Information Elicitation Problem

We now describe in detail the information elicitation problem outlined in the introduction. We consider the case of a centre that wants to acquire a probabilistic forecast or estimate of some event characterised by a continuously valued parameter (e.g. a forecast of tomorrow's temperature, or a prediction of the latency of some online computational service), the unknown true value of which is denoted by x_0. The centre requires this estimate to have a minimum precision, denoted by θ_0. It derives no utility from an estimate whose precision is less than θ_0, and derives no additional benefit if the estimate is of precision greater than θ_0. We assume that the the centre is capable of paying an unbounded amount for this estimate, but it would clearly prefer to acquire an estimate of at least the minimum precision, at the minimum cost possible.

There are $N \geq 2$ rational, risk neutral agents, that can potentially provide the centre with this estimate. Each of these agents is capable of committing a variable amount of some costly resource in order to generate an independent noisy estimate of the parameter in question. As is common within the data fusion literature (see for example Gregory (2005) and DeGroot and Schervish (2002)), we model the estimates of the agents as Gaussian distributions with mean, x_i and precision θ_i, and we assume that these estimates are unbiased such that x_i is a random sample from the Gaussian distribution represented by $\mathcal{N}(x_0, 1/\theta_i)$. Note that this assumption does not

actually constrain the results such that they are only valid for Gaussian distributions. If we only know the mean and variance of a distribution, and nothing more, then the least constraining assumption to work with is that it is a Gaussian distribution (i.e. within the Bayesian framework, for any specific value of mean and variance, the Gaussian distribution is the distribution that exhibits the greatest Shannon entropy). It would be possible to extend the model to other general distributions (for example, using a beta distribution if the parameter is constrained to lie between 0 and 1, or a gamma distribution if it were just constrained to be positive). However, in general, working with Gaussian distributions is both widely applicable and also analytically tractable (since both the fusion and summation of two Gaussian distributions results in another Gaussian distribution).

We assume that the greater the resources committed by the agent, the greater the precision, θ_i, of the estimate that it generates and the greater the cost that it incurs. These costs are private to the agent and are described by the function $c_i(\theta_i)$. We assume that this cost function is double differentiable and convex (i.e. $c_i''(\theta) \geq 0$), and note that this is a realistic assumption in all cases where there are diminishing returns as more resources are committed. Finally, we do not assume that all agents use the same cost function, but we do demand that the costs of different agents and their derivatives do not cross (i.e. the ordering of the agents' costs and their derivatives is the same over all precisions) in order to prove incentive compatibility and individual rationality of the mechanisms that we derive[4]. In the examples that we provide in this paper we shall assume that that cost functions are in fact linear, such that $c_i(\theta_i) = c_i\theta_i$, and we note that this corresponds to the continuous limit of the case where an agent make n independent estimates with fixed precision, $\Delta\theta_i$ and cost Δc_i, and forms its final estimate by fusing these together such that $\theta_i = n\Delta\theta_i$ and $c_i(\theta_i) = \left(\frac{\Delta c_i}{\Delta\theta_i}\right)\theta_i$.

Given this setting, our challenge is then to design a mechanism that enables the centre to identify the agent that can provide the required estimate at the lowest cost, and to provide a payment to this agent such that it is incentivised to generate an estimate with a precision at least equal to that required and to report it truthfully. This payment is conditioned on the true value of the parameter, x_0, which is revealed to both the centre and the agents at some time after the estimate is required.

We then extend this initial setting to consider the case that the precision of the estimates that any agent can generate is constrained such that $\theta_i \leq \theta_i^c$. As before, the maximum precision that any agent can generate, θ_i^c, is the private information of the individual agent. Finally, we also relax the constraint that the value of x_0 is revealed to the centre and agents before the payments must be made. Thus, the centre can now only condition the payment to any agent on the estimates that were received from the other agents (and not from the true parameter value).

3. Strictly Proper Scoring Rules

Given the problem setting described above, we now describe the use of strictly proper scoring rules to incentivise the agents to make and truthfully report their estimates to the centre in the conventional case where the costs of the agents are assumed to be known.

---

[4]In Corollary 1 and Lemma 1, we prove that these assumptions regarding the cost functions and their derivatives are necessary.

Table 1: Comparison of quadratic, spherical, logarithmic and parametric scoring rules.

Scoring Rule:	Quadratic	Spherical
$S(x_0; \widehat{x}, \widehat{\theta})$	$2\mathcal{N}(x_0; \widehat{x}, 1/\widehat{\theta}) - \frac{1}{2}\sqrt{\frac{\widehat{\theta}}{\pi}}$	$\left(\frac{4\pi}{\widehat{\theta}}\right)^{\frac{1}{4}} \mathcal{N}(x_0; \widehat{x}, 1/\widehat{\theta})$
$\overline{S}(\theta)$	$\frac{1}{2}\sqrt{\frac{\theta}{\pi}}$	$\left(\frac{\theta}{4\pi}\right)^{\frac{1}{4}}$
$\overline{S}'(\theta)$	$\frac{1}{4\sqrt{\pi\theta}}$	$\frac{1}{4}\left(\frac{1}{4\pi\theta^3}\right)^{\frac{1}{4}}$
α	$4c'(\theta_0)\sqrt{\pi\theta_0}$	$4c'(\theta_0)\left(4\pi\theta_0^3\right)^{\frac{1}{4}}$
β	$c(\theta_0) - 2\theta_0 c'(\theta_0)$	$c(\theta_0) - 4\theta_0 c'(\theta_0)$

Scoring Rule:	Logarithmic	Parametric
$S(x_0; \widehat{x}, \widehat{\theta})$	$\log\mathcal{N}(x_0; \widehat{x}, 1/\widehat{\theta})$	$k\mathcal{N}(x_0; \widehat{x}, \theta)^{(k-1)} - \frac{k-1}{\sqrt{k}}\left(\frac{2\pi}{\widehat{\theta}}\right)^{\frac{1-k}{2}}$
$\overline{S}(\theta)$	$\frac{1}{2}\log\left(\frac{\theta}{2\pi}\right) - \frac{1}{2}$	$\frac{1}{\sqrt{k}}\left(\frac{2\pi}{\theta}\right)^{\frac{1-k}{2}}$
$\overline{S}'(\theta)$	$\frac{1}{2\theta}$	$\frac{k-1}{2\theta\sqrt{k}}\left(\frac{2\pi}{\theta}\right)^{\frac{1-k}{2}}$
α	$2c'(\theta_0)\theta_0$	$\frac{2c'(\theta_0)\theta_0\sqrt{k}}{k-1}\left(\frac{\theta_0}{2\pi}\right)^{\frac{1-k}{2}}$
β	$c(\theta_0) - 2c'(\theta_0)\theta_0\left(\frac{1}{2}\log\left(\frac{\theta_0}{2\pi}\right) - \frac{1}{2}\right)$	$c(\theta_0) - \frac{2\theta_0}{k-1}c'(\theta_0)$

Where $\mathcal{N}(x_0; \widehat{x}, \widehat{\theta}) = \sqrt{\frac{\widehat{\theta}}{2\pi}}\exp\left(\frac{\widehat{\theta}(\widehat{x}-x_0)^2}{2}\right)$.

3.1. Background

As seen in the introduction, scoring rules incentivise a risk neutral forecaster to truthfully report its forecast by maximising its expected reward. As such, this approach has been widely used as a statistical tool for eliciting personal beliefs and expectations regarding a future event (Brier, 1950; Hendrickson and Buehler, 1971; Savage, 1971). In particular, if an agent actually has an estimate represented by the probability density function $Q(x)$, but reports an estimate to the centre denoted by $R(x)$ and then receives a payment conditioned on this reported estimate and the true value revealed sometime later, $S(x_0|R(x))$, then the agent's expected score will be denoted as follows:

$$\overline{S}(R, Q) = \int_{-\infty}^{\infty} Q(x)S(x|R(x))dx \tag{6}$$

A scoring rule is defined as *strictly proper* if the agent's expected score is maximised when it reports the truth i.e. $R = Q \Leftrightarrow \overline{S}(Q,Q) > \overline{S}(R,Q)$. In this case the agent has an incentive to report the truth in order to maximise its expected utility.

Against this background, much of the literature on strictly proper scoring rules concerns three specific rules (quadratic, spherical and logarithmic) and a parametric family of rules known as the *power rule family* or *k-power scoring rules* (Selten, 1998). These four strictly proper scoring rules are defined in the following way:

1. Quadratic: $\quad S(x_0, R(x)) = 2R(x_0) - \int_{-\infty}^{\infty} R(x)^2 dx$

2. Spherical: $\quad S(x_0, R(x)) = R(x_0) / \sqrt{\int_{-\infty}^{\infty} R(x)^2 dx}$

3. Logarithmic: $\quad S(x_0, R(x)) = \log R(x_0)$

4. Parametric: $\quad S(x_0, R(x)) = kR(x_0)^{(k-1)} - (k-1) \int_{-\infty}^{\infty} R(x)^k dx$

where $k \in (1, 3)$, and when $k = 2$ the parametric rule takes the form of the quadratic rule.

Given that in our setting we are considering estimates in the form of Gaussian distributions, we can re-derive these scoring rules for this specific case[5]. Table 1 shows each of the four strictly proper scoring rules, $S(x_0; \widehat{x}, \widehat{\theta})$, in the case that the agent reports its estimate as a Gaussian distribution with mean $\widehat{x}$ and precision $\widehat{\theta}$. By integrating over this expression, we can also derive the score that the agent expects to derive, $\overline{S}(\theta)$, given that it has generated and truthfully reported (as it is incentivised to do) an estimate of precision θ.

3.2. Eliciting Effort when Costs are Known

Now, although eliciting truthful reports (incentive compatibility) is one very desirable property of the strictly proper scoring rules, it is certainly not the only one. In our setting, agents may decide to commit less than the required resources into the generation of the probabilistic estimate if they expect to increase their utility functions by doing so. To combat this, Miller et al. (2007b) elicit effort through the use of appropriate scaling parameters, noting that any affine transformation of a strictly proper scoring rule does not affect its incentive compatibility property. Given knowledge of an agent's costs, they show that it is possible to induce an agent to make and truthfully report an estimate with a specified precision, θ_0. In this case, the payment that an agent expects to receive, $\overline{P}(\theta)$, is given by:

$$\overline{P}(\theta) = \alpha \overline{S}(\theta) + \beta \tag{7}$$

where α and β are the scaling parameters, and the expected utility of the agent is given by:

$$\overline{U}(\theta) = \alpha \overline{S}(\theta) + \beta - c(\theta) \tag{8}$$

The centre can now choose the value of α such that the agent's utility (its payment minus its costs) is maximised when it produces and truthfully reports an estimate of the required precision, θ_0. To do so, it solves $\left. \dfrac{d\overline{U}}{d\theta} \right|_{\theta_0} = 0$ to give:

$$\alpha = \frac{c'(\theta_0)}{\overline{S}'(\theta_0)} \tag{9}$$

---

[5]Note that although we provide analytical results for the specific case of Gaussian distributions, equivalent results could be derived for any continuous distribution used.

In Table 1 we present this result, and the derivative of the expected score, $\overline{S}'(\theta)$, that is required to calculate it, for each of the four strictly proper scoring rules presented earlier.

Having defined the α parameter of the affine transformation that elicits effort and honest reporting, we calculate parameter β which motivates agents to participate in the mechanism by ensuring that their expected utility is always positive. In more detail, we now note that in order for a self-interested agent to incur the cost of producing a forecast, it must expect to derive positive utility from doing so. Thus, the centre can use the constant β to ensure that it makes the minimum payment to the agent, hence ensuring that the mechanism is individually rational. When costs are known, the centre can do so by making the agents indifferent between producing the forecast or not, by setting $\overline{U}(\theta_0) = 0$, thus giving:

$$\beta = c(\theta_0) - \frac{c'(\theta_0)}{\overline{S}'(\theta_0)}\overline{S}(\theta_0) \tag{10}$$

Again, cells β in Table 1 show this result for each of the four scoring rules.

Finally, it should be noted that the expected score of the quadratic, spherical and logarithmic scoring rules, as a function of the precision θ (expressed as $\overline{S}(\theta)$ in Table 1) is strictly concave, strictly increasing and twice differentiable. While we will show that this property of the expected scores is important to guarantee certain economic properties of the mechanism, it does not hold for all strictly proper scoring rules. For example, in the parametric scoring rule for $k > 3$ the second derivative of the expected score (denoted by Equation 11) becomes positive and therefore the expected score convex, which in turn, as we will show in the following section, results in a payment that fails to incentivise an agent to produce an estimate at the required precision, θ_0.

$$\overline{S}''(\theta) = \frac{(1-k)(3-k)}{4\theta^2\sqrt{k}} \left(\frac{2\pi}{\theta}\right)^{\frac{1-k}{2}} \tag{11}$$

Therefore, and in order to guarantee the concavity of the expected score we will restrict the parameter k to the space $(1,3)$.

4. A Mechanism for Dealing with Unknown Costs

We now consider the setting where the costs involved in the generation of a probabilistic estimate are unknown to the centre, and the centre wants to select a single agent to procure the estimate of the required precision at the lowest cost. As described in Section 2, we assume that there are multiple agents, all of which are capable of producing an estimate of at least the required precision (we shall relax this assumption in Section 5 where we consider agents that have a limitation on the maximum precision of the estimate they produce).

4.1. The Mechanism

We address the above mentioned challenges by designing a two-stage mechanism (Mechanism 1). In the first stage, the centre elicits the agents' true costs and identifies the agent that can provide an estimate of the specified precision at the lowest cost. Then, in the second stage, the centre uses an appropriately scaled strictly proper scoring rule in order to incentivise this agent to generate the estimate with the required precision, and to truthfully report it. At first glance it might seem that this mechanism is akin to a reverse second-price or Vickrey auction (Vickrey,

---

Mechanism 1 The mechanism for dealing with unknown costs:

---

1. First Stage

1.1 The centre announces that it needs an estimate of required precision θ_0, and asks all agents $i \in \{1, \ldots, N\}$, where $N \geq 2$, to report their cost functions $\widehat{c}_i(\theta)$[6].

1.2 The centre assigns the estimate to the agent who reported the lowest cost at the required precision, i.e., agent i such that $\widehat{c}_i(\theta_0) = min_{k \in \{1, \ldots, N\}} \widehat{c}_k(\theta_0)$.

2. Second Stage

2.1 The centre announces a scoring rule $\alpha S(x_0; \widehat{x}, \widehat{\theta}) + \beta$, where:

    (i) $S(x_0; \widehat{x}, \widehat{\theta})$ is a strictly proper scoring rule,

    (ii) $\overline{S}(\theta)$ is strictly concave as a function of precision θ,[7] and

    (iii) α and β are determined using equations 9 and 10 respectively, but now based on the second-lowest reported cost functions (i.e. $\widehat{c}_j(\theta)$ such that $\widehat{c}_j(\theta_0) = min_{k \neq i} \widehat{c}_k(\theta_0)$).

2.2 The agent selected in the first stage produces an estimate with mean x and precision θ, and reports $\widehat{x}$ and $\widehat{\theta}$ to the centre.

2.3 Once the actual outcome has been observed, the centre then gives the following payment to the agent:

$$P(x_0; \widehat{x}, \widehat{\theta}) = \alpha S(x_0; \widehat{x}, \widehat{\theta}) + \beta \tag{12}$$

1961), where the agents' rewards are equal to the second-lowest reported costs. This is indeed the case, however here the selected agent's reward in the second stage is determined by scaling the scoring rule using the second lowest cost identified in the first stage (rather than using the selected agent's reported costs).

4.2. Economic Properties of the Mechanism

Having detailed the mechanism, in the next section we identify and prove its economic properties. Specifically, in this section we show that:

1. The mechanism outlined above is incentive compatible in the first stage regarding the costs. In particularly, truthful revelation of the agents' cost functions is a weakly dominant strategy.

2. The mechanism is incentive compatible regarding the selected agent's reported measurement and precision in the second stage.

3. There can be no incentive compatible mechanism regarding the agents' cost functions revealed when the cost functions overlap.

4. The mechanism is individually rational.

---

[6]We note that in practice the centre only requires $\widehat{c}_i(\theta_0)$ and $c_i'(\theta_0)$, and not the entire functions. However, for notational convenience we request the agents to reveal their entire cost function.

[7]We note that the quadratic, spherical, logarithmic and parametric scoring rules satisfy both of these properties (see row 2 of Table 1).

5. The centre motivates the selected agent to make an estimate with a precision which is at least as high as θ_0, the precision required by the centre. We refer to the actual precision produced as the 'optimal precision' (from the perspective of the agent) θ^*, since for this precision the expected payment is maximised.

In this section, with prove the economic properties of the mechanism. Initially, we derive two lemmas which are then used in the proofs of the theorems that follow. The first of these lemmas shows that if the true costs of the agent performing the estimate are greater than the costs which are used to scale the scoring rule, then the agent's utility will always be negative, regardless of the precision.

Lemma 1. *If $c_t(\theta)$ and $c_s(\theta)$ are convex functions with $c_t(\theta) > c_s(\theta)$, $c_t'(\theta) > c_s'(\theta)$ and $c_t(0) = c_s(0) = 0$, where $c_t(\theta)$ is the agent's true cost function, $c_s(\theta)$ is the cost function used to scale the scoring function and $c_t'(\theta)$ and $c_s'(\theta)$ their respective derivatives, then $\overline{U}(\theta) < 0$ for any θ.*

Proof. Concavity of the expected score $\overline{S}(\theta)$ implies:

$$\overline{S}'(\theta_0)(\theta - \theta_0) \geq \overline{S}(\theta) - \overline{S}(\theta_0) \tag{13}$$

Similarly, convexity of the cost function $c_s(\theta)$ gives:

$$c_s'(\theta_0)(\theta - \theta_0) \leq c_s(\theta) - c_s(\theta_0) \tag{14}$$

Given that by definition $\overline{S}(\theta)$ and $c_s(\theta)$ are strictly increasing (as stated in the model description in Section 2), dividing with $\overline{S}'(\theta_0)$ and $c_s'(\theta_0)$ maintains the sign in inequalities 13 and 14. Therefore, from:

$$(\theta - \theta_0) \geq \frac{\overline{S}(\theta) - \overline{S}(\theta_0)}{\overline{S}'(\theta_0)}$$

and

$$(\theta - \theta_0) \leq \frac{c_s(\theta) - c_s(\theta_0)}{c_s'(\theta_0)}$$

it follows that:

$$\frac{\overline{S}(\theta) - \overline{S}(\theta_0)}{\overline{S}'(\theta_0)} \leq \frac{c_s(\theta) - c_s(\theta_0)}{c_s'(\theta_0)}$$

or

$$\frac{c_s'(\theta_0)}{\overline{S}'(\theta_0)}(\overline{S}(\theta) - \overline{S}(\theta_0)) + c_s(\theta_0) - c_s(\theta) \leq 0 \tag{15}$$

Now, the expected utility, is given by $\overline{U}(\theta) = \alpha\overline{S}(\theta) + \beta - c(\theta)$ (Equation 8), with the scaling parameters α and β already defined using Equations 9 and 10 as $\alpha = \dfrac{c'(\theta_0)}{\overline{S}'(\theta_0)}$ and $\beta = c(\theta_0) - \dfrac{c'(\theta_0)}{\overline{S}'(\theta_0)}\overline{S}(\theta_0)$. Therefore, an agent's expected utility is given by:

$$\overline{U}(\theta) = \frac{c_s'(\theta_0)}{\overline{S}'(\theta_0)}(\overline{S}(\theta) - \overline{S}(\theta_0)) + (c_s(\theta_0) - c_t(\theta)) \tag{16}$$

11

Therefore, since $c_t(\theta) > c_s(\theta)$, *for any* θ the following holds:

$$\frac{c'_s(\theta_0)}{\overline{S}'(\theta_0)}(\overline{S}(\theta) - \overline{S}(\theta_0)) + c_s(\theta_0) - c_s(\theta) \leq 0 \Rightarrow$$

$$\frac{c'_s(\theta_0)}{\overline{S}'(\theta_0)}(\overline{S}(\theta) - \overline{S}(\theta_0)) + c_s(\theta_0) - c_t(\theta) < 0 \Rightarrow$$

or

$$\overline{U}(\theta) = \frac{c'_s(\theta_0)}{\overline{S}'(\theta_0)}(\overline{S}(\theta) - \overline{S}(\theta_0)) + c_s(\theta_0) - c_t(\theta) < 0$$

$\square$

The next lemma shows that if the true costs of the agent performing the estimate are less than the costs used to scale the scoring rule, then the optimal precision θ^* will be greater than θ_0.

Lemma 2. *If $c_t(\theta)$ and $c_s(\theta)$ are convex functions with $c_t(\theta) < c_s(\theta)$, $c'_t(\theta) < c'_s(\theta)$ and $c_t(0) = c_s(0) = 0$, where $c_t(\theta)$ is the agent's true cost function, $c_s(\theta)$ is the cost function used to scale the scoring function and $c'_t(\theta)$ and $c'_s(\theta)$ their respective derivatives, then $\theta^* > \theta_0$.*

Proof. The agent's optimal precision, θ^*, which maximises its expected utility is formally denoted by $\theta^* = \text{argmax}_\theta \overline{U}(\theta)$, with $\overline{U}'(\theta^*) = 0$. Now, the agent's expected utility is already defined by Equation 16 as:

$$\overline{U}(\theta) = \frac{c'_s(\theta_0)}{\overline{S}'(\theta_0)}(\overline{S}(\theta) - \overline{S}(\theta_0)) + (c_s(\theta_0) - c_t(\theta))$$

Given that the optimal precision, θ^*, maximises the expected score, we have $\overline{U}'(\theta^*) = 0$, and hence, after replacing θ with θ^* and calculating the derivative of the expected utility (Equation 16):

$$\frac{c'_s(\theta_0)}{\overline{S}'(\theta_0)}\overline{S}'(\theta^*) - c'_t(\theta^*) = 0 \Leftrightarrow \frac{\overline{S}'(\theta^*)}{\overline{S}'(\theta_0)} = \frac{c'_t(\theta^*)}{c'_s(\theta_0)} \tag{17}$$

Let $f(\theta) = \overline{S}'(\theta)/\overline{S}'(\theta_0)$ and $g(\theta) = c'_t(\theta)/c'_s(\theta_0)$. Now, since $\overline{S}(\theta)$ is (strictly) concave, strictly increasing and twice differentiable, then $f'(\theta) \leq 0$ for all θ_0. Furthermore, since we also assume that the cost functions, and their derivatives, maintain the same ordering, without overlapping for all θ, then since $c''_t(\theta) \geq 0$ (due to the convexity of the cost) and $c'_s(\theta) \geq 0$ (since cost functions are strictly increasing), then $g'(\theta) \geq 0$ for all θ, and since $c'_t(\theta) < c'_s(\theta)$ for all θ, then $g(\theta_0) < 1$. Finally, since $g(\theta)$ is strictly increasing and $f(\theta_0) = 1$, then $f(\theta)$ and $g(\theta)$ must cross (i.e. $f(\theta^*) = g(\theta^*)$) at $\theta^* > \theta_0$. $\square$

Based on these two key lemmas, we now proceed to prove the four economic properties of our mechanism.

Theorem 1. *Truthful revelation of the agents' cost functions in the first stage of the mechanism is a weakly dominant strategy.*

Proof. We prove this by contradiction. Let $c_t(\theta)$ and $\widehat{c}(\theta)$ denote an agent's true and reported cost functions respectively. Furthermore, let $c_s(\theta)$ denote the cost function used to scale the scoring function *if the agent wins* (i.e. if $\widehat{c}(\theta_0) < c_s(\theta_0)$).

First, suppose that the agent misreports, but this does not affect whether it wins or not. In this case, since the costs are based on the second-lowest costs, this does not affect the scoring rule if the agent wins. Moreover, if the agent loses, the payoff is always zero. Therefore, there is no incentive to misreport.

Second, suppose that the agent's misreporting affects whether that agent is pre-selected or not. There are now two cases:

1. The agent wins by misreporting, but would have lost when truthful.
2. The agent loses by misreporting, but would have won when truthful.

In this context:

- Case (1) can be formally denoted as $c_t(\theta_0) > c_s(\theta_0)$ and $\widehat{c}(\theta_0) < c_s(\theta_0)$. Now, since the true cost $c_t(\theta_0) > c_s(\theta_0)$, it follows directly from Lemma 1 that the expected utility $\overline{U}(\theta)$ is strictly negative, irrespective of θ. Therefore, the agent could do strictly better by reporting truthfully in which case the expected utility is zero.

- Case (2) can be formally denoted as $c_t(\theta_0) < c_s(\theta_0)$ and $\widehat{c}(\theta_0) > c_s(\theta_0)$. In this case the agent would have won by being truthful, but now receives a utility of zero. To show that this type of misreporting is suboptimal, we need to show that, when $c_t(\theta_0) < c_s(\theta_0)$, an agent benefits from being selected and generating the (optimal) estimate (i.e. $U(\theta^*) > 0$ when $c_t(\theta_0) < c_s(\theta_0)$). Now, since θ^* is optimal by definition, then $U(\theta^*) \geq U(\theta_0)$. From the expected utility in equation 16, we have $\overline{U}(\theta_0) = c_s(\theta_0) - c_t(\theta_0) > 0$ when $c_t(\theta_0) < c_s(\theta_0)$, and hence $U(\theta^*) > 0$ at true costs reporting.

$\square$

Corollary 1. *Incentive compatibility with respect to agents' reported costs and precisions does not hold if the agents' cost functions cross at θ'.*

Proof. The proof regarding the agents' reported costs comes directly from the above theorem, as we need to show only one example where an agent is incentivised to misreport its cost function. Following the same notation as above, let $c_t(\theta)$ and $\widehat{c}(\theta)$ denote an agent's true and reported cost functions respectively, while $c_s(\theta)$ denotes the cost function used to scale the scoring function and θ' is the point where two cost functions (suppose $c_s(\theta)$ and $c_t(\theta)$) intersect. In this context, we intend to show that an agent can do better by misreporting and losing, rather than by reporting truthfully and winning. In more detail, since $c_s(\theta)$ and $c_t(\theta)$ overlap at θ', $c_s(\theta) < c_t(\theta)$, for every $\theta > \theta'$. Therefore, according to Lemma 1, the expected utility will be strictly negative. If the agent misreports its cost function so it is not selected, its utility will be zero. Therefore, the mechanism is no longer incentive compatible with respect to reported cost functions.

Now, regarding the agent's reported precision, if at θ' $c_t(\theta)$ and $c_s(\theta)$ intersect then $c_t(\theta') > c_s(\theta')$ and therefore $\overline{U}(\theta') < 0$. Given that this agent is the cheapest one, at least for $\theta \leq \theta'$ it is in its best interest to report a precision lower that θ' even if it makes an estimate with precision greater than θ', in order to maintain positive utility. That is, $\widehat{\theta} < \theta'$, while $\theta > \theta'$. Therefore, the mechanism is no longer incentive compatible with respect to reported precision. $\square$

Theorem 2. *The mechanism is incentive compatible regarding the agent's reported forecast and precision in the second stage.*

Proof. The proof for this theorem follows directly from the definition of the strictly proper scoring rules (see Section 3). □

Theorem 3. *The two-stage mechanism is individually rational.*

Proof. Having shown in Theorem 1 that the true reporting of cost functions in the first stage is a weakly dominant strategy, we only have to examine whether the selected agent is incentivised to participate into the the second stage of the mechanism and report its estimate and its precision to the centre. Since agents that do not win in the first stage receive zero utility, we only consider the case of the selected agent. For that agent, its true cost function is less than or equal to the cost function used for the scaling of the expected score (i.e. $c_t(\theta) \le c_s(\theta)$). Given that the selected agent's expected utility, $\overline{U}(\theta)$, is: $\frac{c_s'(\theta_0)}{\overline{S}'(\theta_0)}(\overline{S}(\theta) - \overline{S}(\theta_0)) + c_s(\theta_0) - c_t(\theta)$ (equation 16), it follows that $\overline{U}(\theta_0) = c_s(\theta_0) - c_t(\theta_0) \ge 0$. In Lemma 2, we have shown that the agent will produce an estimate $\theta^* > \theta_0$. By definition, $\overline{U}(\theta^*) \ge \overline{U}(\theta_0)$, and thus, $\overline{U}(\theta^*) \ge 0$. □

Theorem 4. *For the agent selected in the first stage of the mechanism, it is optimal to produce an estimate with a precision equal to or higher than the precision required by the centre, i.e., $\theta^* \ge \theta_0$.*

Proof. This proof follows directly from Lemma 2 where we show that there is an optimal precision, θ^*, such that $\theta^* \ge \theta_0$ if $c_t(\theta)$ and $c_s(\theta)$ are convex functions with $c_t(\theta) < c_s(\theta)$, $c_t'(\theta) < c_s'(\theta)$ and $c_t(0) = c_s(0) = 0$. Given that the mechanism is incentive compatible in costs, then all these conditions hold, and thus, $\theta^* \ge \theta_0$. □

Note that these proofs indicate that the two stages of the mechanism are inextricably linked and cannot be considered in isolation of one another. Indeed, apparently small changes to the second stage of the mechanism can destroy the incentive compatibility property of the first stage. For example, it is important to note that our mechanism is more precisely known as *interim* individually rational (Mas-Colell et al., 1995), since the utility is positive in expectation. In any specific instance, the payment could actually be negative if the prediction turns out to be far from the actual outcome. An alternative choice for the second stage of the mechanism would be to set β such that the payments are always positive, thus making the mechanism ex-post individually rational. However, this would then violate the incentive-compatibility property since the agents could then receive positive pay-offs by misreporting their cost functions. Likewise, it might be tempting to imagine that the centre could use the revealed costs of the agents in order to request a lower precision, confident in the knowledge that the selected agent will actually produce an estimate of the required precision. However, by effectively using the lowest revealed cost within the payment rule in this way, the incentive-compatibility property of the mechanism would again be destroyed.

4.3. Numerical Simulations

Having proven the economic properties of the mechanism in the general case with any convex cost function, we now consider a specific scenario in which costs are linear functions, given by

$c_i(\theta) = c_i\theta$, where the value of c_i is drawn from a uniform distribution $c_i \sim \mathcal{U}(1,2)$ and $\theta_0 = 1$. Note that while we can derive analytical expressions for the expected payment that the centre will make in any specific instance of this case (i.e. when the lowest and second-lowest cost functions are known), we cannot do so in the general case where these cost functions are drawn from some distribution since this requires that we integrate over the joint probability distribution of the first and second agent's costs, and the mean and variance of the estimate that the agent with the lowest cost actually generates. Performing this integration numerically is also problematic as expressing this joint probability distributions becomes increasingly complex as the mechanisms described in this paper become more complex. Thus, instead we effectively draw independent samples from this joint probability distribution by actually simulating the processes of the mechanism (i.e. the revelation of the agents' costs, the selection of the agent with the lowest cost in the first stage of the mechanism, the choice of precision that this agent makes given the scoring rule presented to it, its generation of an estimate, and finally the score that it receives after the true unknown value is revealed). To this end, for a range from 2 to 20 agents participating in the first stage, we simulate the mechanism 10^6 times and, for each iteration, record the payment made to the agent that provided the forecast and the precision of this forecast. Due to the number of iterations that we perform, the standard error in the mean values plotted are much smaller than the symbol size shown in the plot, and thus for clarity, we omit them.

The payment the agent expects to derive, $\overline{P}$, and its actual precision, θ^*, for every value of $N \in [2,20]$ are shown in Figure 1. As expected, as the number of agents increases, the mean payment, shown in Figure 1a, decreases toward the lower limit of the uniform distribution from which the costs were drawn. Furthermore, note that there is a fixed ordering over the entire range, with the payment resulting from the quadratic scoring rule being the highest, and that of the logarithmic scoring rule being the lowest. The reason for this can be seen in Figure 1(b) where the precision of the forecasts that were actually made are shown. Note that the logarithmic scoring rule induces agents to produce forecasts closer to the required precision than both the spherical and the quadratic scoring rules.

Figure 1a also shows the mean of the lowest and second lowest costs evaluated at the required precision θ_0 (denoted by $\overline{c}_1\theta_0$ and $\overline{c}_2\theta_0$ respectively). The first cost represents the minimum payment that could have been made if the costs of the agents were known to the centre. The second represents the payment that would have been made, had the agent produced a forecast of the required precision θ_0 rather than its own optimal precision θ^*. The gap between $\overline{c}_1\theta_0$ and $\overline{c}_2\theta_0$ is the extra amount that must be paid as a result of the costs being unknown and is the same regardless the scoring rule used. On the other hand, the gap between $\overline{c}_2\theta_0$ and the mean payment of any particular scoring rule, depends on the choice of the scoring rule as it represents the loss that the centre has to cover, as a result of the agent producing an estimate at its optimal precision, θ^* (rather than one at the minimum precision required, θ_0). The goal in selecting scoring rules is clearly to minimise this gap, and it can be seen that the logarithmic scoring rule is closest to achieving this goal.

We also derive the analytical expressions of the expected payment, $\overline{P}$, and the optimal precision, θ^*, as a function of the required precision, θ_0, for a single run of the mechanism in this specific setting where cost functions are represented by linear functions, and the costs of the cheapest and second cheapest agents (denoted by c_1 and c_2) are known. These results are represented in the first two rows of Table 2, and show that the pattern observed in the empirical evaluation (where we effectively average over the distribution of the first and second lowest
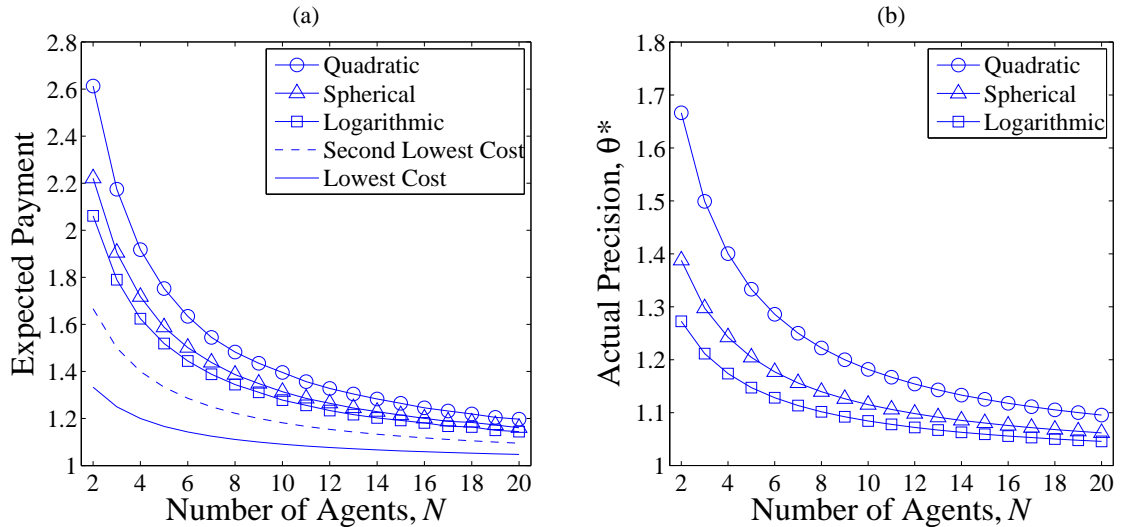
15

Figure 1: Selected agent's expected payment and optimal precision.

costs) is shown in the individual analytical results. That is, when the payment is based on the logarithmic scoring rule, the agent's expected payment is less than the other two scoring rules, and the precision that the agent actually reports is closest to that requested.

Furthermore, in Figure 2 we also apply the parametric scoring rule to the case where $N = 10$, and compare it to the three fixed scoring rules. Note that in the case of the parametric scoring rule, as $k \to 1$, the expected payment of the centre, and its variance, is asymptotically equal to that of the logarithmic scoring rule. Likewise, for $k = 2$, the parametric scoring rule is exactly the quadratic scoring rule (since it takes the same mathematical form). For $k = 1.5$, the expected payment of the parametric scoring rule is equal to that of the spherical rule, but the variance in the payments is not. From these plots, it would appear that the logarithmic scoring rule would be the optimum choice for the centre since it will minimise the amount that must be paid to the agents. It also displays the minimum variance in this payment which is an important criteria since it reflects the uncertainty in the payment that the agent is expecting to receive. However, further analysis in the next section indicates that the parametric scoring rule has a significant advantage over the logarithmic; that is, the existence of a finite lower bound on the payment.

4.4. Analysis of Payment Lower Bound

In more detail, row 3 of Table 2 shows the analytically calculated lower bound of the scaled payment, P^-, based on the principle that the lower bound is derived from the scoring rule $S(x_0; \widehat{x}, \widehat{\theta})$, when the value of the probability density function describing the actual outcome is 0 (i.e. $\mathcal{N}(x_0; \widehat{x}, 1/\widehat{\theta}) = 0$). Note that the logarithmic scoring rule does not have a finite lower bound. Thus, if the agent's estimate is far from the actual outcome, then a payment based on the logarithmic scoring rule will go to $-\infty$, and the agent will actually be required to pay an unbounded penalty to the centre. Likewise, in the limit as $k \to 1$, the payment based on the

Table 2: Analytical calculation of the expected payment, optimal precision and lower bound on the payment for quadratic, spherical, logarithmic and parametric scoring rules with linear cost functions for an instance of the mechanism.

SR:	Quadratic	Spherical	Logarithmic	Parametric
$\overline{P}(\theta_0)$	$c_2\theta_0\left[2\frac{c_2}{c_1}-1\right]$	$c_2\theta_0\left[4\left(\frac{c_2}{c_1}\right)^{\frac{1}{3}}-3\right]$	$c_2\theta_0\left[1+\log\left(\frac{c_2}{c_1}\right)\right]$	$\frac{c_2\theta_0}{k-1}\left[2\left(\frac{c_2}{c_1}\right)^{\frac{k-1}{2}}+k-3\right]$
θ^*	$\left(\frac{c_2}{c_1}\right)^2\theta_0$	$\left(\frac{c_2}{c_1}\right)^{\frac{4}{3}}\theta_0$	$\left(\frac{c_2}{c_1}\right)\theta_0$	$\left(\frac{c_2}{c_1}\right)^{\frac{2}{3-k}}\theta_0$
P^-	$-c_2\theta_0\left[1+2\frac{c_2}{c_1}\right]$	$-3c_2\theta_0$	$-\infty$	$c_2\theta_0\left[1-\frac{2}{k-1}-2\left(\frac{c_2}{c_1}\right)^{\frac{k-1}{3-k}}\right]$

Costs are given by linear functions, $c(\theta)=c\theta$, and c_1 and c_2 are the lowest and second lowest costs.

scaled parametric scoring rule also has no finite lower bound[8]. However, for values of $k \neq 1$, the parametric scoring rule is bounded, and the appropriate choice of the parameter, k, allows the overall performance of the scoring rule (in terms of the expected total payment of the centre and the variance in this payment) to be traded-off against the value of this bound.

In Figure 3 we plot the lower bound of the payments based on the quadratic, spherical and parametric scoring rules (we omit the logarithmic scoring rule as it goes to $-\infty$); noting that the lower bound occurs when $c_1 = 1$ and $c_2 = 2$ (the lower and upper support of the cost function distribution). Note that for some values of k, the lower bound of the parametric scoring rule is greater than that of the quadratic rule, but it is always less than that of the spherical rule. Based on this result, we select k to be equal to 1.2 in our future experiments. This value results in an expected payment and variance that is close to the logarithmic scoring rule, whilst not penalising the agent excessively in the worst case. The choice of parameter value here is is somewhat arbitrary, and in practice, it will depend on the details of the particular application domain.

4.5. Discussion

In this section we introduced a two-stage mechanism based on strictly proper scoring rules that motivates self-interested rational agents to make a costly forecast of a specified precision and report it truthfully to a centre. The mechanism was applied in a setting in which a centre is faced with multiple agents but has no knowledge about the costs involved in the generation of the probabilistic estimates. We first proved that the mechanism was incentive compatible and individually rational. Then we empirically evaluated the mechanism by comparing the quadratic, spherical, logarithmic and parametric scoring rules, and showed that the logarithmic and the parametric (for $k \to 1$) rules minimise the centre's expected payment, the variance in this payment, and the selected agent's optimal precision. However, given that payments derived from the logarithmic scoring rule payment have no finite lower bound, the parametric scoring rule is

---

[8]In this case, this is due to the scaling parameters being unbounded. The score has a finite lower bound for all values of k.
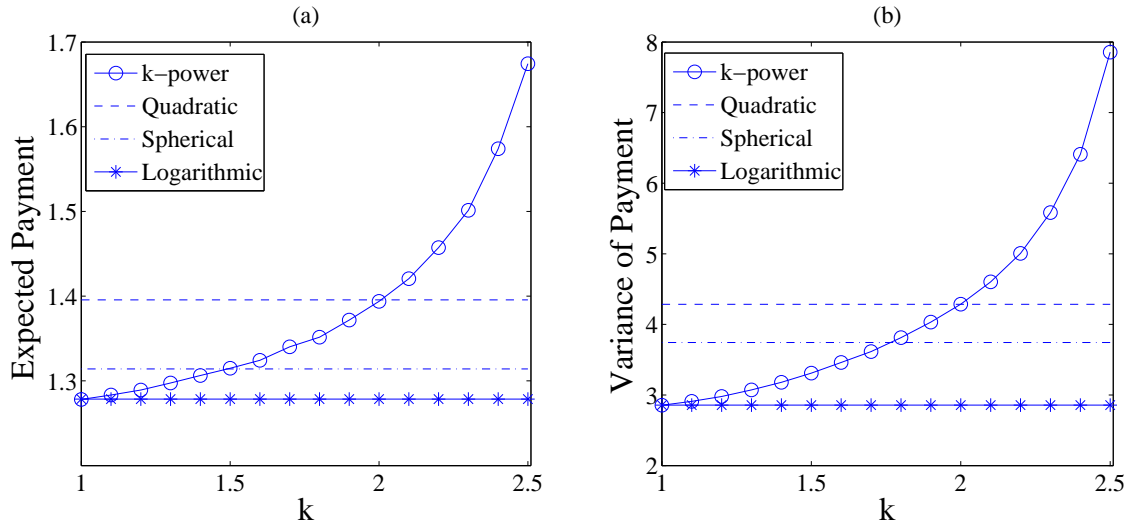
Figure 2: The mean and variance of the centre's payment.

a more appropriate choice for a centre that does not want to severely punish agents that inadvertently provide inaccurate observations. Hence, we will be using it for the numerical evaluations of the mechanisms we develop in the remainder of this paper.

5. A Mechanism for Dealing with Multiple Agents that have a Limited Degree of Precision

In the previous section we considered the case where any single agent is able to generate an estimate of the required precision. Now, as already mentioned in Section 1, this may not always be the case in situations where agents have limited resources with which to produce these estimates, and thus, the centre may have to procure estimates from multiple agents and fuse them together in order to achieve a sufficiently high precision. To this end, we revise the mechanism in the previous section by relaxing the assumption that any single agent is capable of producing the required estimate. In doing so, we propose a parametrised iterative mechanism (Mechanism 2), which is similar to the previous mechanism (i.e. two stages, first stage to elicit costs, second to calculate payments), but that uses a significantly different process to elicit those costs and calculate the payments. In more detail, in the first stage the centre pre-selects M from N agents through a series of selection steps which elicits their costs. In the second stage, it elicits the pre-selected agents' probabilistic estimates, after sequentially approaching them in a random order. As with the previous mechanism, we formally prove that this novel mechanism is incentive compatible regarding the costs, maximum precisions and estimates, and that it is individually rational. Finally, we introduce a family of processes by which the centre may pre-select M from N agents and show both empirically and analytically that the centre will minimise its expected payments by forming a single group of agents in the first stage of the mechanism.

5.1. Eliciting Information from Multiple Sources

As described in Section 2, we consider the same model as in the previous section, however we additionally assume that there is a limit in the maximum precisions of the agents' estimates,
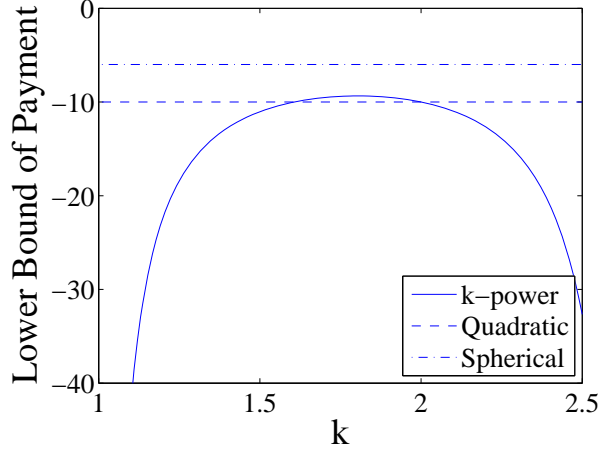
Figure 3: Calculated lower bound of the payment, P^-, for linear functions, given by $c_i(\theta) = c_i\theta$, where the value of c_i is drawn from a uniform distribution $c_i \sim \mathcal{U}(1,2)$ and $\theta_0 = 1$.

denoted by θ_i^c. Thus, agents can produce estimates of any precision up to and including this maximum value (i.e. $0 \leq \theta_i \leq \theta_i^c$). Given this limit, the centre may not be able to rely on a single agent to achieve its required precision, and may have to combine estimates from multiple agents, in order to achieve the desired degree of accuracy, and thus, must fuse k conditionally independent and unbiased probabilistic estimates, $\{\widehat{x}_1, \ldots, \widehat{x}_k\}$ of possibly different precisions $\{\widehat{\theta}_1, \ldots, \widehat{\theta}_k\}$, into a single estimate with mean $\bar{x}$ and precision $\bar{\theta}$. To do so, the centre uses the standard result (see DeGroot and Schervish (2002)) for fusing independent Gaussian distributions such that:

$$\bar{x}\bar{\theta} = \sum_{i=1}^{k} \widehat{x}_i\widehat{\theta}_i \qquad \text{and} \qquad \bar{\theta} = \sum_{i=1}^{k} \widehat{\theta}_i \qquad (18)$$

By fusing the agents' precisions, the centre manages to acquire an estimate of higher precision than the precision of any of the individual agents. Indeed, it can be seen that $\bar{\theta} \geq \theta_i$ for any agent i. Note that for this fusion to be appropriate, agents must be incentivised to truthfully report both the means and precisions of their estimates. Now, given this model, the challenge is to design a mechanism in which the centre will be able to initially identify those agents that can provide their estimates at the lowest cost, then motivate these agents to truthfully report their maximum precisions and finally generate and truthfully report their estimates with precisions equal to their reported maximum precisions.

5.2. The Mechanism

In Mechanism 2 we extend the mechanism discussed in the previous section by relaxing the assumption that the centre can select a single agent that can provide the estimate at the required precisions. The centre can now elicit estimates from multiple agents which have limited precisions in the estimates they can provide. In order to address this issue, the centre in the first stage, iteratively pre-selects M of the N available agents based on their reported costs. There are a number of ways in which this may be done; most generally, by dividing all the available agents, N, into groups of $n \leq N$ agents and then by sequentially asking the agents of each group

to reveal their costs. The centre then selects the *m* cheapest agents, with $m < n$. We shall shortly show that one combination of *n* and *m* dominates all others.

In the second stage, the centre then sequentially asks the *M* pre-selected agents to reveal their private maximum precision, in a random order that is independent of their reported costs, until it achieves its required precision, θ_0, at which point it discards the remaining pre-selected agents. Then, those that are not discarded are provided with a payment rule that incentivises them to generate estimates at their reported maximum precisions and to truthfully report these estimates to the centre.

We now proceed to prove that this mechanism leads the agents to truthfully reveal their costs in the first stage (so that those which can produce the estimate at the lowest cost can be identified), and that the *M* pre-selected agents are incentivised to truthfully report their maximum precisions to the centre and subsequently make and truthfully report estimates of these precisions in the second stage. These properties are not obvious, and as in the single agent section, they depend rather subtly on the details of the mechanism. For example, we note that if after asking all *M* agents for their maximum precisions, the centre does not achieve its required precision, the mechanism must proceed to the payment phase (step 5 in second stage). That is, the centre must commit to paying all pre-selected agents for their estimates at their reported maximum precisions, even if it does not acquire its required precision. Failure to observe this policy would lead agents to over-report their maximum precision, in order that some payment was received, and thus, the mechanism would no longer be incentive compatible in terms of maximum precisions.

Furthermore, note that in step 2 of the first stage, the centre chooses the *m* agents with the lowest reported costs, and discards the remaining $n - m$ agents. If these agents were not discarded, but were placed back into the pool of available agents, then the mechanism would no longer be incentive compatible in terms of costs; agents would have an incentive to over-report their costs, such that when they are eventually pre-selected, their payment rule will be calculated using a higher cost. Finally, in step 2 of the second stage, the centre must randomly ask the pre-selected agents to report their maximum precisions using an ordering which is independent of their reported costs. Failing to do so will undermine incentive compatibility in terms of costs of the first stage of the mechanism, thereby illustrating how the two stages interact. Even in the case where only one agent participates in the second stage, as a result of the available agents being two ($N = 2$) or the number of the pre-selected agents being set to one by the centre ($M = 1$), the incentive compatibility is maintained. In this case, the agent with the higher reported cost will not be asked for its precision in the second stage since, since the $N - M$ agents are discarded in the first stage.

5.3. Economic Properties of the Mechanism

Having detailed the mechanism, we now identify and provide its economic properties. Specifically, we show that:

1. The mechanism is incentive compatible with respect to the pre-selected agents' reported maximum precisions and reported estimates.
2. The mechanism is incentive compatible with respect to the agents' reported costs.
3. The mechanism is individually rational.

Mechanism 2 The mechanism for dealing with multiple agents that can provide estimates of a limited precision:

1. First Stage

1.1 The centre selects $n \geq 2$ agents from the available N and asks them to report their cost functions $\widehat{c}_i(\theta)$ with $i \in \{1, \ldots, n\}$.

1.2 The centre selects the m, $(1 \leq m < n)$, agents with the lowest costs, associates the $(m+1)^{th}$ cost with these agents and discards the remaining $n - m$ agents.

1.3 The centre repeats the above two steps until it has asked all N agents to report their cost functions. Note that when N is not exactly divisible by n and we have a single remainder, it is discarded. Otherwise in the final round the centre modifies n and m such that $n = N \bmod n$ and $m = \min(m, n - 1)$.

1.4 We denote the total number of the agents pre-selected in this stage as M and note that its value depends on N, n and m.

2. Second Stage

2.1 The centre sets its required precision θ_r equal to θ_0.

2.2 The centre randomly selects one of the pre-selected agents and asks it to report its maximum precision $\widehat{\theta}_j^c$, with $j \in \{1, \ldots, M\}$.

2.3 The centre asks the agent j to produce an estimate of this precision and presents this agent with a scaled strictly proper scoring rule. The scaling parameters α and β are determined using equations 9 and 10. However, within these expressions $\widehat{\theta}_j^c$ is used instead of θ_0, and c_s (the cost associated with this agent in the preceding stage – $(m+1)^{th}$ cost in the group from which it was selected) is used instead of c_t. Hence, the scaling parameters are given by:

$$\alpha_j = \frac{c_s'(\widehat{\theta}_j^c)}{\overline{S}'(\widehat{\theta}_j^c)} \quad \text{and} \quad \beta_j = c_s(\widehat{\theta}_j^c) - \frac{c_s'(\widehat{\theta}_j^c)}{\overline{S}'(\widehat{\theta}_j^c)} \overline{S}(\widehat{\theta}_j^c) \tag{19}$$

2.4 The centre sets $\theta_r = \theta_r - \min(\theta_r, \widehat{\theta}_j^c)$ and if $\theta_r > 0$ it repeats step two of the second stage.

2.5 The agents that were asked to do so, produce an estimate x_j with precision θ_j and report $\widehat{x}_j$ and $\widehat{\theta}_j$ to the centre[9], which after observing the actual outcome, x_0, issues the following payments:

$$P_j(x_0; \widehat{x}_j, \widehat{\theta}_j) = \alpha_j S_j(x_0; \widehat{x}_j, \widehat{\theta}_j) + \beta_j \tag{20}$$

with α and β being already determined in step two of the second stage.

Theorem 5. *The mechanism is incentive compatible with respect to the pre-selected agents' reported maximum precisions and reported estimates.*

Proof. Given the mechanism described above, when the agent reports its estimate, it must do so with the precision that it claimed was its maximum. Thus, $\widehat{\theta} = \widehat{\theta}^c$. Now, given the scaling of the scoring rules described in step 2 in the second stage of the mechanism, the expected utility of the agent, if it reports its maximum precision as $\widehat{\theta}^c$, and subsequently produces an estimate of precision θ, which it reports with precision $\widehat{\theta}^c$, is denoted by $\overline{U}(\theta, \widehat{\theta}^c)$, and is given by:

$$\overline{U}(\theta, \widehat{\theta}^c) = \frac{c_s'(\widehat{\theta}^c)}{\overline{S}'(\widehat{\theta}^c)}\left(\overline{S}(\theta, \widehat{\theta}^c) - \overline{S}(\widehat{\theta}^c)\right) + c_s(\widehat{\theta}^c) - c_t(\theta) \tag{21}$$

where $\overline{S}(\theta, \widehat{\theta}^c)$ is the agent's expected score for producing an estimate of precision θ and reporting its precision as $\widehat{\theta}^c$. Furthermore, $\overline{S}(\widehat{\theta}^c)$ is the agent's expected score for producing and truthfully reporting an estimate of precision $\widehat{\theta}^c$, $c_t(.)$ is the true cost function of the agent, and $c_s(.)$ is the cost function used to produce the scoring rule (i.e. the $(m+1)$th lowest revealed cost in the group from which the agent was pre-selected). Taking the first derivative of this expression with respect to $\widehat{\theta}^c$ gives:

$$\frac{d\overline{U}(\theta, \widehat{\theta}^c)}{d\widehat{\theta}^c} = \frac{d}{d\widehat{\theta}^c}\left(\frac{c_s'(\widehat{\theta}^c)}{\overline{S}'(\widehat{\theta}^c)}\right)\left(\overline{S}(\theta, \widehat{\theta}^c) - \overline{S}(\widehat{\theta}^c)\right) + \frac{c_s'(\widehat{\theta}^c)}{\overline{S}'(\widehat{\theta}^c)}\overline{S}'(\theta, \widehat{\theta}^c) \tag{22}$$

Now, since S is a strictly proper scoring rule, then $\overline{S}(\theta, \widehat{\theta}^c) = \overline{S}(\widehat{\theta}^c)$ and $\overline{S}'(\theta, \widehat{\theta}^c) = 0$ when $\theta = \widehat{\theta}^c$. Hence:

$$\left.\frac{d\overline{U}(\theta, \widehat{\theta}^c)}{d\widehat{\theta}^c}\right|_{\widehat{\theta}^c = \theta} = 0 \tag{23}$$

and thus, the utility of the agent is maximised when it reveals as its maximum precision, the precision of the estimate that it subsequently produces[10].

We now show that it will actually produce an estimate of precision equal to its reported maximum precision. To this end, we note that when $\widehat{\theta}^c = \theta$, the expected utility of the agent is given by:

$$\overline{U}(\theta) = c_s(\theta) - c_t(\theta) \tag{25}$$

Since $c_s(.)$ and $c_t(.)$ do not cross or overlap, and $c_s'(\theta) > c_t'(\theta)$, then $\overline{U}(\theta)$ is a strictly increasing function. Thus the agent will maximise its expected utility by producing an estimate at its maximum precision, and thus, $\theta = \theta^c$, and hence, $\widehat{\theta}^c = \widehat{\theta} = \theta^c$, as required. $\qquad\square$

---

[9]Note that we could restrict agents to report their estimates with precision $\widehat{\theta}_j^c$. However, as we shall show in Section 5.3, under this mechanism the agents are automatically incentivise to report $\widehat{\theta}_j = \widehat{\theta}_j^c$ anyway.

[10]For completeness, we confirm that the second derivative is negative at $\theta = \widehat{\theta}^c$. To this end, the second derivative is given by:

$$\frac{d^2\overline{U}(\theta, \widehat{\theta}^c)}{d(\widehat{\theta}^c)^2}(\widehat{\theta}^c = \theta) = \frac{c_s'(\widehat{\theta}^c)}{\overline{S}'(\widehat{\theta}^c)}\overline{S}''(\theta, \widehat{\theta}^c) - c_s''(\widehat{\theta}^c) + \frac{c_s'(\widehat{\theta}^c)}{\overline{S}'(\widehat{\theta}^c)}\overline{S}''(\widehat{\theta}^c) \tag{24}$$

Now, the first term of equation 24 is negative because S is strictly proper, and this implies that $\overline{S}''(\theta, \widehat{\theta}^c)$ is negative at $\theta = \widehat{\theta}^c$. Furthermore, $c_s''(\widehat{\theta}^c)$ is positive, assuming convexity of the cost function, and $\overline{S}''(\widehat{\theta}^c)$ is negative assuming concavity of the scoring rule. Hence, the second derivative is negative at $\widehat{\theta}^c = \theta$.

Theorem 6. *The mechanism is incentive compatible with respect to the agents' reported costs.*

Proof. We prove this by contradiction and consider two cases depending on whether or not an agent is pre-selected in the first stage of the mechanism as a result of its misreporting. Let $c_t(.)$ and $\widehat{c}(.)$ denote an agents' true and reported cost functions respectively. Furthermore, let $c_s(.)$ denote the cost function used to scale the scoring rule if that agent is among the m agents with the lowest reported costs in its group of n agents in the first stage of the mechanism (i.e. $c_s(.)$ is the $(m+1)^{th}$ cost of that group).

First, suppose that the agent's misreporting does not affect whether it is pre-selected or not. In this case, had the agent been pre-selected, its payment would have been based on the $(m+1)^{th}$ cost of its group and therefore independent of its own report. Conversely, had the agent not been pre-selected, it would have received zero utility, since the remaining $n-m$ agents, of a group of initially n agents, that are not pre-selected are discarded. Hence, there is no incentive to misreport.

Second, suppose that the agent's misreporting affects whether that agent is pre-selected or not. There are now two cases: (1) the agent is pre-selected by misreporting but would have not been if it was truthful, (i.e. $c_t(\widehat{\theta}^c) > c_s(\widehat{\theta}^c)$ and $\widehat{c}(\widehat{\theta}^c) < c_s(\widehat{\theta}^c)$), and (2) the agent is not pre-selected by misreporting but would have been if truthful (i.e. $c_t(\widehat{\theta}^c) < c_s(\widehat{\theta}^c)$ and $\widehat{c}(\widehat{\theta}^c) > c_s(\widehat{\theta}^c)$).

Case (1). Since the true cost $c_t(\widehat{\theta}^c) > c_s(\widehat{\theta}^c)$, it follows directly from Theorem 5 that the expected utility $\overline{U}(\theta) = c_s(\theta) - c_t(\theta)$ is strictly negative, irrespective of θ. Therefore, the agent could do strictly better by reporting truthfully in which case the expected utility is zero.

Case (2). In this case the agent would have been pre-selected if it was truthful, but now receives a utility of zero since it has not been pre-selected due to its misreporting. To show that this type of misreporting is suboptimal, we need to show that, when $c_t(\widehat{\theta}^c) < c_s(\widehat{\theta}^c)$, an agent benefits from being pre-selected, since it may then be asked to generate an estimate at its reported maximum precision, $\widehat{\theta}^c$. It follows directly from Theorem 5 that $\overline{U}(\widehat{\theta}^c) = c_s(\widehat{\theta}^c) - c_t(\widehat{\theta}^c) > 0$ when $c_t(\widehat{\theta}^c) < c_s(\widehat{\theta}^c)$, and therefore there is no incentive for an agent that would have been pre-selected to misreport its cost function. $\square$

Theorem 7. *The mechanism is interim individually rational.*

Proof. Due to Theorem 6, we can assume that all agents, and consequently those pre-selected, will report their true cost functions, and therefore $c_t(\theta) \leq c_s(\theta)$. In Theorem 5, we show that the expected utility $\overline{U}(\theta) = c_s(\theta) - c_t(\theta)$ is strictly non-negative, irrespective of θ. Therefore, the expected utility of a pre-selected agent that generates an estimate of precision equal to its reported maximum precision $\widehat{\theta}^c$, is strictly non-negative (i.e. $\overline{U}(\widehat{\theta}^c) \geq 0$), and hence the mechanism is *interim* individually rational. $\square$

5.4. Numerical Simulations

Having proven the economic properties of the mechanism, we present empirical results for a specific scenario in order to explore the effect that the parameters n and m have on the centre's total payments, and on the probability of achieving its required precision. In more detail, as before, the cost functions are represented by linear functions, given by $c_i(\theta) = c_i\theta$, where c_i are independently drawn from a uniform distribution $c_i \sim \mathcal{U}(1,2)$. The maximum precisions of the selected agents, θ_i^c, is independently drawn from another uniform distribution $\theta_i^c \sim \mathcal{U}(0,1)$ and finally the centre's required precision, θ_0, is equal to 1.7 in order to generate representative
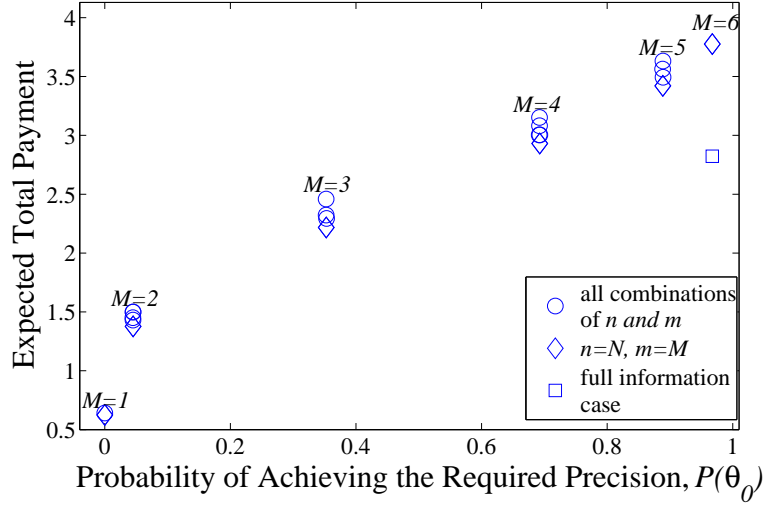
Figure 4: Centre's probability of achieving the required precision and the mean total payment it has to issue.

results whereby the probability of achieving the required precision, $P(\theta_0)$, covers a broad range of values in $[0, 1]$. Finally, we restrict our analysis to the use of the parametric scoring rule for $k = 1.2$ as we have shown in Section 4.4, that among the common rules and for various values of the parameter k, this rule is a good choice for a centre intending to issue low payments with low variance, that still remain bounded.

Given this, and for $N = 7$, we explore all possible combinations of n and m given the constraints that $2 \leq n < N$ and $1 \leq m < n$. For each combination, we simulate the mechanisms 10^7 times and for each iteration we record whether the centre was successful in acquiring an estimate at its required precision, and the sum of all the payments it issued to those agents that were asked to produce an estimate. In Figure 4 we plot, for each possible combination of n and m, the probability of acquiring the required precision against the total payment made by the centre. We note that again the standard error of the mean values are much smaller than the plotted symbols, and thus, for clarity we omit it.

With regard to this figure, the squares indicates the case where the centre has full information of the agents' costs, and therefore represents an upper bound for the mechanism. It results in significantly lower total payments to the agents since the centre is able to select those agents with the lowest costs to generate the estimates and it use payment rules that are scaled using the known costs of these agents. The circles depict the results for all possible combinations of n and m, except where $n = N$ and $m = M$, which is indicated by a diamond (the reason for this will become clear shortly). We first note that many possible combinations of n and m give rise to the same value of $P(\theta_0)$, and thus the family of possible pre-selection methods fall into 6 distinct columns. This is because this probability depends only on the number of agents that are pre-selected (denoted by M) and many of these combinations result in the same number of agents being pre-selected (e.g. if $N = 7$, both $n = 4$, $m = 2$ and $n = 5$, $m = 3$ result in $M = 4$). Second,

note that for each possible value of M, the case where $n = N$ and $m = M$ dominates all other combinations of n and m (i.e. it results in the lowest mean total payment). This case corresponds to a single selection stage in which M agents are pre-selected directly from the original N in a single step. We more formally analyse these two observations in the following section.

5.5. Analysis of Pre-Selection Schemes

With regard to the observation above that the probability of achieving θ_0 is dependent on the number of agents pre-selected, we can see that this is so since within our mechanism the maximum precisions of the pre-selected agents are independent of their costs. In the numerical simulations described above we have M independent and uniformly distributed random variables $\theta_i^c \sim \mathcal{U}(0,1)$ which denote the agents' maximum precisions. If we denote the sum of these as Θ such that $\Theta = \sum_{i=1}^{M} \theta_i^c$, then its cumulative probability distribution allows us to calculate $P(\Theta \geq \theta_0)$ as follows:

$$P(\Theta \geq \theta_0) = \begin{cases} 1 - \frac{1}{M!} \sum_{i=0}^{\lfloor \theta_0 \rfloor} (-1)^i \begin{pmatrix} M \\ i \end{pmatrix} (\theta_0 - i)^M & 0 \leq \theta_0 \leq M \\ 0 & \theta_0 > M \end{cases} \tag{26}$$

Although not immediately obvious from its analytical form, this is an increasing function in M as demonstrated in the numerical simulations.

The second observation is that a single selection stage in which M agents are pre-selected directly from the original N in a single step dominates all other selection schemes. This is perhaps more surprising and for this reason we provide a formal proof for the case where costs are linear functions of precision below.

Theorem 8. *In a setting with linear cost functions, where agents' costs and maximum precisions are independently drawn from uniform distributions, for a given probability of achieving θ_0, the centre minimises its expected total payment when $n = N$ and $m = M$.*

Proof. Given the mechanism and setting described above, we first note that when the costs of the agents are represented by linear functions, then $c_i(\theta) = c_i\theta$, and hence, $c_i'(\theta) = c_i$. Using this result within the scaling parameters of the payment rule described in Step 2.4, gives the result:

$$\alpha_j = \frac{c_s}{\overline{S}'(\widehat{\theta}_j^c)} \quad \text{and} \quad \beta_j = c_s\widehat{\theta}_j^c - \frac{c_s}{\overline{S}'(\widehat{\theta}_j^c)}\overline{S}(\widehat{\theta}_j^c) \tag{27}$$

Thus, both α and β are proportional to c_s, and hence the payment to any agent is also proportional to the cost used in the calculation of the scaling parameters. Secondly, we note that due to the random selection of agents within the second stage of the mechanism, the precision of the estimate generated by any agent is independent of the cost used to generate its payment rule. Hence, the expected total payment to the agents is proportional to the mean cost used to generate their payment rules.

Now, the costs used to generate the payment rule of any agent is the $(m+1)^{th}$ lowest reported cost when m agents are pre-selected from n. Thus, in order to show that setting $n = N$ and $m = M$ minimises the expected total payment of the centre, we must simply show that the expected value of the $(M+1)^{th}$ cost when pre-selecting M agents from N, is lower than any other combination.

25

To do so, we note that if the costs of the agents are i.i.d. from the standard uniform distribution[11], and the agents report truthfully (as they are incentivised to do), then the density function that describes the $(m+1)^{th}$ cost, denoted by $C_{m+1:n}$, is given by:

$$c_{m+1:n}(u) = \frac{n!}{m!(n-m-1)!}u^m(1-u)^{n-m-1}, \quad 0 \le u \le 1 \tag{28}$$

and Arnold et al. (2008) show that $c_{m+1:n}(u) \sim B(m+1, n-m)$ and therefore the mean of this distribution is simply $\frac{m+1}{n+1}$. Thus, we now prove that the $(M+1)^{th}$ cost when pre-selecting all M agents directly from N is less than the expected cost that results from first pre-selecting m agents from n and then pre-selecting the remaining $M-m$ agents from $N-n$. Therefore, and given that $c_{M+1:N}(u) \sim B(M+1, N-M)$ and $c_{M-m+1:N-n}(u) \sim B(M-m+1, N-n-M+m)$, we must prove the inequality:

$$\left(\frac{M+1}{N+1}\right) \le \frac{m}{M}\left(\frac{m+1}{n+1}\right) + \frac{M-m}{M}\left(\frac{M-m+1}{N-n+1}\right) \tag{29}$$

subject to the constraints that $M < N$, $m < n$ and $N-n > M-m$, and we note that if it holds in this case, then it holds for all possible combinations of n and m.

A first step towards the proof of equation 29, is performing the following substitutions:

$$a = m, \quad b = M-m, \quad c = n, \quad d = N-n$$

such that equation 29 now takes the following form:

$$\frac{(a+b)(a+b+1)}{c+d+1} \le \frac{a(a+1)}{c+1} + \frac{b(b+1)}{d+1} \tag{30}$$

with $a, b, c, d \ge 0$, $a < c$, and $b < d$.

Now, by multiplying all fractions in equation 30 to obtain a common denominator, $(c+1)(d+1)(c+d+1)$, and noting that this denominator is positive, translates equation 30 into the following condition:

$$
\begin{aligned}
(a+b)(a+b+1)(c+1)(d+1) & \\
- a(a+1)(c+d+1)(d+1) & \\
- b(b+1)(c+d+1)(c+1) & \le 0
\end{aligned}
\tag{31}
$$

We can rearrange this expression into the form:

$$F_1(a,b,c,d) + F_2(a,b,c,d) + F_3(a,b,c,d) \le 0 \tag{32}$$

where:

$$F_1(a,b,c,d) = -(d \cdot a - b \cdot c)^2 \tag{33}$$
$$F_2(a,b,c,d) = -b(c-a)^2 - b^2(c-a) - a(d-b)^2 - a^2(d-b) \tag{34}$$
$$F_3(a,b,c,d) = a(b-d) + b(a-c) \tag{35}$$

Now, it is easy to verify that F_1, F_2, and F_3 are all negative given the initial constraints that $a, b, c, d \ge 0$, $a < c$ and $b < d$. Hence, it follows that equation 32 is negative. $\quad\square$

---

[11] For notational simplicity we shall assume that the costs are drawn from $\mathcal{U}(0,1)$, but we note that the proof is valid for a uniform distribution of any support.

5.6. Discussion

In this section, we extended our original two-stage mechanism by relaxing the assumption that a single agent can provide an estimate of infinite precision and have introduced limitations on the agents' maximum precisions. As a result of this, an agent might not be able to generate estimates of a sufficient precision to individually meet the centre's needs, hence leaving the centre no option but to combine multiple such estimates and fuse them into a more accurate one. In addition to that, the centre now needs to elicit the maximum precision from the agents. In order to address these challenges, in this extended mechanism a centre pre-selects M from the N available agents by eliciting their cost functions in the first stage. Then, in the second stage, it approaches some of these M agents and asks them to report their maximum precision and make a costly probabilistic estimate or forecast of that precision. We proved that this mechanism is incentive compatible and individually rational, and we empirically evaluated the mechanism for various values of the parameters m and n and showed that for a given probability, $P(\theta_0)$, the centre minimises its mean total payment if it pre-selects M agents directly from a single group of N agents. These results showed that while it is always preferable to set $n = N$ (i.e. no agents should be excluded from the mechanism), the choice of the value of m is determined by the trade-off between the total payment made by the centre and the probability of it acquiring an estimate of its required precision. If the distributions of cost and maximum precisions are known, this can be evaluated prior to running the mechanism through simulation. However, if these distributions are unknown, setting $m = N - 1$ ensures that the probability of acquiring the required precision is maximised (but doing so will also incur the greatest expected payment).

6. A Mechanism Addressing the Centre's Lack of Access to Knowledge of the Outcome

So far, in both mechanisms we have considered, we have assumed that the centre has access to the actual outcome of the estimated event in order to calculate the payment to the agents in the second stage. In this section we remove this assumption. This means that in the second stage the centre must now calculate the score of each individual agent based upon the reports of the other agents. To do so, we have to modify the standard strictly proper scoring rules because, as we will show, to use them directly results in a scoring rule which no longer motivates agents to truthfully report the precision of their estimates (as as we have seen in the last section, failing to truthfully report the precision means that multiple independent estimates cannot be correctly fused). In particular, we show that our ensuing mechanism is incentive compatible with respect to maximum precisions and estimates, and that it is individually rational. We empirically evaluate our mechanism and compare it with the one we introduced in the previous section, in which the centre has access to the actual outcome and with a modified version of the 'peer prediction mechanism' proposed by Miller et al. (2007b). We show both analytically and empirically that for all the mechanisms we simulate, the agents expect to derive the same payment, which means the centre incurs no additional cost as a result of its lack of knowledge of the outcome. However, we identify a significant difference between the fusion and the peer prediction methods, by showing that in our mechanism the variance of the total payment issued to the selected agents by the centre (and thus the variance of the individual payments) is significantly lower than the total payment's variance in the peer prediction mechanism for $M > 2$ (although both are greater than the case where the outcome is known). This is important since this variance represents the uncertainty in the payments of both the centre and the individual agents (i.e. although the expected payments

may be calculated before hand, the actual payments will depend on the individual reports of the agents).

6.1. Evaluating Information without Knowledge of the Outcome

As described in Section 2, here we consider the same model as in the previous section, however we additionally assume that the centre will have no access to any knowledge of the outcome of the estimated event at the time at which is must make the payments to the agents. In doing so, the centre now has to rely solely on the estimates that it receives from the agents in order to scale the scoring rule, and is has two options is doing so: peer prediction or fusion.

In more detail, in the peer prediction mechanism (Miller et al., 2007b) the centre scores each agent's estimate directly against each of the other agents' estimates and then calculates its payment by averaging the resulting scaled scores. However, in the mechanism we introduce in this section, the centre uses the fused reported estimates of all the other agents (excluding from the fusion process the agent that is currently receiving the payment), and repeats this process for each individual agent. The exclusion of the agent that is receiving the payment from the fusion process is important, otherwise agents would have an incentive to exaggerate the precision of their estimates in order that the fused estimate corresponds more closely with their own report.

To this end, when there are $K \geq 2$ available agents, the centre calculates the payment to agent i, after fusing the agents' conditionally independent and unbiased probabilistic estimates with mean $\{\widehat{x}_1, .., \widehat{x}_{i-1}, \widehat{x}_{i+1}, .., \widehat{x}_k\}$ and precision $\{\widehat{\theta}_1, .., \widehat{\theta}_{i-1}, \widehat{\theta}_{i+1}, .., \widehat{\theta}_k\}$, into one estimate with mean $\overline{x}$ and precision $\overline{\theta}$ by using the standard result (see DeGroot and Schervish (2002)):

$$\overline{x}_{-i}\overline{\theta}_{-i} = \sum_{j \in K_{-i}} \widehat{x}_j\widehat{\theta}_j \qquad \text{and} \qquad \overline{\theta}_{-i} = \sum_{j \in K_{-i}} \widehat{\theta}_j \qquad (36)$$

where $K_{-i} = \{1, .., i-1, i+1, .., k\}$ is the set that contains all k agents besides agent i, which is the agent that is receiving the payment from the centre.

In this case, it is in a selected agent's best interest to consider its belief about the fused observations of all the other agents when reporting its precision. Now, this means that agent i's expected score $\overline{S}(\overline{x}; x_i, \theta_i)$ is maximised not at $\widehat{\theta}_i = \theta_i$ but at $\widehat{\theta}_i = \theta_i + \overline{\theta}_{-i}$. Indeed, if $N(\overline{x}_{-i}; x_i, 1/\theta_i + 1/\overline{\theta}_{-i})$ and $\mathcal{N}(\overline{x}_i; \widehat{x}_i, 1/\widehat{\theta}_i)$ are Gaussian distributions with mean and variance $(x_i, 1/\theta_i + 1/\overline{\theta}_{-i})$ and $(\widehat{x}_i, 1/\widehat{\theta}_i)$, which represent agent i's true and reported estimate's distributions, agent i's expected score, which is given by:

$$\overline{S}(\overline{x}_{-i}; \widehat{x}_i, \widehat{\theta}_i) = \int_{-\infty}^{\infty} \mathcal{N}(\overline{x}_{-i}; x_i, 1/\theta_i + 1/\overline{\theta}_{-i}) S(\overline{x}_{-i} | \mathcal{N}(\overline{x}_i; \widehat{x}_i, 1/\widehat{\theta}_i)) d\overline{x}_{-i} \qquad (37)$$

will be maximised at $\widehat{\theta}_i = \theta_i + \overline{\theta}_{-i}$, since for that value of the reported precision, $\widehat{\theta}_i$, the two distributions become identical.

Subsequently, an agent wanting to maximise its expected score (equation 37), will have to report $\theta_i + \overline{\theta}_{-i}$ instead of θ_i, which is impossible since it does not have access to other agents' precisions ($\overline{\theta}_{-i}$). However, given that the centre, when calculating the payments, has access to both θ_i and $\overline{\theta}_{-i}$, it can modify the strictly proper scoring rule such that the agent is only required to report θ_i but the payment is calculating using $\theta_i + \overline{\theta}_{-i}$. Therefore, within our mechanism we use this modified *modified strictly proper scoring rule*, $S(\overline{x}_{-i}; \widehat{x}_i, \widehat{\theta}_i + \overline{\theta}_{-i})$ and in Section 6.3 we prove that payments based on this modified scoring rule result (when scaled appropriately) in

---
Mechanism 3 The mechanism for dealing with the case where the centre does not have access to the actual outcome:

1. First Stage

    1   Identical to Stage 1 of Mechanism 2.

2. Second Stage

  2.1   and 2.2 are identical to steps 2.1 and 2.2 of Mechanism 2.

  2.3   The centre asks the agent j to produce an estimate of this precision and presents this agent with a scaled strictly proper scoring rule. Scaling parameters α_j and β_j are now based on the expected value of the *modified scoring rule*, $\overline{S}(\widehat{\theta}^c_j, \overline{\theta}_{-j})$, and its derivative. The parameters are given by:

$$\alpha_j = \frac{c'_s(\widehat{\theta}^c_j)}{\overline{S}'(\widehat{\theta}^c_j, \overline{\theta}_{-j})} \quad \text{and} \quad \beta_j = c_s(\widehat{\theta}^c_j) - \frac{c'_s(\widehat{\theta}^c_j)}{\overline{S}'(\widehat{\theta}^c_j, \overline{\theta}_{-j})}\overline{S}(\widehat{\theta}^c_j, \overline{\theta}_{-j}) \tag{38}$$

where, c_s is the cost associated with this agent in the first stage and $\overline{\theta}_{-j}$ is the fused precisions[11] of all the agents that are asked to produce an estimate except agent j and is defined in Equation 36.

  2.4   Identical to step 2.4 of Mechanism 2.

  2.5   The agents that were asked to do so, produce an estimate with mean x_j and precision θ_j, and report $\widehat{x}_j$ and $\widehat{\theta}_j$ to the centre, which in turn issues the following payment:

$$P_j(\overline{x}_{-j}; \widehat{x}_j, \widehat{\theta}_j + \overline{\theta}_{-j}) = \alpha_j S_j(\overline{x}_{-j}; \widehat{x}_j, \widehat{\theta}_j + \overline{\theta}_{-j}) + \beta_j \tag{39}$$

where $\overline{x}_{-j}$ and $\overline{\theta}_{-j}$ are the fused estimates and precisions of all the selected agents except agent j as defined in Equation 36.

---

truthful revelation of an agent's estimate being a Nash equilibrium, since the agent will maximise its expected payment if it reports truthfully, assuming that all other agents also report their true estimates.

6.2. The Mechanism

Having defined our *modified strictly proper scoring rules*, in this section, we describe how we extend the two-stage mechanism introduced in Section 5.2 for the setting in which the centre will not have access to the actual outcome when calculating the payments to the pre-selected agents (Mechanism 3). To this end, in the first stage the centre pre-selects M out of N agents and identifies their cost functions, while in the second stage it calculates their payments. Although the first stage is identical to the previous mechanism, in the second stage the centre fuses the pre-selected agents' reports into a single estimate (excluding the agent for which the payment is being calculated) and then uses an appropriately scaled modified scoring rule to calculate each agent's payment.

6.3. Economic Properties of the Mechanism

Having detailed the mechanism we now prove that truthful reporting of the preselected agents' maximum precisions and estimates is a Nash equilibrium. Note that truthful reporting of agents' cost functions in the first stage is still a dominant strategy and that this mechanism is also individually rational, like in all the previous mechanisms in this paper, and given that the proofs are identical to those of Theorems 6 and 7 respectively we refrain from re-writing them here. However, we now formally prove that truthful reporting of the agents' estimates is a Nash equilibrium when using our modified strictly proper scoring rule, and that truthful reporting of the agents' maximum precisions and estimates is a Nash equilibrium under our mechanism.

Theorem 9. *Truthful revelation of an agent's estimate and precision is a Nash Equilibrium under our modified strictly proper scoring rule.*

Proof. Given that an agent i's estimate is represented by the Gaussian distribution $\mathcal{N}(x_0; x, 1/\theta)$, under the modified strictly proper scoring rules, the score it expects to derive, is the following:

$$\overline{S}(\overline{x}_{-i}; \widehat{x}_i, \widehat{\theta}_i + \overline{\theta}_{-i}) = \int_{-\infty}^{\infty} \mathcal{N}(\overline{x}_{-i}; x_i, 1/\theta_i + 1/\overline{\theta}_{-i}) S(\overline{x}_{-i} | \mathcal{N}(\overline{x}_i; \widehat{x}_i, 1/\widehat{\theta}_i + 1/\overline{\theta}_{-i})) d\overline{x}_{-i} \quad (40)$$

where $\mathcal{N}(\overline{x}_{-i}; x_i, 1/\theta_i + 1/\overline{\theta}_{-i})$ and $\mathcal{N}(\overline{x}_i; \widehat{x}_i, 1/\widehat{\theta}_i + 1/\overline{\theta}_{-i})$ are Gaussian distributions with mean and variance $(x_i, 1/\theta_i + 1/\overline{\theta}_{-i})$ and $(\widehat{x}_i, 1/\widehat{\theta}_i + 1/\overline{\theta}_{-i})$ respectively, which will be denoted as Q and R respectively. Now equation 40 takes the following form:

$$\overline{S} = \int_{-\infty}^{\infty} Q(\overline{x}_{-i}) S(\overline{x}_{-i} | R(\overline{x}_{-i})) d\overline{x}_{-i} \quad (41)$$

Since S is a strictly scoring rule, as defined by Hendrickson and Buehler (1971) and Savage (1971), its expected value is maximised when $Q = R$. Furthermore, given the definition of Q and R, for $\widehat{x}_i = x_i$ and $\widehat{\theta}_i = \theta_i \Rightarrow Q = R$. Therefore, for $\widehat{x}_i = x_i$ and $\widehat{\theta}_i = \theta_i$, a payment based on the modified strictly scoring rule, $S(\overline{x}_{-i}; \widehat{x}_i, \widehat{\theta}_i + \overline{\theta}_{-i})$, is incentive compatible, since an agent will maximise its expected payment if it reports truthfully its estimate, assuming that all other agents also report their true estimates. The latter makes truthful reporting a Nash equilibrium since an agent will maximise its utility, thus making this strategy the optimal, if all other agents report truthfully their estimates too. $\square$

Theorem 10. *Truthful reporting of the maximum precisions and estimates is a Nash equilibrium under our mechanism.*

Proof. In the mechanism described above, when agent j reports its estimate, its reported precision, $\widehat{\theta}_j$, is equal to its reported maximum precision, $\widehat{\theta}_j^c$. Indeed, $\widehat{\theta}_j > \widehat{\theta}_j^c$ is not possible given that the centre is already informed of the agent's maximum precision, and $\widehat{\theta}_j < \widehat{\theta}_j^c$ would not be in the agent's best interest since under-reporting its precision would lead to a smaller payment. Therefore, $\widehat{\theta}_j = \widehat{\theta}_j^c$. Now, given the scaling of the scoring rules described in step 3 in the second stage of the mechanism, the expected utility of the agent, if it reports its maximum precision as

---

[11]However, it is important to note that whilst the scoring rule can be described at this point in the mechanism, the value of some of these precisions is still unknown, and is only known after the final iteration of the mechanism.

$\widehat{\theta}_j^c$, and subsequently produces an estimate of precision θ_j, which it reports with precision $\widehat{\theta}_j^c$, is denoted by $\overline{U}_j(\theta_j, \widehat{\theta}_j^c)$, and given by:

$$\overline{U}_j(\theta_j, \widehat{\theta}_j^c) = \frac{c_s'(\widehat{\theta}_j^c)}{\overline{S}_{f,j}'(\widehat{\theta}_j^c, \overline{\theta}_{-j})} \left(\overline{S}_{f,j}(\theta_j, \widehat{\theta}_j^c, \overline{\theta}_{-j}) - \overline{S}_{f,j}(\widehat{\theta}_j^c, \overline{\theta}_{-j}) \right) + c_s(\widehat{\theta}_j^c) - c_t(\widehat{\theta}_j^c) \qquad (42)$$

where $\overline{S}_{f,j}(\theta_j, \widehat{\theta}_j^c, \overline{\theta}_{-j})$ is agent j's expected score for producing an estimate of precision θ and reporting to the centre, $\widehat{\theta}^c$ and $\overline{S}_{f,j}(\widehat{\theta}_j^c, \overline{\theta}_{-j})$ is agent j's expected score for producing and truthfully reporting an estimate of precision $\widehat{\theta}_j^c$. Furthermore, $c_t(.)$ is the true cost function of the agent, and $c_s(.)$ is the cost function used to produce the scoring rule (i.e. the $(m+1)^{th}$ lowest revealed cost in the group from which the agent was pre-selected).

Note that $\overline{S}_{f,j}$ is the expected value of the modified scoring rule $\overline{S}(\overline{x}_{-j}; \widehat{x}_j, \widehat{\theta}_j + \overline{\theta}_{-j})$, already defined in Theorem 9:

$$\overline{S}(\overline{x}_{-j}; \widehat{x}_j, \widehat{\theta}_j + \overline{\theta}_{-j}) = \int_{-\infty}^{\infty} \mathcal{N}(\overline{x}_{-j}; x_j, 1/\theta_j + 1/\overline{\theta}_{-j}) S(\overline{x}_{-j} | \mathcal{N}(\overline{x}_j; \widehat{x}_j, 1/\widehat{\theta}_j + 1/\overline{\theta}_{-j})) d\overline{x}_{-j} \quad (43)$$

Now, taking the first derivative of expected utility (Equation 42) with respect to $\widehat{\theta}_j^c$ gives:

$$\frac{d\overline{U}_j(\theta_j, \widehat{\theta}_j^c)}{d\widehat{\theta}_j^c} = \frac{d}{d\widehat{\theta}_j^c} \left(\frac{c_s'(\widehat{\theta}_j^c)}{\overline{S}_{f,j}'(\widehat{\theta}_j^c, \overline{\theta}_{-j})} \right) \left(\overline{S}_{f,j}(\theta_j, \widehat{\theta}_j^c, \overline{\theta}_{-j}) - \overline{S}_{f,j}(\widehat{\theta}_j^c, \overline{\theta}_{-j}) \right) +$$

$$+ \frac{c_s'(\widehat{\theta}_j^c)}{\overline{S}_{f,j}'(\widehat{\theta}_j^c, \overline{\theta}_{-j})} \overline{S}_{f,j}'(\theta_j, \widehat{\theta}_j^c, \overline{\theta}_{-j}) \qquad (44)$$

We have already shown that truthful revelation is a Nash equilibrium for the modified scoring rule, $S_{f,j}$ (Theorem 9). Hence, when $\theta_j = \widehat{\theta}_j^c$:

$$\overline{S}_{f,j}(\theta_j, \widehat{\theta}_j^c, \overline{\theta}_{-j}) = \overline{S}_{f,j}(\widehat{\theta}_j^c, \overline{\theta}_{-j}) \quad \text{and} \quad \overline{S}_{f,j}'(\theta_j, \widehat{\theta}_j^c, \overline{\theta}_{-j}) = 0$$

and subsequently:

$$\left. \frac{d\overline{U}_j(\theta_j, \widehat{\theta}_j^c)}{d\widehat{\theta}_j^c} \right|_{\widehat{\theta}_j^c = \theta_j} = 0 \qquad (45)$$

Therefore, a preselected agent's expected utility is maximised when it reveals as its maximum precision, the precision of the estimate that it subsequently produces, given that all other agents do the same.

We now show that it will actually produce an estimate of precision equal to its reported maximum precision. To this end, we note that when $\widehat{\theta}_j^c = \theta_j$, the expected utility of the agent is given by:

$$\overline{U}_j(\theta) = c_s(\theta_j) - c_t(\theta_j) \qquad (46)$$

Since $c_s(.)$ and $c_t(.)$ do not cross or overlap, and $c_s'(\theta_j) > c_t'(\theta_j)$, then $\overline{U}_j(\theta_j)$ is a strictly increasing function. Thus the agent will maximise its expected utility by producing an estimate at its maximum precision, and thus, $\theta_j = \theta_j^c$, and hence, $\widehat{\theta}^c_j = \widehat{\theta}_j = \theta_j^c$, as required. $\qquad \square$

6.4. Numerical Simulations

Having introduced our mechanism and proven its economic properties, we present empirical results for a specific scenario. We do so in order to compare our mechanism (that uses the fused outcome of all the other agents to score each individual agent) with Miller et al.'s peer prediction mechanism (which calculates the average of pairwise comparisons). As in the previous simulations, the cost functions are represented by linear functions, given by $c_i(\theta) = c_i\theta$, where c_i are independently drawn from a uniform distribution $c_i \sim \mathcal{U}(1, 2)$. Also, the maximum precisions of the selected agents, θ_i^c, are independently drawn from another uniform distribution $\theta_i^c \sim \mathcal{U}(0, 1)$ and, as before, the centre's required precision, θ_0, is equal to 1.7. Finally, as before, we use the parametric family of scoring rules and set the parameter equal to 1.2.

For the purposes of this analysis, the peer prediction mechanism had to be slightly modified in order to eliminate the assumption that the centre has knowledge of the agents' costs. Hence, we transform the peer prediction mechanism to a *two-stage peer prediction mechanism*, in which the centre in the first stage asks all agents, N, to report their cost functions and then pre-selects M of them, while in the second it allocates the payments to the agents providing the estimates. The first stage is identical to the first stage of the mechanism presented in Section 5.2, while in the second stage the centre calculates the payment to an agent not by using the fused reported estimates, but by scoring that agent against each one of the remaining $M - 1$ agents and then by averaging over the $M - 1$ respective payments.

To this end, for $N = 7$, we evaluate our mechanisms (both that from Section 5 which assumes access to the real outcome, and that from this section which does not) against the two-stage peer prediction mechanism. As before, we define the upper bound on the performance of the mechanism as that in which the centre has access to the agents' cost functions (denoted as the 'full information' case) and thus can optimally allocate the estimate to the agent it needs in order to achieve the required precision. We simulate the mechanisms 10^6 times and for each iteration we record whether the centre was successful in acquiring its required precision, and the sum of the payments issued to the selected agents. In Figure 5(a) we show for $M \in \{1, 2, 3, 4, 5, 6\}$, the probability with which the centre has achieved the required precision and the total payment made by the centre (again we omit error bars since, given the number of the iterations, the standard error in the plotted values is smaller than the symbol size), while in Figure 5(b) we show the variance of the total payment the centre issues to the selected agents.

Considering Figure 5(a), we note that for every value of M, the sum of the expected payment for each of the three mechanisms is the same. This means that the centre expects to derive no additional penalties as a result of its lack of knowledge of the actual outcome. Effectively this result shows that the uncertainty that has been introduced in our setting due to the lack of knowledge of the actual outcome has no impact on the expected payments. The explanation of this result lies within the properties of the mechanism itself and we provide a formal proof in the next section, where we calculate the total expected payment for the three evaluated mechanisms for the general case of any convex cost function and show that they are all equal.

However, Figure 5(b), shows that lack of knowledge about the actual outcome does impact the variance in the total payment that the centre makes. In all cases, the variance of total payments is lowest when the centre has access to the actual outcome. Furthermore, for $M = 2$ the variance of the payments is the same in both mechanisms since our mechanism becomes identical with the peer prediction mechanism in this case. However, for $M > 2$ the variance of the payments the centre issues in the peer prediction mechanism is much greater than the variance of payments
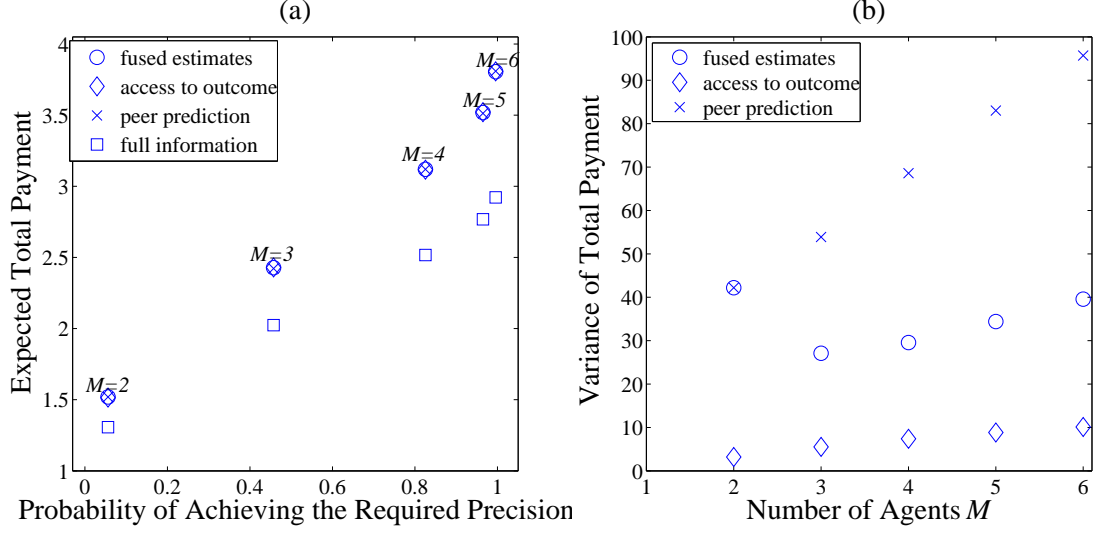
Figure 5: Centre's probability of achieving the required precision and the mean total payment, and the total payment's variance.

in our mechanism. This results from the peer predictions methods increased sensitivity to agents whose estimates diverge from the consensus. In the peer prediction method the impact of such an outlier is somewhat minimised by averaging over all the pair-wise calculated scores. However, as we show here, the impact is further minimised in our approach where all the estimates, apart from that of the agent whose payment is being calculated, are fused together prior to applying the scoring rule.

6.5. Analysis of Expected Total Payment

In the discussion above, we noted that the expected total payment of the centre under all three mechanisms is equal, and is not affected by the lack of knowledge of the actual outcome. To see why this is so, we first note that the mechanism by which an agent is selected to produce an estimate is identical in each case, and thus, we need only show that the expected payment of any of the selected agents is identical in each mechanism. Thus, to this end, we first consider the mechanism in which the centre has access to the actual outcome. In this case, the payment agent j expects to derive, after the centre observes the actual outcome, is given by:

$$\overline{P}_j(\theta_j, \widehat{\theta}_j^c) = \frac{c_s'(\widehat{\theta}_j^c)}{\overline{S}_j'(\widehat{\theta}_j^c)} \left(\overline{S}_j(\theta_j, \widehat{\theta}_j^c) - \overline{S}_j(\widehat{\theta}_j^c) \right) + c_s(\widehat{\theta}_j^c) \tag{47}$$

In this context, given that agents produce estimates with precisions equal to their reported maximum precisions (theorem 5), $\theta_j = \widehat{\theta}_j^c$. Thus, agent j's expected payment is:

$$\overline{P}_j(\theta_j) = c_s(\theta_j) \tag{48}$$

where $c_s(\theta_j)$ is the scaling cost used for in the calculation of agent j's payment.

33

Now, in the mechanism presented in this section in which the centre has no access to the actual outcome, the payment the selected agent j expects to derive is given by:

$$\overline{P}_j(\theta_j, \widehat{\theta}_j^c) = \frac{c_s'(\widehat{\theta}_j^c)}{\overline{S}_{f,j}'(\widehat{\theta}_j^c, \overline{\theta}_{-j})} \left(\overline{S}_{f,j}(\theta_j, \widehat{\theta}_j^c, \overline{\theta}_{-j}) - \overline{S}_{f,j}(\widehat{\theta}_j^c, \overline{\theta}_{-j}) \right) + c_s(\widehat{\theta}_j^c) \qquad (49)$$

Thus, the payment agent j expects to derive, given that $\theta_j = \widehat{\theta}_j^c$, is again given by:

$$\overline{P}_j^f(\theta_j) = c_s(\theta_j) \qquad (50)$$

Finally, in the peer prediction mechanism the centre scores agent j against every other agent in pairs, and then calculates its payment after averaging over the $M-1$ payments that correspond to each one of the selected agents. Hence, the payment agent j expects to derive in the peer prediction mechanism is the following:

$$\overline{P}_j(\theta_j, \widehat{\theta}_j^c) = \frac{1}{M-1} \sum_{i \in M_i} \frac{c_s'(\widehat{\theta}_j^c)}{\overline{S}_{p,j}'(\widehat{\theta}_j^c, \overline{\theta}_i)} \left(\overline{S}_{p,j}(\theta_j, \widehat{\theta}_j^c, \overline{\theta}_i) - \overline{S}_{p,j}(\widehat{\theta}_j^c, \overline{\theta}_i) \right) + c_s(\widehat{\theta}_j^c) \qquad (51)$$

where $\overline{S}_{p,j}$ is the expected score of our modified scoring rule $\overline{S}(\overline{x}_i; \widehat{x}_j, \widehat{\theta}_j + \overline{\theta}_i)$ which scores agent j against the estimate of agent i (unlike our approach which would score agent j against the fused estimate of all other agents). Miller et al. (2007b) have shown that this scoring rule is incentive compatible, therefore $\overline{S}_{p,j}(\theta_j, \widehat{\theta}_j^c, \overline{\theta}_i) = \overline{S}_{p,j}(\widehat{\theta}_j^c, \overline{\theta}_i)$ when $\theta_j = \widehat{\theta}_j^c$. Hence, the payment agent j expects to derive is:

$$P_j^p(\theta_j) = \frac{1}{M-1} \sum_{i \in M_{-j}} c_s(\theta_j) = c_s(\theta_j) \qquad (52)$$

In each of the three mechanisms, the agents are incentivised to truthfully report their estimates, and we find that all three payments are equal, $\overline{P}_j(\theta_j) = \overline{P}_j^f(\theta_j) = \overline{P}_j^p(\theta_j) = c_s(\theta_j)$, as required.

6.6. Discussion

In this section, we provided a non-trivial extension to our previous mechanism by eliminating the assumption that the centre has access to the realised outcome of the estimated event when calculating the payments to the pre-selected agents. As a result of this, now the centre uses the fused reported estimates of the selected agents when calculating its payment. We modified the existing strictly proper scoring rules so they can incentivise agents to truthfully report their estimate, given that they will be scored against all other agents' fused reported estimates and proved that in our mechanism, truthful reporting of the maximum precisions and the estimates is a Nash equilibrium. Finally we empirically evaluated our mechanism for various values of M and compared it with a modified peer prediction mechanism, while using the mechanism which assumes knowledge of the actual outcome as a benchmark. We showed that the centre's total mean payment is the same among all three mechanisms and that for the two mechanisms that do not rely on knowledge of the actual outcome, the variance of the centre's total payment is minimised when it calculates payments using the fused reported agents' estimates. This result indicates that the increase of uncertainty in the system, due to lack of knowledge of the realised outcome, restricts its impact only to the variance of the payments, and does not affect the payments that the agents expect to derive.

7. Related Work

The main scoring rules literature has already been described in Sections 1 and 2, so here we only review other approaches for addressing the problems described within this paper. We first consider mechanism design, and in particularly the VCG mechanism which has been widely used to incentivise truth-telling in dominant strategies when allocating goods or tasks (Vickrey, 1961; Clarke, 1971; Groves, 1973; Krishna, 2002). If we first consider the setting in which any selected agent is capable of providing the centre with the requested estimate, then we can see the two stage mechanism that we introduce as an extension of the VCG mechanism (or more precisely, a second price reverse auction in this case) in which the payments are not directly determined by the types of the agents, but are conditioned on the actual outcome, such that the agents are incentivised to actually commit resources to generating the estimate.

The case is more complex in the second two mechanisms since procuring estimates from multiple agents results in an interdependent valuation setting, with so-called allocative externalities, for which it has been shown that no standard mechanism exists which is both efficient and incentive compatible (Jehiel and Moldovanu, 2001). This has been addressed to a certain degree by Mezzetti (2004, 2007) who shows that efficiency can be achieved in a two-stage mechanism, where in the first stage the final outcome is determined (i.e. the allocation of some goods or tasks), while in the second stage the agents and the centre observe their utilities, and hence receive the final transfers from the centre. Such two stage mechanisms have also been demonstrated in interdependent valuation settings by Klein et al. (2008) who present an application for allocating communication bandwidth, by Porter et al. (2008) for a task allocation setting where agents have some fixed probability of completing a task which the centre must elicit, and Ramchurn et al. (2009) who extend the previous setting by allowing agents to report on the probability with which other agents may be able to complete the requested tasks.

However, these settings depart from the one considered here in one important way. In our setting the agents are able to manipulate the resources that they can commit to generate their estimates. In the setting described above, this is not possible. In the mechanisms of Mezzetti (2004) and Klein et al. (2008) fixed goods are being allocated. While in those of Porter et al. (2008) and Ramchurn et al. (2009), the probability that a task is completed is fixed and is not under the control of the agent. Thus, agents in our setting are able to manipulate their utility by misreporting the precision of their estimates. Crucially, this misreporting cannot be observed by the centre, and it is dependent on the type (their cost function and maximum precision) that they would report to the centre within the VCG mechanism[13]. For this reason, the two stage mechanisms described address do not fully address the challenges of our domain, and thus, rather than seeking an efficient and incentive-compatible mechanism, we settle for an in-efficient mechanism which is still incentive-compatible.

At this point, it should be noted that scoring rules have found other applications not always directly relevant to mechanism design. In more detail, there are many similarities between the logarithmic scoring rule, used in this research, and the *market scoring rules* (Hanson, 2003, 2007) used in prediction markets (Berg and Rietz, 2003; Wolfers and Zitzewitz, 2004). In such systems, agents trade information on probabilistic events and receive pay-offs which depend on

---

[13]Note that in the case of the mechanism that selects a single agent, this manipulation is independent of the agents' types (since their types now only constitute their cost functions), and thus, an efficient mechanism is still possible.

the outcome of these events. In another line of research, strictly proper scoring rules are not used as stand alone payments, but in conjunction with prediction markets (Goel et al., 2009). Although their contribution is significant, since they merge prediction markets and strictly proper scoring rules, in their setting they assume knowledge of the common prior of the agents' subjective beliefs on the estimated parameter. This issue is addressed by Prelec (2004) who propose a mechanism which does not depend on knowledge of the common prior but only on its existence.

However, in all these applications (i.e. market scoring rules and strictly proper scoring rules prediction markets), agents can change their initial reported prediction, if they have new evidence, and their payments also have to be adjusted in order to take into consideration the difference in the agents' reports. This constant flow of information makes these approaches particularly appealing in dynamic systems with rich interactions among the participating agents. Although we are interested in a similar setting where there is also a lack of knowledge about the state of the world, we adopted a less complex approach, in which agents communicate their estimates to a centre and then receive their payments. We believe that our approach is more appropriate for a setting in which a centre wants to acquire a single estimate of a probabilistic event, since it only has to calculate the agents' payments in a single round simply by comparing it with the fused reported estimates or the outcome (if that knowledge is available), instead of implementing the more dynamic and complex process described above.

More importantly, these approaches fail to account for the costs of the agents, since they do not explicitly model the case that agents must invest resources to generate their estimates. This is a key assumption within our setting, and thus, it is difficult to see how these mechanisms can be applied in this setting, since any agent that commits resources to generating an estimate in order to participate within the mechanism must expect to receive at least this amount in payment in order to ensure individual rationality. Extending such mechanisms to address this challenge would certainly be an interesting contribution, but it is outside of the scope of this paper.

8. Conclusions and Future Work

In this paper we contributed to the state of the art by introducing the first mechanism that elicits costly probabilistic estimates from multiple agents in a setting where the centre has no knowledge of the costs involved in the generation of these estimates or the outcome of the estimated event. We achieved this by gradually relaxing the assumptions in two more specific mechanisms. In the first mechanism, the agents faced no restrictions on the precision of the estimates they could provide, while in the second the agents had limitations on the maximum precisions of their estimates. In the both the first and second mechanisms the payment could be conditioned on the actual outcome, whereas in the third mechanism the agents' payments are conditioned on the reports of the other agents. However, it should be noted that although the third mechanism is a generalisation of the first two, both preliminary mechanisms are contributions in their own right and would be used in preference to the more general one in specific settings. Indeed, the first mechanism is the optimal choice in a setting where all agents can provide estimates of precisions that are higher than the required by the centre (with out necessarily being infinite), and therefore the centre does not have to acquire multiple estimates. Moreover, for the second mechanism truthful reporting is a stronger solution concept (i.e. dominant strategy) when compared to the final mechanism, and the payments are more robust since their variance is minimised.

For all three mechanisms we provided both theoretical and empirical results. Specifically, we proved that the first two mechanisms are incentive compatible with respect to all the reported parameters (including costs, estimates and precisions). Moreover, we further modified the strictly proper scoring rules so that they incentivise agents to truthfully report their estimates when they are scored against the fused reports of all the other agents, and we proved that truthful reporting is a Nash equilibrium in this case. In addition, we thoroughly compared the quadratic, spherical and logarithmic scoring rules with a parametric family of strictly proper scoring rules, both analytically and empirically, and provided criteria for the choice of the parameter k. Our final contribution was to show that the increase of uncertainty in our model, as introduced by the centre's lack of knowledge of the realised outcome, did not have an impact on the centre's expected total payment, but it was restricted only to the total payment's variance.

Our future work will address two of the current limitations of our mechanisms. In particular, first we would like to investigate the vulnerability of the mechanisms to collusion among the pre-selected agents in the second stage of the mechanism. Jurca and Faltings (2007) describe a number of potential manipulations of mechanisms whereby all the agents, or a sub-group of the agents, agree on a specific strategy, or where agents deploy 'pseudo-agents' which they can control and impose their strategies (commonly referred to false-name bidding). A number of authors have described auction mechanism which are resistant to collusion and/or false-naming bidding (Day and Milgrom, 2008; Yokoo et al., 2004), and in general these operate by imposing additional constraints on the payments made to the agents. However, such constraints change the economic properties of the auction (for example, the core-selecting package auctions designed by Day and Milgrom (2008) reduce opportunities for bidders to collude but render it only approximately incentive-compatible), and their use in our setting must be carefully evaluated.

Second, we would like to extend our model and mechanism to cases where the cost functions and the exchanged information cannot be modelled by continuous distributions. This will allow us to address a wider set of problems where the exchanged information can be represented by not only continuous but also discrete probability distributions; as used within the task allocation problems of Czumaj and Ronen (2004), or for rating and ranking of information in recommender systems by Adomavicius and Tuzhilin (2005). However, several of the assumptions we make regarding the order of the cost functions and their derivatives will not hold in the case of discrete probability distributions, and thus, it is not obvious that an incentive compatible and individually rational mechanism can still be derived.

9. Acknowledgements

References

G. Adomavicius and A. Tuzhilin. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge and*

Data Engineering, 17(6):734–749, 2005.

B. C. Arnold, N. Balakrishnan, and H. N. Nagaraja. *A First Course in Order Statistics*. SIAM, 2008.

J. E. Berg and T. A. Rietz. Prediction markets as decision support systems. *Information Systems Frontiers*, 5(1):79–93, 2003.

G. W. Brier. Verification of forecasts expressed in terms of probability. *Monthly Weather Review*, 78:1–3, 1950.

E. Clarke. Multipart pricing of public goods. *Public Choice*, 11(1):17–33, 1971.

A. Czumaj and A. Ronen. On the expected payment of mechanisms for task allocation. In *Proceedings of the Twenty-Third Annual ACM Symposium on Principles of Distributed Computing*, pages 98–106, 2004.

R. Day and P. Milgrom. Core-selecting package auctions. *International Journal of Game Theory*, 36(3):393–407, 2008.

M. H. DeGroot and M. J. Schervish. *Probability and Statistics*. Addison Wesley, 2002.

S. Goel, D. M. Reeves, and D. M. Pennock. Collective revelation: A mechanism for self-verified, weighted, and truthful predictions. In *Proceedings of the ACM Conference on Electronic Commerce*, pages 265–274, Stanford, California, USA, 2009.

P. C. Gregory. *Bayesian logical data analysis for the physical sciences: A comparative approach with Mathematica support*. Cambridge Univ Pr, 2005.

S. J. Grossman and O. D. Hart. An analysis of the principal-agent problem. *Econometrica*, 51 (1):7–45, 1983.

T. Groves. Incentives in teams. *Econometrica*, 41(4):617–631, 1973.

R. Hanson. Logarithmic market scoring rules for modular combinatorial information aggregation. *The Journal of Prediction Markets*, 1(1):3–15, 2007.

R. Hanson. Combinatorial information market design. *Information Systems Frontiers*, 5(1): 107–119, 2003.

J. K. Hart and K. Martinez. Environmental sensor networks: A revolution in the earth system science? *Earth-Science Reviews*, 78:177–191, 2006.

A. D. Hendrickson and R. J. Buehler. Proper scores for probability forecasters. *The Annals of Mathematical Statistics*, 42(6):1916–1921, 1971.

P. Jehiel and B. Moldovanu. Efficient design with interdependent valuations. *Econometrica*, 69 (5):1237–1259, 2001.

A. Jøsang, R. Ismail, and A. Boydb. A survey of trust and reputation systems for online service provision. *Decision Support Systems*, 43(2):618–644, 2007.

R. Jurca and B. Faltings. Reputation-based service level agreements for web services. In *Service Oriented Computing*, volume 3826 of *Lecture Notes in Computer Science*, pages 396–409. Springer Berlin / Heidelberg, 2005.

R. Jurca and B. Faltings. Minimum payments that reward honest reputation feedback. In *Proceedings of the ACM Conference on Electronic Commerce*, pages 190–199, Ann Arbor, Michigan, USA, 2006.

R. Jurca and B. Faltings. Collusion resistant, incentive compatible feedback payments. In *Proceedings of the ACM Conference on Electronic Commerce*, pages 200–209, San Diego, California, USA, 2007.

M. Klein, G. A. Moreno, D. C. Parkes, D. Plakosh, S. Seuken, and K. Wallnau. Handling interdependent values in an auction mechanism for enhanced bandwidth allocation in tactical data networks. In *Proceedings of the 3rd international workshop on Economics of networked systems, (NetEcon 2008)*, pages 73–78, New York, USA, 2008.

V. Krishna. *Auction Theory*. Academic Press, 2002.

A. Mas-Colell, M. D. Whinston, and J. R. Green. *Microeconomic Theory*. Oxford University Press, 1995.

J. E. Matheson and R. L. Winkler. Scoring rules for continuous probability distributions. *Management Science*, 22(10):1087–1096, 1976.

C. Mezzetti. Mechanism design with interdependent valuations: Efficiency. *Econometrica*, 72 (5):1617–1626, 2004.

C. Mezzetti. Mechanism design with interdependent valuations: Surplus extraction. *Economic Theory*, 31(3):473–488, 2007.

N. H. Miller, J. W. Pratt, R. J. Zeckhauser, and S. Johnson. Mechanism design with multidimensional, continuous types and interdependent valuations. *Journal of Economic Theory*, 136: 476–496, 2007a.

N. H. Miller, P. Resnick, and R. J. Zeckhauser. Eliciting honest feedback: The peer prediction method. *Management Science*, 51(9):1359–1373, 2007b.

A. Papakonstantinou, A. Rogers, E. H. Gerding, and N. R. Jennings. A truthful two-stage mechanism for eliciting probabilistic estimates with unknown costs. In *Proceedings of the Eighteenth European Conference on Artificial Intelligence*, pages 448–452, Patras, Greece, 2008.

A. Papakonstantinou, A. Rogers, E. H. Gerding, and N. R. Jennings. Mechanism design for eliciting probabilistic estimates from multiple suppliers with unknown costs and limited precision. In *Proceedings of the Eleventh Worhshop in Agent Mediated Electronic Commerce*, pages 111–124, Budapest, Hungary, 2009.

R. Porter, A. Ronen, Y. Shoham, and M. Tennenholtz. Fault tolerant mechanism design. *Artificial Intelligence*, 172(15):1783–1799, 2008.

D. Prelec. A Bayesian truth serum for subjective data. *Science*, 306(5695):462–466, 2004.

S. D. Ramchurn, C. Mezzetti, A. Giovannucci, J. A. Rodriguez, R. K. Dash, and N. R. Jennings. Trust-based mechanisms for robust and efficient task allocation in the presence of execution uncertainty. *Journal of Artificial Intelligence Research*, 35:119–159, 2009.

W. P. Rogerson. The first-order approach to principal-agent problems. *Econometrica*, 53(6): 1357–1367, 1985.

L. J. Savage. Elicitation of personal probabilities and expectations. *Journal of the American Statistical Association*, 66(336):783–801, 1971.

R. Selten. Axiomatic characterization of the quadratic scoring rule. *Experimental Economics*, 1 (1):43–61, 1998.

W. Vickrey. Counterspeculation, auctions and competitive sealed tenders. *The Journal of Finance*, 16(1):8–37, 1961.

G. Werner-Allen, J. Johnson, M. Ruiz, J. Lees, and M. Welsh. Monitoring volcanic eruptions with a wireless sensor network. In *Proceedings of the Second European Workshop on Wireless Sensor Networks*, pages 108–120, Instanbul, Turkey, 2005.

J. Wolfers and E. Zitzewitz. Prediction markets. *Journal of Economic Perspectives*, 18(2):107–126, 2004.

M. Xue, D. Wang, J. Gao, and K. Brewster. The advanced regional prediction system (ARPS): Storm-scale numerical weather prediction and data assimilation. *Meteorology and Atmospheric Physics*, 82(1-4):139–170, 2004.

M. Yokoo, Y. Sakurai, and S. Matsubara. The effect of false-name bids in combinatorial auctions: New fraud in internet auctions. *Games and Economic Behavior*, 46(1):174–188, 2004.

J. Zhou and D. De Roure. FloodNet: Coupling adaptive sampling with energy aware routing in a flood warning system. *Journal of Computer Science and Technology*, 22(1):121–130, 2007.

A. Zohar and J. S. Rosenschein. Mechanisms for information elicitation. *Artificial Intelligence*, 172(16-17):1917–1939, 2008.