Crowd-sourcing with uncertain quality - an auction approach

Papakonstantinou, A. and Bogetoft, P.

Copenhagen Business School, Department of Economics

6 February 2013

Crowd-sourcing with uncertain quality - an auction approach

Athanasios Papakonstantinou
and
Peter Bogetoft*

Copenhagen Business School, Department of Economics,
Porcelaenshaven 16 A, DK-2000 Frederiksberg, Denmark

May 30, 2013

Abstract

This article addresses two important issues in crowd-sourcing: ex ante uncertainty about the quality and cost of different workers and strategic behaviour. We present a novel multi-dimensional auction that incentivises the workers to make partial enquiry into the task and to honestly report quality-cost estimates based on which the crowd-sourcer can choose the worker that offers the best value for money. The mechanism extends second score auction design to settings where the quality is uncertain and it provides incentives to both collect information and deliver desired qualities.

1 Introduction

The unprecedented scale of social interaction in the Internet has allowed people from different parts of the world to collaborate or compete for the completion of various projects. The process of enlisting humans on-line to complete tasks has been labelled '*crowd-sourcing*', and there are several Internet platforms supporting such processes, cf. e.g. [12]. They include Topcoder for software coding, Freelancer for photo moderation and tagging, MTurk for data clean-up and translations among others, and Innocentive for scientific research.

Simplifying a detailed definition in [13], we will think of crowd-sourcing *' .. as a process whereby individuals propose to a group of individuals, via a flexible open call, the voluntary undertaking of a task'.* Those proposing a task are typically referred as 'crowd-sourcers' and their target group as the 'crowd' or the 'workers'. Those of the crowd who end up participating in the project can receive a type of reward depending on the terms of the crowd-sourcers, while the crowd-sourcers get to utilise the crowd's labour. Tasks vary

*Corresponding author. Email address pb.eco@cbs.dk (P. Bogetoft)

with respect to the used platform, with some platforms providing the development tools for breaking a complex task to its components. For example, the CrowdForge framework manages the breakdown of complex tasks in MTurk, such as the writing of an article or market research to its sub-tasks, while it takes care of the emerging dependencies among the different tasks and their corresponding workers [24].

Although crowd-sourcing can increase productivity by turning the world into one virtual working place, it also has some less favourable aspects. Crowd-sourcing, shares with the rest of the Internet, the existence of many layers of *malicious behaviour*. The most common manifestation of such behaviour involves workers choosing to produce work of sub-standard quality i.e. deliberately not meeting the crowd-sourcers publicly announced requirements. Some of the several documented cases where such behaviour occurs include Taskcn, the Chinese crowd-sourcing platform, where it is common for members of the crowd to mis-represent their quality by biding for difficult and complex tasks beyond their capabilities. They speculate that it will be easier to get selected due to lower competition occurring in complex projects, while others will bid for several projects hoping that they will be selected for a few [37]. Other observed cases include MTurk, where a comparison between crowd-workers and specialists revealed that crowd-workers tend to generate results of lower quality [23].

Solutions to the crowd's strategic behaviour are already in place, albeit very basic. For example, in MTurk the crowd-sourcers can reject a completed assignment and consequently refuse payment to the worker if they are not satisfied by the final result. Still, existing crowd-sourcing systems have not benefited as much as other sectors of internet-based commerce by the advancements in trust and reputation systems (a survey of the related literature is [32, 19]), but they also lack the structure that will allow such breakthroughs. We introduce such a structure through the use of Mechanism Design [26] and particularly Auction Theory [25] to model the interactions between the crowd-sourcer (principal) and the workers (agents). We design a payment scheme that incentivises honest reporting and production of appropriate quality, after the workers have invested sufficient resources (i.e. time) in determining their quality. We assume that workers operate in an environment of uncertainty where they report to the crowd-sourcer a probabilistic estimate of their production.

As a starting point, we address the crowd-sourcing problem as a multi-dimensional procurement auction. Single-dimensional auctions have been widely used to procure a given service from the supplier with the lowest cost[35, 16, 10]. However, multi-dimensional auctions are more useful when the service can take many forms, since they take into account not only the price but also the service characteristics or quality when selecting a winner. This is well-suited for crowd-sourcing, where even simple tasks may have several parameters. A software application may for example depend on responsiveness, usability of interface and resource management. In his seminal paper Che [8] designed a series of multi-dimensional auctions (first score, second score and second preferred score) to address cases where the quality of a product is of equal importance to its cost. In these auctions, suppliers report their production quality and the associated costs, and the mechanism maps the multi-dimensional bid into a single-dimensional quantity, named as 'score'. All three auctions are incentive compatible, and based on the assumption that costs are independently distributed, while the first and second score auctions implement the socially optimal (allocatively efficient) outcome. The assumption regarding the distribution of costs was relaxed by Branco [5] who introduced a two-stage optimal multi-dimensional auction in a setting in which there was correlation among suppliers' costs. A mechanism proposed by Bogetoft and Nielsen [4] further exploited the correlations among the costs of different agents through the introduction of a Data Envelopment Analysis (DEA [6, 7]) based competition.

Despite the fact that score auctions emerged from the need for more efficient government procurement, there are also several links with Computer Science and in particular with applications in multi-agent systems and e-commerce [17]. Bichler[2] paves the way for possible e-commerce applications of multi-dimensional auctions by showing that they result in significantly higher utility when compared to single-dimensional auctions in a web-based experimental setting. Furthermore, Beil and Wein [1] propose an iterative mechanism in which the buyer sequentially estimates each bidder's cost function through a series of score auctions. Parkes and Kalagnanam [31] also propose an iterative multi-attribute price-based procurement auction in which suppliers in each round submit their bids and a winner maximizing the buyer's preference is selected. They show that their mechanism terminates with a modified Vickrey-Clarke-Groves allocation. Furthermore, multi-dimensional auctions can also be applied in settings where multiple suppliers are necessary to satisfy the principal's demand [3].

Now, although these approaches address effectively specific issues, they do not combine all the elements we require. Most importantly, the literature does not take into account the real world challenge of ensuring truthful reporting when there is uncertainty about the quality and of ensuring the final production when this cannot be fully controlled. In cases where there is no uncertainty, it is assumed that the principal can enforce the agents to truthfully report their production quality, through the use of external means. In the few cases where the possibility of misreporting quality is considered it is explicitly stated that the auction will be cancelled, or an extremely heavy fine will be issued to the winner of the auction if the observed output deviates from its report. Obviously, such an approach does not work when quality is uncertain.

This multi-layered challenge can be addressed by incorporating a *strictly proper scoring rule* payment in a multi-dimensional auction. Strictly proper scoring rules are designed to elicit accurate predictions by rewarding forecasters based on how close the actual outcome is to their prediction[33, 18, 14]. Strictly proper scoring rules have been widely used in mechanism design to elicit accurate information and in particular for the design of reputation systems to promote truthful reporting of feedback regarding the quality of a service experienced [20, 21, 22]. Furthermore, Miller et al. [27, 28] have shown how an appropriately scaled strictly proper scoring rule can be used to incentivise agents to invest costly resources when generating their forecasts. Extensions are given in [30] and [36], and a brief summary of the main insights used in this paper is provided in Section 3.

In this paper we combine elements from multi-dimensional auctions and information elicitation mechanisms. We consider a setting where the worker is not certain of the quality of its future production when reporting it to the crowd-sourcer. Workers base their beliefs on initial expectations and costly investigations modelled as the observation of a sample of independent Gaussian distributions. After the auction is completed the winner starts working on its assigned task and the crowd-sourcer observes the outcome after the work is finished. Based on this observation and the initial report, the crowd-sourcer penalises the selected worker (winner) for any deviation from its report, while compensating the worker's actual cost and quality.

We provide solid theoretical results as we prove the economic properties of our mechanism i.e. incentive compatibility and individual rationality. We also show that in expectation our mechanism achieves the outcome of the second score auction in which agents are able to directly report their actual quality outcomes, and that agents invest the maximum amount of resources available to them in order to generate precise estimates of their qualities. Finally, we numerically evaluate our mechanism though simulations, where we discuss its computational aspects and demonstrate its convergence to the outcome of the second score auction, under the strong assumption that there is no uncertainty about the agents' qualities.

The rest of the paper is organised as follows: In Section 2 we describe the setting in more details, and in Section 3 we provide the relevant background in strictly proper scoring rules. In Section 4 we define the mechanism, while in Section 5 we outline the economic properties. In Section 6 we evaluate the mechanism though numerical simulations and in Section 7 we conclude.

2 The Context

We consider a principal (the crowd-sourcer) interested in procuring a task or a service from one of N rational and risk neutral agents (the crowd or the workers). The provided task or service may be an independent task or part of a more comlex one, without this affecting our analysis, and is characterised by multiple parameters defined by an s-dimensional vector of qualities $y_0^i \in \mathbb{R}^s$ with $y_0^i > 0$ and $i \in I = \{1, \dots N\}$. To simplify the analysis, we assume that for each agent the parameters of its service can be aggregated in one variable, hence each agent has a single quality profile denoted by y_0^i.

We depart from the existing literature by introducing uncertainty regarding the agent's qualities. We model uncertainty by assuming that each agent i attempts to estimate his individual production (quality) y_0^i by generating a sample of M independent observations $y_j^i, j \in \{1, ..., M\}$. In more detail, an individual agent does not know his quality ex ante; instead he has an a priori belief and he can collect additional information. Since Gaussian distributions are commonly used in the data fusion literature [15, 11], we will also use them here. We therefore assume that agent i's a priori belief about y_0^i is given as $y_0^i \sim \mathcal{N}(y_\mu^i, 1/\theta_\mu^i)$, and that he is able to collect further information about y_0^i by generating M independent and identically distributed random observations $\{y_1^i, y_2^i, \dots, y_M^i\}$ with $y_j^i \sim \mathcal{N}(y_0^i, 1/\theta_0^i)$. Using these observations, the agent can update the a priori beliefs to the posterior belief

$$y_0^i \sim \mathcal{N}\left(\frac{y_\mu^i \theta_\mu^i + \theta^i \bar{y}^i}{\theta_\mu^i + \theta^i}, \frac{1}{\theta_\mu^i + \theta^i}\right) \tag{1}$$

where $\bar{y}^i$ is the mean of the observations $\{y_1^i, y_2^i, \dots, y_M^i\}$ and θ^i the resulting precision of the sample average $\bar{y}^i$, equal to $M\theta_0^i$.

It is natural to assume that the cost of collecting information about the likely quality increases as the precision θ^i increases. we will therefore model data collection cost $c^i(\theta^i)$ as a convex, increasing and double differentiable function such as $c(\theta) = C^i\theta^2$, where $C^i > 0$

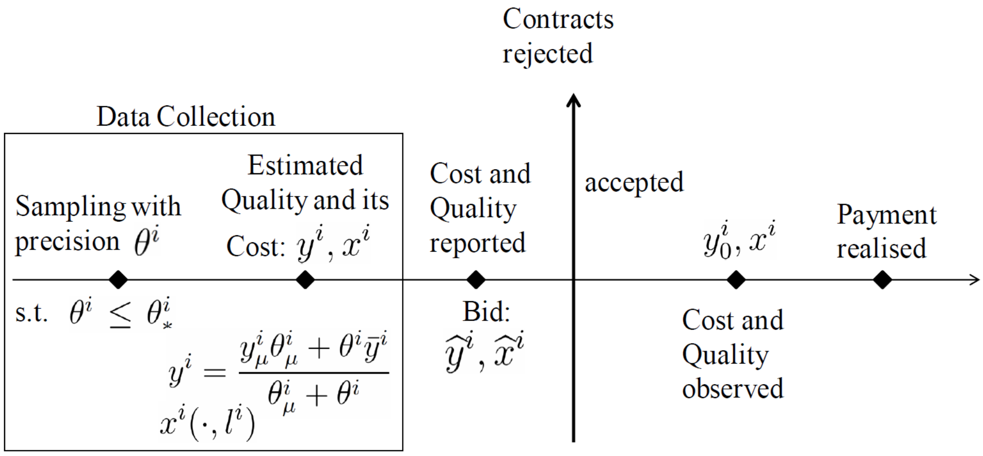

Figure 1: The time-line of the game.

is a parameter which represents different base costs for each agent. Typically, costs in data collection introduce constraints in the overall precision since it will be impossible to have an infinite sample, or to have a finite sample of very costly observations. This constraint is denoted as θ_*^i, with $\theta^i \leq \theta_*^i$.

Now, regarding the production costs, we follow the existing literature [8] by assuming that agents are capable of producing different levels of outputs, and that in order to produce the quality y_0^i agent i needs inputs which depend on each agent's efficiencies. These inputs are the costs involved in production and should not be confused with the costs involved in the estimation of the quality. Here, costs are private information to each agent and cannot be verified by any third party. The cost agent i faces in the production of its quality y_0^i is denoted $x^i(y_0^i, l^i)$, where l^i represents the agent's private information about his production cost (in)efficiency. While agents are aware of their cost parameters, the principal has only access to their distribution. We assume that l^i is independently and identically distributed over $[\underline{l}, \bar{l}]$ with $0 < \underline{l} < \bar{l} < +\infty$ according to a distribution with positive and continuously differentiable density function. Finally, the cost function is increasing in both quality and the cost (in)efficiency parameter and it is convex in the quality.

Based on the above, the time-line (Figure 1) of the game is as follows. Initially, each agent collects information about his likely production quality and production costs. By sampling with precision θ^i, and spending information collection costs $c^i(\theta^i)$ it is able to predict its quality y^i as $(y_\mu^i \theta_\mu^i + \bar{y}^i \theta^i)/(\theta_\mu^i + \theta^i)$ with precision $\theta_\mu^i + \theta^i$, and based that prediction, the cost of the production as $x^i(y^i, l^i)$. We assume that the agent can send possibly manipulated signals about his production (quality) level, his production costs, and the precision of his prediction to the principal before the principal decides on the provider. Let the signalled production be $\widehat{y}^i$, the signalled data collection effort be $\widehat{\theta}^i$, and the signalled cost be $\widehat{x}^i$. The principal can use these signals to choose the provider and he can use this information together with the realised production y_0^i to determine reimbursement. If the principal picks agent i as the provider, his value of the realised quality y_0^i will be given by $V(y_0^i)$ where $V(\cdot)$ is an increasing, concave and twice differentiable function of the quality.

To sum up, in this setting the principal has to deal with poor quality and costs estimates generated by agents not committing significant resources to the pre-bidding information collection, with misreporting of the estimates and with incentivising the selected agent to actually produce the final outputs.

3 Strictly Proper Scoring Rules

Before turning to the details of the mechanism, it is convenient to discuss the simpler problem of inducing agents to collect information about their production and to reveal their findings.

So-called strictly proper scoring rules are used as a tool for eliciting forecasters' beliefs of future events in various domains ranging from meteorology and weather forecasting to computer science and online trust and reputation systems. Such scoring rules incentivise a risk neutral forecaster to truthfully report his forecast by maximizing his expected reward. Imagine that a forecaster has to predict an event y and that he reports a probability distribution $R(y)$. If the realized outcome is y^*, his score is then $S(y^*|R)$. If the forecasters belief is that y is generated by a probability density function $Q(y)$, his expected score when he reports R is:

$$\overline{S}(Q, R) = \int_{-\infty}^{\infty} Q(y) S(y|R) dy \tag{2}$$

and we say that the scoring rule $S(y|R)$ is *strictly proper* if its expected value is maximised by truthful reporting i.e. $\overline{S}(Q, Q) \geq \overline{S}(Q, R)$ for all R. Due to this property, a payment based on such a scoring rule can create incentives for truthful reporting.

The four most popular strictly proper scoring rules in the literature are the quadratic, spherical, logarithmic and the parametric power rule family [34]. For the special case where an agent's posterior belief of his quality y_0 is represented by a Gaussian distribution $\mathcal{N}(y, 1/\theta)$ and his report is $(\widehat{y}, \widehat{\theta})$ the four rules, $S(y_0; \widehat{y}, \widehat{\theta})$, take the following forms:

1. Quadratic: $\quad 2\mathcal{N}(y_0; \widehat{y}, 1/\widehat{\theta}) - \frac{1}{2}\sqrt{\frac{\widehat{\theta}}{\pi}}$
2. Spherical: $\quad \left(\frac{4\pi}{\widehat{\theta}}\right)^{\frac{1}{4}} \mathcal{N}(y_0; \widehat{y}, 1/\widehat{\theta})$
3. Logarithmic: $\quad \log \mathcal{N}(y_0; \widehat{y}, 1/\widehat{\theta})$
4. Parametric: $\quad k\mathcal{N}(y_0; \widehat{y}, \theta)^{(k-1)} - \frac{k-1}{\sqrt{k}} \left(\frac{2\pi}{\widehat{\theta}}\right)^{\frac{1-k}{2}}$

where $k \in (1, 3)$. When $k = 2$ the parametric rule coincide with the quadratic rule.

It is interesting to note that strictly proper scoring rules can guarantee not only truthful reporting, but also sufficient data collection effort on behalf of the agents. This process is described by Miller et. [27] who note that making payment an affine function $\alpha + \beta S$ of a strictly proper scoring rule S it is possible to induce an agent to make and truthfully report an estimate at a specific precision.

An agent's expected payment , $\overline{P}(\theta)$, is:

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

where α and β are the scaling parameters in the affine transformation, θ is the agent's true precision and $\overline{S}(\theta)$ is the expected score. Parameter α guarantees the estimate will be generated at the appropriate precision, while β compensates the agent for the cost of his estimate.

In our model, θ is equal to $\theta_\mu + \theta'$, where θ_μ is the precision of the agent's prior belief and θ' the precision of the sample average. The expected utility to an agent net of data collection costs is therefore:

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

Imagine now that there is a constraint θ^* on the agent's precision (i.e. $\theta' \leq \theta^*$). It it is in the best interest of a principal solely interested in data collection to elicit an estimate at that maximum precision. Hence, the principal will choose a value for α so that the agent's precision is equal to θ^*. That is, the principal selects an α which maximises the agent's expected utility at θ^*. To do so, the principal solves $\left.\frac{\mathrm{d}\overline{U}}{\mathrm{d}\theta'}\right|_{\theta^*} = 0$ to give:

$$\alpha = \frac{c'(\theta^*)}{\overline{S}'(\theta^*)} \tag{3}$$

The β parameter serves only to ensure participation in the mechanism by ensuring that the agent's expected utility is positive. Presuming that the expected utility from the data collection and reporting alone shall be at least 0 we get

$$\beta = c(\theta^*) - \frac{c'(\theta^*)}{\overline{S}'(\theta^*)}\overline{S}(\theta^*) \tag{4}$$

Based on Equations 3 and 4 we calculate the specific values of α and β. Clearly, the parameters will depend on which one of the strictly proper scoring rule is used. This raises the important issue of which rule should be selected by the principal. Indeed, each one of the aforementioned four strictly proper scoring rules has additional properties that can be considered in addition to incentivising truthful reporting and eliciting sufficient effort. For example, the logarithmic scoring rule and the parametric one for $k \to 1$ lead to the lowest expected payments, but they have no lower bounds. It is suggested in [29] that the parametric scoring rule offers a good compromise. Selecting a value for the parameter k within $(1, 1.5)$ keeps the payment relatively low for the majority of the agents, and the finite lower bound protects the agents who generate inaccurate estimates ($\mathcal{N} \to 0$). For the parametric rule, the parameters α and β become:

$$\alpha = \frac{2c'(\theta^*)(\theta_\mu + \theta^*)\sqrt{k}}{k-1}\left(\frac{\theta_\mu + \theta^*}{2\pi}\right)^{\frac{1-k}{2}} \tag{5}$$

and

$$\beta = c(\theta^*) - \frac{2(\theta_\mu + \theta^*)}{k-1}c'(\theta^*) \tag{6}$$

4 The Mechanism

Our proposed mechanism implements a two-step payment to the winner of a second score auction based on the agents' reported beliefs of their qualities. The first payment to the winner is equal to the second score auction's payment based on that reported belief and is received before the actual production. Once that agent produces his quality and it is observed by the principal, he receives his secondary payment. This payment consists of the three following parts:

1. A symmetric penalty if the selected agent produced an inaccurate report.
2. A compensation for the costs involved in the generation of the estimate based on its accuracy.
3. A compensation for the selected agent's production based on the realised quality.

We introduce a scaled strictly proper scoring rule to evaluate the selected agent's probabilistic estimate. Although the use of scoring rules does not guarantee that the selected agent's reported belief will be close to his actual production, since an agent's sample can always include a significant number of poor observations, it does motivate the agent to invest sufficient resources when generating his estimate and then to truthfully report it.

The mechanism is formally defined as follows:

1. Principal invites N agents to participate in the procurement auction and ask them to report their precision constraint θ_*^i.
2. Agents generate and report estimates of their outputs $\widehat{y}^i$, their precision $\widehat{\theta}^i$, and their costs $\widehat{x}^i$, for $i \in \{1, ..., N\}$.
3. Each bid is assigned a score $\widehat{S}^i = S_(\widehat{x}^i, \widehat{y}^i) = V(\widehat{y}^i) - \widehat{x}^i$, for $i \in \{1, ..., N\}$
4. The agent with the highest score wins the auction and is allocated the project.
5. The winner[1] receives its **first payment** from the principal: $\widehat{P} = V(\widehat{y}) - \widehat{S}_{(2)}$ similar to the payment in a second score auction.
6. Winning agent produces quality y_0.
7. Principal observes winning agent's quality production and issues the **second payment**:

$$B(y_0; \widehat{y}, \widehat{\theta}) = d(y_0; \widehat{y}, \widehat{\theta})[V(\widehat{y}) - \widehat{S}_{(2)}] + \alpha S(y_0; \widehat{y}, \widehat{\theta}) + \beta + [V(y_0) - \widehat{S}_{(2)}]$$

   where $d(y_0; \widehat{y}, \widehat{\theta})$ is a function that evaluates the selected agent's reported estimate based on the observed actual production, parameters α and β are the effort inducing parameters from the scaled strictly proper scoring rule $S(y_0; \widehat{y}, \widehat{\theta})$ incentives, and $S_{(2)}$ is the score of the runner up agent in the initial second score auction (Step 5).

[1] In order to simplify our notation we omit the use of subscript (1) to denote the winner of the auction, while we maintain the use of (2) for the runner-up agent.

The function $d(\cdot)$ serves to guarantee truthful reporting by penalising deviation from truth telling. Since an agent's report can deviate from his actual production due to unforeseen circumstances (inherent poor observations) and due to strategic behaviour, we let the deviation function $d(\cdot)$ be based on a scaled strictly proper scoring rule which elicits truthful behaviour and maximises agent's effort. The function is defined as following:

$$d(y_0; \widehat{y}, \widehat{\theta}) = S(y_0; \widehat{y}, \widehat{\theta}) - \overline{S}(\widehat{\theta}^*) - 1 \tag{7}$$

where $\widehat{\theta}^*$ is the agent's reported constraint, $S(y_0; \widehat{y}, \widehat{\theta})$ is the scoring rule and $\overline{S}(\widehat{\theta}^*)$ is the expected score as a function of the reported constraint $\widehat{\theta}^*$.

The *total* payment a truthful agent expects to derive by this mechanism is the following:

$$\overline{P}(\theta) = [\overline{S}(\theta) - \overline{S}(\widehat{\theta}^*)][V(y) - S_{(2)}] + \alpha\overline{S}(\theta) - \beta + V(y_0) - S_{(2)} \tag{8}$$

In the following section, where we prove the mechanism's economic properties, we also show in detail how the above expression is derived.

5 Economic Properties

Having described in detail the mechanism, we now develop its economic properties. Specifically we show that:

1. Agents are incentivised to generate quality estimates at their maximum precisions, and to truthfully report the estimates and precisions.
2. Truthful revelation of the production costs is a weakly dominant strategy given a truthful report of the quality estimate.
3. The mechanism is immune to the effects of combined misreporting of quality and cost.
4. The mechanism is individually rational for the selected agent (auction winner).

Lemma 1. *It is a dominant strategy for an agent to generate a quality estimate at his maximum precision and to honestly reveal this to the principal.*

Proof. We prove this Lemma by showing that the winner of the auction (referred as 'selected agent') will truthfully reveal his quality estimate at his maximum precision.

The agent's utility when he reports $(\widehat{y}, \widehat{\theta})$ and realized quality is y_0, is:

$$U(y_0 \mid \widehat{y}, \widehat{\theta}) = V(\widehat{y}) - S_{(2)} + [S(y_0; \widehat{y}, \widehat{\theta}) - \overline{S}(\widehat{\theta}^*) - 1][V(\widehat{y}) - S_{(2)}] + \alpha S(y_0; \widehat{y}, \widehat{\theta}) + \beta + V(y_0) + S_{(2)} - x(y_0) - c(\theta)$$

where α and β are the strictly proper scoring rules scaling parameters defined in Section 3.

By integrating over the set of possible outputs y_0 we derive the winner's expected utility from reporting $(\widehat{y}, \widehat{\theta})$:

$$\overline{U}(\widehat{y}, \widehat{\theta}) = \int_{-\infty}^{\infty} \mathcal{N}(y_0; y, 1/\theta)[V(\widehat{y}) - S_{(2)}]dy_0$$

$$+\int_{-\infty}^{\infty} \mathcal{N}(y_0; y, 1/\theta)[S(y_0; \widehat{y}, \widehat{\theta}) - \overline{S}(\widehat{\theta}^*) - 1][V(\widehat{y}) - S_{(2)}]dy_0$$

$$+\int_{-\infty}^{\infty} \mathcal{N}(y_0; y, 1/\theta)[\alpha S(y_0; \widehat{y}, \widehat{\theta}) + \beta]dy_0$$

$$+\int_{-\infty}^{\infty} \mathcal{N}(y_0; y, 1/\theta)[V(y_0) + S_{(2)} - x(y_0) - c(\theta)]dy_0$$

Since the initial payment does not depend on the final outcome and since $\int_{-\infty}^{\infty} \mathcal{N}(y_0; y, 1/\theta)dy_0 = 1$, a simpler expression is:

$$\overline{U}(\widehat{y}, \widehat{\theta}) = [V(\widehat{y}) - S_{(2)}] \int_{-\infty}^{\infty} \mathcal{N}(y_0; y, 1/\theta)[S(y_0; \widehat{y}, \widehat{\theta}) - \overline{S}(\widehat{\theta}^*)]dy_0$$

$$+\int_{-\infty}^{\infty} \alpha \mathcal{N}(y_0; y, 1/\theta) S(y_0; \widehat{y}, \widehat{\theta})dy_0 + \beta - c(\theta) + \int_{-\infty}^{\infty} \mathcal{N}(y_0; y, 1/\theta)[V(y_0) - x(y_0)]dy_0 + S_{(2)}$$

The above expression can be further simplified by using the notation of the expected score:

$$\overline{S}(\widehat{\mathcal{N}}, \mathcal{N}) = \int_{-\infty}^{\infty} \mathcal{N}(y_0; y, 1/\theta) S(y_0; \widehat{y}, 1/\widehat{\theta})dy_0$$

where $\widehat{\mathcal{N}}$ represents the reported distribution of y_0 and $\mathcal{N}$ the distribution of his true estimate.

To sum up, the selected agent's expected utility from estimating and reporting his quality is:

$$\overline{U}(\widehat{y}, \widehat{\theta}) = [V(\widehat{y}) - S_{(2)}][\overline{S}(\widehat{\mathcal{N}}, \mathcal{N}) - \overline{S}(\widehat{\theta}^*)] + \alpha \overline{S}(\widehat{\mathcal{N}}, \mathcal{N}) + \beta$$

$$+\int_{-\infty}^{\infty} \mathcal{N}(y_0; y, 1/\theta)[V(y_0) - x(y_0)]dy_0 + S_{(2)} - c(\theta)$$

Having defined the selected agent's expected utility function we proceed to show that it is maximised when the agent reports its true estimate and its precision, given that that precision will be its maximum (i.e. constraint). Initially, it easy to see that due to the use of a strictly proper scoring rule, the expected scoring rule $\overline{S}(\widehat{\mathcal{N}}, \mathcal{N})$ is maximised at $\widehat{\mathcal{N}} = \mathcal{N}$, hence (y, θ) is a local maximum for the expected score.

Now, in Section 3 we have defined the parameters α and β so that the agent is incentivised to make an estimate with maximal[2] precision, θ^*. Based on these two properties, the partial derivatives of $\overline{S}(\widehat{\mathcal{N}}, \mathcal{N})$ w.r.t $\widehat{y}$ and $\widehat{\theta}$ are equal to 0, for $(\widehat{y}, \widehat{\theta}) = (y, \theta^*)$, when $\theta = \theta^*$.

In order to show that (y, θ^*) is a maximum point for $\overline{U}$, we first show that it is a critical one:

$$\frac{\partial \overline{U}}{\partial \widehat{y}} = V'(\widehat{y})[\overline{S}(\widehat{\mathcal{N}}, \mathcal{N}) - \overline{S}(\widehat{\theta}^*)] + [V(\widehat{y}) - S_{(2)} + \alpha] \frac{\partial \overline{S}(\widehat{\mathcal{N}}, \mathcal{N})}{\partial \widehat{y}} = 0 \quad (9)$$

$$\frac{\partial \overline{U}}{\partial \widehat{\theta}} = [V(\widehat{y}) - S_{(2)} + \alpha] \frac{\partial \overline{S}(\widehat{\mathcal{N}}, \mathcal{N})}{\partial \widehat{\theta}} = 0 \quad (10)$$

[2] The use of the term 'maximal' implies that although the selected agent reports his own maximum precision, that precision may not be the maximum one reported given the reports of the other agents.

Since S is a strictly proper scoring rule, we have $\dfrac{\partial\overline{S}(\widehat{\mathcal{N}},\mathcal{N})}{\partial\widehat{y}} = \dfrac{\partial\overline{S}(\widehat{\mathcal{N}},\mathcal{N})}{\partial\widehat{\theta}} = 0$ for $(\widehat{y},\widehat{\theta}) = (y,\theta^*)$. Hence the first order conditions for $\overline{U}$ are fulfilled by (y,θ^*).

Moreover, the determinant of the Hessian matrix of $\overline{U}$ is:

$$\mathrm{Det}(\mathcal{H}(\overline{U}))(y,\theta^*) = [V(y) - S_{(2)} + \alpha]^2[\frac{\partial^2\overline{S}}{\partial\widehat{y}^2}\frac{\partial^2\overline{S}}{\partial\widehat{\theta}^2} - \frac{\partial^2\overline{S}}{\partial\widehat{y}\partial\widehat{\theta}}\frac{\partial^2\overline{S}}{\partial\widehat{\theta}\partial\widehat{y}}] =$$

$$= [V(y) - S_{(2)} + \alpha]^2\mathrm{Det}(\mathcal{H}(\overline{S}))(y,\theta^*) \tag{11}$$

which is positive given that $[V(y) - S_{(2)} + \alpha]^2 > 0$ and $\mathrm{Det}(\mathcal{H}(\overline{S}))(y,\theta^*) > 0$ since (y,θ) is a maximum for the expected score $\overline{S}$ and $\theta = \theta^*$.

Hence (y,θ^*) is a maximum for $\overline{U}(\widehat{y},\widehat{\theta})$ given that $\widehat{\theta} = \theta$. We have therefore shown that truthful revelation of a selected agent's quality and its precision is a dominant strategy and that the agent is incentivised to generate his estimate at a precision equal to his reported constraint. □

Lemma 2. *Agents report truthfully their constraints in the initial stage of the mechanism.*

Proof. It is obvious that the fact that the agents have a constraint in the precision of their estimated qualities suggests that they will not be able to generate an estimate at a higher precision.

In addition to this, if an agent generates his estimate at a precision lower than his constraint, θ', s.t. $\theta' < \theta^*$, there will be a loss in the expected score, since $\overline{S}(\theta') < \overline{S}(\theta^*)$. Hence the penalty in the secondary payment will increase, which in turn will reduce his overall utility.

It becomes apparent that agents are incentivised to truthfully reveal their constraints in the opening of the mechanism, given that not doing so is either counter-intuitive, or results in loss of utility. □

Lemma 3. *It is a weakly dominant strategy for an agent to truthfully reveal production costs given that his reported quality estimate is equal to his true estimate.*

Proof. A selected agent who was truthfully reported his quality estimate and its precision expects the following utility:

$$\overline{U}(y) = \int_{-\infty}^{\infty}\mathcal{N}(y_0; y, 1/\theta)[V(y_0) - x(y_0)]dy_0 - S_{(2)}$$

Representing the Gaussian probability distribution as the Dirac delta function, leads to a transformation which simplifies the above expression. The transformation is based on the property of the Dirac delta function: $\int_{-\infty}^{\infty} f(y_0)\delta(y_0 - y)dy_0 = f(y)$, where $f(y_0)$ is equal to $V(y_0) - x(y_0)$ and $\delta(y_0 - y) = \dfrac{1}{\frac{\sqrt{2}}{\sqrt{\theta}}\sqrt{\pi}}\exp\left(-\theta(y_0 - \widehat{y})^2/2\right)$, with the Dirac delta function $\delta_\alpha(y) = \dfrac{1}{\alpha\sqrt{\pi}}\exp(-y^2/\alpha^2)$, and $\alpha = \frac{\sqrt{2}}{\sqrt{\pi}}$.

Now it is possible to replace $\int_{-\infty}^{\infty}\mathcal{N}(y_0; y, 1/\theta)[V(y_0) - x(y_0)]dy_0$ with $V(y) - x(y)$ which is in fact the selected agent's true parameters. Hence:

$$\overline{U}(y) = V(y) - x(y) - S_{(2)} = S_{(1)} - S_{(2)}$$

Given this insight, we prove the Lemma by contradiction:

Let x and y be an agent's true cost and quality, and S the score that corresponds to these true values, and let $\widehat{x}, \widehat{y}$ and $\widehat{S}$ be the reported ones. Furthermore, let $x_{(2)}, y_{(2)}, S_{(2)}$ be the bids, and the score of the runner up agent (i.e. $\widehat{S} > S_{(2)}$).

First, let the agent's misreporting have an effect on the outcome of the auction. We consider the following two cases:

1. Agent wins by misreporting while it would have lost if truthful.

2. Agent loses by misreporting while it would have won if truthful.

- In Case (1) agent reports his cost s.t $\widehat{S} > S_{(2)}$ given that $S < S_{(2)}$. The agent achieves this by reporting a lower cost than his actual one i.e. $\widehat{x} < x$. Under optimal reporting of quality, the utility of an agent misreporting his cost in Case (1) will be negative i.e. $\overline{U}(y) = V(y) - x(y) - S_{(2)} = S_{(1)} - S_{(2)} < 0$.

- In Case (2) agent reports his cost s.t. $\widehat{S} < S_{(2)}$ given that $S > S_{(2)}$. The agent would have won the auction, but instead reports a cost greater than his actual one i.e. $\widehat{x} > x$. As a result, the agent loses the auction and consequently receives negative utility (since he still faces the costs of determining his quality).

Second, we assume that the agent misreports his cost of production without this affecting whether he wins the auction or not. If the agent had already lost the auction, misreporting would have no additional effect given that the utility would be negative due to the cost of determining his quality without any dependence on the cost of production. Had the agent already won the auction, misreporting would not result in additional benefits. Specifically, both payments depend on the second lowest score and, the reported and actually produced (for the second stage) quality and the compensation for his estimate. □

Theorem 1. *The mechanism is immune to combined misreporting of quality and cost.*

Proof. In the above proofs we showed that truthful reporting of the production quality is an optimal strategy if the agent reports truthfully his cost, and that the same holds for an agent's costs, given that he generated an accurate estimate of his quality by investing the maximum amount of resources in determining it. However, given the multi-dimensional nature of the bids an agent could attempt to manipulate the principal by misreporting both costs and the precision of his quality estimate.

In this proof we examine agents' strategic behaviour as a whole. We will show that even when it is possible for some type of misreporting to occur, there is no negative impact on the principal.

In order to demonstrate how it is not optimal for an agent to deviate from truthful behaviour we consider the four following general cases of misreporting:

1. Agent wins the auction by misreporting both his estimate of quality and production cost

2. Agent wins the auction with the misreporting having no effect on the auction's outcome

3. Agent loses the auction due to his misreporting

4. Agent loses the auction despite his misreporting

- In Case (1) the agent reports his estimate of quality and cost s.t. $\widehat{S} > S_{(2)}$, while $S < S_{(2)}$, with his precision not necessarily equal to his reported constraint. We will show that the misreporting agent's expected utility $\overline{U}(\widehat{y}, \widehat{\theta})$ will always be less or equal to the utility of a truthful agent $\overline{U}(y, \widehat{\theta}^*)$:

$$\overline{U}(\widehat{y}, \widehat{\theta}) - \overline{U}(y, \widehat{\theta}^*) = [V(\widehat{y}) - S_{(2)}][\overline{S}(\widehat{\mathcal{N}}, \mathcal{N}) - \overline{S}(\widehat{\theta}^*)] + \alpha \overline{S}(\widehat{\mathcal{N}}, \mathcal{N}) + \beta - c(\theta) \quad (12)$$

  Regarding $V(\widehat{y}) - S_{(2)}$ we have assumed that it is a positive quantity since $\widehat{S} > S^{(2)} \Rightarrow V(\widehat{y}) - \widehat{x}(\widehat{y}) > S_{(2)} \Rightarrow V(\widehat{y}) > S_{(2)}$, while $\overline{S}(\widehat{\mathcal{N}}, \mathcal{N}) - \overline{S}(\widehat{\theta}^*)$ is negative since $\overline{S}(\widehat{\mathcal{N}}, \mathcal{N}) \leq \overline{S}(\widehat{\theta}^*)$ given that S is a strictly proper scoring rule.

  Finally, after replacing α and β it can be shown that $\alpha \overline{S}(\widehat{\mathcal{N}}, \mathcal{N}) + \beta - c(\theta) < 0$:

$$\frac{c'(\widehat{\theta}^*)}{S'(\widehat{\theta}^*)}[\overline{S}(\widehat{\mathcal{N}}, \mathcal{N}) - \overline{S}(\widehat{\theta}^*)] + c(\widehat{\theta}^*) - c(\theta)$$

  which is negative since $\overline{S}(\widehat{\mathcal{N}}, \mathcal{N}) \leq \overline{S}(\widehat{\theta}^*)$ and $c(\widehat{\theta}^*) - c(\theta) < 0$ since it is not optimal for an agent to report a constraint lower than his intended precision if it knows that it will be paid based on his constraint, and consequently lose by doing so.

- In Case (2) the agent would have won the auction anyway, and although misreporting of cost and quality will have no impact on the outcome of the auction, it may have on the secondary payment. Still, such a manipulation is not attractive since we have from Case (1) that $\overline{U}(\widehat{y}, \widehat{\theta}) \leq \overline{U}(y, \widehat{\theta}^*)$. Even if we assume that the estimate's precision is equal to the reported constraint, it is still the misreporting of the estimate and the production cost which makes this strategy sub-optimal.

Cases (3) and (4) are simpler. For both cases it is obvious that the utility of an agent not winning the initial auction will solely consist of the cost of data collection. In Case (3) the agent deliberately misreports his estimate and his production cost in order to lose. It would be in his best interest to invest minimum resources in generating his estimate, so that he can minimise his inevitable loss. However, that is not a straightforward decision. Estimates of low precision may end up winning the auction and inflicting additional losses, while estimates of high precision will increase his losses. Effectively, an agent who wants to lose the auction has no reason to participate in the auction. Now, in Case (4) the agent misreports with the intention to win but ends up losing the auction. Had the agent won, it would result in negative utility as shown in Case (1) and given that the agent intends to win, it will invest maximum resources in generating his estimate, as shown in Lemma 2.

Having shown that combined misreporting of costs, estimates of qualities and their precision leads to either negative utility or a non-optimal outcome, we proved that the mechanism is immune to this type of strategic behaviour. □

Theorem 2. *The mechanism is individually rational for the winning agent.*

Proof. The utility an agent which has truthfully reported his estimates, his precisions and the productions costs is given by:

$$\overline{U}(y) = V(y) - x(y) - S_{(2)} = S_{(1)} - S_{(2)}$$

Given that $V(y) - x(y)$ is the selected agent's true score, the expected utility is positive and consequently the mechanism individually rational. □

6 Numerical Evaluation

In this section we initially demonstrate how this mechanism works through an example and proceed to undertake a series of simulations to get a better understanding of its performance. In order to highlight the costs of the uncertainty regarding the agents' predictions of their output we use two benchmark cases. In more detail, in the case 'Second Score: Outcome', we compare our mechanism with the standard second score auction (SSA) under the assumption that there is no uncertainty and agents can directly report their actual productions y_0^i. In the case labelled as 'Second Score: Belief' we introduce uncertainty in the model and compare our mechanism with the second score auction using the agents' beliefs of their qualities instead of the actual outcomes of the previous case, under the assumption that the principal can elicit their truthful reporting without external means.

We consider a specific case in which the parameters y_μ^i of the agents' prior beliefs of their production qualities are drawn from the uniform distribution $\mathcal{U}(2, 5)$, while we assume that the agents' precisions in both priors and individual observations during data collection are equal to 1. Consequently, given our model, the actual production quality level follows the Gaussian distribution $\mathcal{N}(y_\mu^i, 1)$. Furthermore, the agents' production cost functions are given by $x^i(y) = X^i y^2$, where $X^i \sim \mathcal{U}(0, 1)$, while the costs of data collection are linear functions, given by $c^i(\theta) = C^i\theta$, where $C^i \sim \mathcal{U}(0.001, 0.002)$. Note that the bounds in the distribution of the data collection cost parameter are selected so that even for relative large samples the overall cost is relative small compared to the actual production cost. A scenario whereby data collection cost would be higher than the production cost is not considered to be interesting nor realistic.

The principal's value function is given by $V(y) = B(1 - e^{-y})$, with $B = 20$ guaranteeing that there will be some agents with positive scores $V(y) - x(y)$ in the range of qualities we use. This particular value function is both increasing and concave and it provides some curvature, as opposed to more conventional approaches such as $V(y) = B\sqrt{y}$ which are almost linear when B is selected in order to achieve similar results in terms of the sign of the score.

The mechanism is simulated 10^5 times, while the precision of each agent's sample average, and consequently his sample of observations, M, ranges from 1 to 100. For each iteration we record the selected agent's utility, his payment, his prediction and production costs and whether the agent selected by our mechanism is the agent that would have been selected had there been no uncertainty (we refer to such a winner as a 'proper winner'). For the calculations that involve a lack of uncertainty, the agents will report their actual outcome y_0^i directly. In a given iteration all agents face underlying cost functions of the same form, but their priors, sample observations and cost parameters differ. Due to the number of iterations the standard error in the mean values plotted is in the range of 10^{-4} to 10^{-5} and thus we omit the use of errorbars for clarity.

6.1 A Snapshot of the Mechanism

For a single iteration of the mechanism, we calculate several of the mechanism's elements i.e. winners, payments and costs as the sample's precision increases. Specifically, in Table 1 we list the winner of our mechanism and the winner of the second score auction with no uncertainty (i.e. Second Score: Outcome), denoted as w and w' respectively. We also calculate the parts of the secondary payment i.e. the d function: $S(y_0; \widehat{y}, \widehat{\theta}) - \overline{S}(\widehat{\theta}^*) - 1$ and the penalty for inaccuracies: $d(y_0; \widehat{y}, \widehat{\theta})[V(\widehat{y}) - \widehat{S}_{(2)}]$, while listing the first and secondary payments of the mechanism (Steps 5 and 7 respectively), the total payment

Table 1: A single iteration of the mechanism.

θ	w	w'	d()	Penalty	1st Pay	2nd Pay	Total P	Utility	Cost ratio
1	11	14	-1.5862	-1.7624	1.1111	-3.4723	-2.3612	-2.5621	110.09
4	14	14	-0.9166	-0.5276	0.5756	-0.1340	0.4417	0.0692	49.88
10	14	14	-0.9125	-0.9636	1.0560	-0.0171	1.0389	0.6554	19.95
16	14	14	-0.8846	-0.7820	0.8840	0.1577	1.0417	0.6472	12.47
22	14	14	-0.9417	-0.6751	0.7169	0.2874	1.0043	0.5989	9.07
28	14	14	-0.8786	-0.7433	0.8460	0.2266	1.0726	0.6562	7.13
34	14	14	-0.9168	-0.9277	1.0119	0.0906	1.1024	0.6751	5.87
40	14	14	-0.9382	-0.9601	1.0233	0.0581	1.0814	0.6430	4.99
46	14	14	-0.9213	-0.9190	0.9975	0.1251	1.1225	0.6732	4.34
52	14	14	-0.8727	-0.8030	0.9201	0.3023	1.2224	0.7621	3.84
58	14	14	-0.9004	-0.7504	0.8334	0.3573	1.1907	0.7194	3.44
64	14	14	-0.8758	-0.7663	0.8750	0.3885	1.2635	0.7812	3.12
70	14	14	-0.8820	-0.7550	0.8561	0.4070	1.2631	0.7698	2.85
76	14	14	-0.8658	-0.7755	0.8958	0.4239	1.3197	0.8154	2.63
82	14	14	-0.8671	-0.7665	0.8840	0.4551	1.3391	0.8239	2.43
88	14	14	-0.8638	-0.7765	0.8990	0.4678	1.3668	0.8406	2.27
94	14	14	-0.8645	-0.7864	0.9096	0.4791	1.3887	0.8515	2.12
100	14	14	-0.8687	-0.7578	0.8724	0.5239	1.3964	0.8482	2.00

and the winner's utility. Finally, in the last column, we list the ratio between the cost of production $x(y)$ and data collection $c(\theta)$.

From Table 1 it can be seen that in this particular instance, at a sample precision of 4 our auction's winner is the winner of the second score auction with no uncertainty, $w = w'$. This shows that our mechanism identified the 'proper' winner, i.e. the agent who should have won based solely on actual production, after he generated a sample of 4 observations. However, d() function is not equal to -1, as it is on expectation, which leads to a heavier fine for the winner of the auction. Hence the 2nd Pay, total payment and utility are negative at some precisions. Specifically regarding the winner's utility, it is interesting to observe the loss of an imprecise agent, and the relation with our

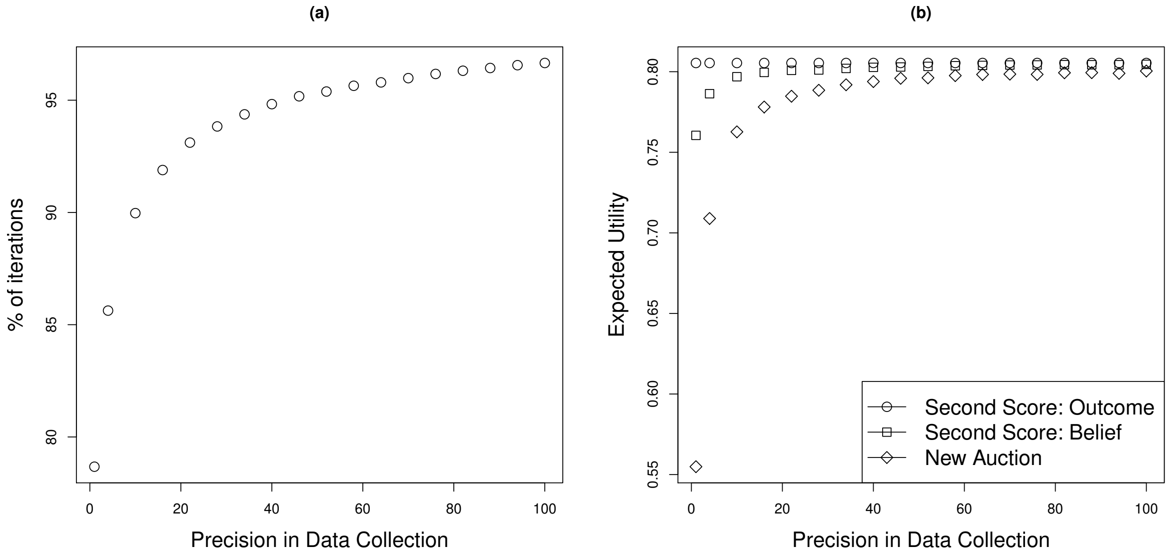

Figure 2: The selected agent's percentage of winning the auction under no uncertainty and his expected utility.

theoretical results in Section 5, where we discussed how estimates of low precision may end up winning the auction but inflicting additional losses instead of gains (Theorem 1, Case (3)). Still, despite the good intuition that this analysis provides for our mechanism, it should be noted that these results are from a single iteration, hence exposed to heavy bias from the random inputs (i.e. costs and qualities).

6.2 Numerical Simulations

Having detailed the simulation's input parameters and analysed a snapshot of the mechanism, we now present our numerical findings after simulating the mechanism 10^5 times. In Fig. 2 we summarise the behaviour of the our mechanism. It can be seen that for the specific scenario we consider, it takes a relatively small sample precision, for the outcome of our mechanism to be the same as the outcome of the second score auction with no uncertainty, where agents directly report their realised qualities (i.e. Second Score: Outcome).

In fact, after around 50 observations the winner of our auction is the winner of the second score auction in more than 95% of the iterations of the mechanism (Fig. 2(a)). In addition to this, our analytical findings in Section 5 are validated in Fig. 2(b), where we notice that the selected agent's expected utility increases as the precision of the sampling increases. The utility the winner of our auction expects to derive is less than the second score auction's winner expected utility (Second Score: Belief), had it been able to generate and report his belief of his quality freely. As it is expected, as the precision increases both auctions approach the second score auction in a setting with no uncertainty where the winner can report his actual production from the beginning (Second Score: Outcome). The differences that appear are attributed to those cases where the winners of the two auctions do not coincide, hence the winner faces losses.

The payment the selected agent expects to derive and his average costs for precision $\theta \in [1, 100]$ are shown in Fig. 3. There is a clear effect of the penalties for inaccuracies,

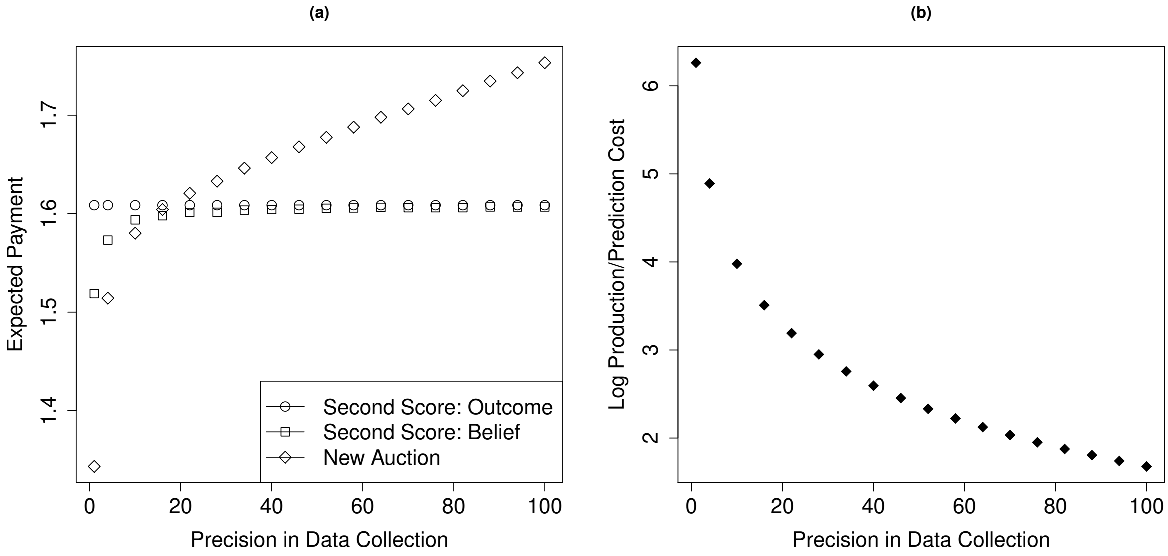

Figure 3: The selected agents' expected payment and average costs.

but also of the principal's compensation for the data collection costs in the expected payment shown (Fig. 3(a)). The expected payment for our mechanism starts lower than the two benchmark auctions, but it increases as the precision increases. The stability in the payments of the other auctions is to be expected since there is no data collection before the auction hence no compensation, while the higher payments for our auction will not be an issue in realistic applications since the data collection cost tends to be significantly lower than the production costs; also note that the payments' differences are highlighted in the plot due to its scale. In fact, this issue is related to the particular implementation of the simulations and not to the mechanism itself, since even after setting the upper bound of the cost collection parameter equal to 0.002 and using a linear cost function, for relatively high precisions that cost ends up very close to some agents' production costs. We demonstrate this data sensitivity in Fig. 3(b), where we plot the logarithmic ratio of the production to the prediction costs.

7 Conclusions

There are many benefits attached to crowd-sourcing. However, before this technology can meet its full potential there are several issues related to both crowd-sourcers' and workers' behaviour which must be addressed. For example, crowd-sourcers rarely focus on anything else than the final cost of the project, and they also lack the means to assess the workers, besides unsophisticated procedures such as discarding a completed task. On the other hand, workers expecting minimum or circumstantial rewards are inclined to dedicate the least of their time or other resources, if any at all, in completing their tasks.

In this article we present a conceptual mechanism for addressing these challenges based on a multi-dimensional procurement auction modified so it can address effectively workers' strategic behaviour. The use of a multi-dimensional auction allows crowd-sourcers to focus on other elements of the workers' output and therefore gives them incentives to improve these while balancing the costs. Furthermore, we introduced uncertainty on how workers determine their production qualities by modelling them as probabilistic estimates, and we assumed that each worker generates a sample of independent estimates of a certain precision. We further departed from standard multi-dimensional approaches by denying the crowd-sourcer of the ability to enforce truthful reporting of agents' qualities through external means (i.e. cancelling the auction or large arbitrary fines).

Initially the crowd-sourcer procures a task from the crowd by implementing a standard second score auction, only now the workers' ranking is calculated based on their reported estimates of their qualities and costs. The winner of the auction receives the second score payment and after he fulfils his part of the contract he receives a secondary payment based on both the reported estimate and the actual production, production costs and costs involved in generating the estimate. The secondary payment uses a strictly proper scoring rule to evaluate the worker's posterior belief of his quality once the task is finished and the crowd-sourcer can witness the outcome.

We showed that the mechanism is immune to workers' combined misreporting i.e. with respect to the reported estimates of their outputs and the reported costs. In addition to that, we showed that workers invest the maximum of the resources available to them when generating that estimate, while individual rationality is maintained for the winner of the auction.

However, there are some limitations regarding practical elements of the mechanism. Although we proved analytically that our mechanism implements the standard second score auction's outcome in terms of the selected worker's expected utility, numerical simulations demonstrated how sensitive the mechanism is to the prediction of the worker's quality, and hence to the resources invested in preparing his bid. The importance of this issue is highlighted by observations which suggest that crowd-sourcers may attempt to manipulate the workers' restrictions during their preparation stage. Specifically, in MTurk crowd-sourcers may attempt to manipulate the position of their task in the search queries, in order to attract more workers and hence have the task completed faster and for less money. This strategy can be effective, at least for the crowd-sourcers, given that workers rarely browse after page 10 in the search results[9]. Although our mechanism includes incentives for investing sufficient resources at the preparation stage (i.e. data collection), it is only the winner of the auction who is compensated with the cost of generating his estimate. To overcome these problems, workers must be rewarded so they can participate in the mechanism, irrespective of if they win or not. These rewards could be monetary or special privileges, such as moderator status, cosmetic customisation or certificates of specialisation. Experience of the past combined with the rate of advancements in internet based commerce suggests that these drawbacks will be addressed.

References

[1] D. R. Beil and L. Wein. An inverse-optimization-based auction mechanism to support a multiattribute rfq process. *Management Science*, pages 1529–1545, 2003.

[2] M. Bichler. An experimental analysis of multi-attribute auctions. *Decision Support Systems*, 29:249–268, 2000.

[3] M. Bichler and J. Kalagnanam. Configurable offers and winner determination in multi-attribute auctions. *European Journal of Operational Research*, 160:380–394, 2003.

[4] P. Bogetoft and K. Nielsen. Dea based auctions. *European Journal of Operational Research*, 184:685–700, 2008.

[5] F. Branco. The design of multidimensional auctions. *RAND Journal of Economics*, 28(1):63–81, 1007.

[6] A. Charnes, W. W. Cooper, and E. Rhodes. Short communication: Measuring the efficiency of decision making units. *European Journal of Operational Research*, 3:339, 1979.

[7] A. Charnes, W. W. Cooper, and E. Rhodes. Measuring the efficiency of decision making units. *European Journal of Operational Research*, 2(6):429–444, 1987.

[8] Yeon-Koo Che. Design competition through multidimensional auctions. *RAND Journal of Economics*, 24(4):668–680, 1993.

[9] L. B. Chilton, J. J. Horton, R. C. Miller, and S. Azenkot. Task search in a human computation market. In *Proceedings of the ACM SIGKDD Workshop on Human Computation*, HCOMP '10, pages 1–9, 2010.

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

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

[12] A. Doan, R. Ramakrishnan, and A. Y. Halevy. Crowdsourcing systems on the world-wide web. *Communications of the ACM*, 2011.

[13] E. Estellés-Arolas and F. González-Ladrón-de Guevara. Towards an integrated crowdsourcing definition. *Journal of Information Science*, 38(2):189–200, 2012.

[14] D. Friedman. Effective scoring rules for probabilistic forecasts. *Management Science*, 29(4):447–454, 1983.

[15] P. C. Gregory. *Bayesian Logical Data Analysis for the Physical Sciences: A Comparative Approach with Mathematica Support*. Cambridge University Press, 2005.

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

[17] M. He, N. R. Jennings, and H.-F. Leung. On agent-mediated electronic commerce. *IEEE Transactions on knowledge and data engineering*, 15(4):985–1002, 2003.

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

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

[20] 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.

[21] 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.

[22] 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.

[23] A. Kittur, E. H. Chi, and S. Bongwon. Crowdsourcing user studies with mechanical turk. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '08, pages 453–456, 2008.

[24] A. Kittur, B. Smus, S. Khamkar, and R. E. Kraut. Crowdforge: crowdsourcing complex work. In *Proceedings of the 24th annual ACM symposium on User interface software and technology*, UIST '11, pages 43–52, 2011.

[25] V. Krishna. *Auction Theory.* Academic Press, 2002.

[26] A. Mas-Collel, M. D. Whinston, and J. R. Green. *Mircoeconomic Theory.* Oxford University Press, 1995.

[27] N. Miller, P. Resnick, and R. Zeckhauser. Eliciting honest feedback: The peer prediction method. *Management Science*, 51(9):1359–1373, 2005.

[28] 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, 2007.

[29] A. Papakonstantinou. *Mechanism Design for Eliciting Costly Observations in Next Generation Citizen Sensor Networks.* PhD thesis, University of Southmapton, School of Electronic and Computer Science, 2010.

[30] A. Papakonstantinou, A. Rogers, E. H. Gerding, and N. R. Jennings. Mechanism design for the truthful elicitation of costly probabilistic estimates in distributed information systems. *Artificial Intelligence*, 175(2):648–672, 2011.

[31] D. Parkes and J. Kalagnanam. Models for iterative multiattribute procurement auctions. *Management Science*, 51(3):435–451, 2005.

[32] Sarvapali D. Ramchurn, D. Huynh, and N. R. Jennings. Trust in multi-agent systems. *The Knowledge Engineering Review*, 19:1–25, 2004.

[33] L. J. Savage. Elicitation of personal probabilities and expectations. *Journal of the American Statistical Association*, 66(336):783–801, 1977.

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

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

[36] J. Witkowski and D. C. Parkes. Peer prediction without a common prior. In *Proceedings of the 13th ACM Conference on Electronic Commerce*, EC '12, pages 964–981, 2012.

[37] J. Yang, L. A. Adamic, and M. S. Ackerman. Crowdsourcing and knowledge sharing: strategic user behavior on taskcn. In *Proceedings of the 9th ACM conference on Electronic commerce*, EC '08, pages 246–255. ACM, 2008.