

Optimum Statistical Estimation with Strategic Data Sources

Yang Cai*

Computer Science, McGill University

CAI@CS.MCGILL.CA

Constantinos Daskalakis†

EECS, MIT

COSTIS@MIT.EDU

Christos Papadimitriou‡

Computer Science, U.C. Berkeley

CHRISTOS@CS.BERKELEY.EDU

Abstract

We propose an optimum mechanism for providing monetary incentives to the data sources of a statistical estimator such as linear regression, so that high quality data is provided at low cost, in the sense that the weighted sum of payments and estimation error is minimized. The mechanism applies to a broad range of estimators, including linear and polynomial regression, kernel regression, and, under some additional assumptions, ridge regression. It also generalizes to several objectives, including minimizing estimation error subject to budget constraints. Besides our concrete results for regression problems, we contribute a mechanism design framework through which to design and analyze statistical estimators whose examples are supplied by workers with cost for labeling said examples.

1. Introduction

Statistical estimation, from data, of the parameters of a model of reality, is the spirit, essence, and workhorse of modern science and business. In today’s complex world of science and industry, data for an estimator deployed by a particular research group or enterprise is often provided by other entities; furthermore, the quality of the data is crucial for the accuracy of the estimator, and can vary widely. It is reasonable to assume that, with appropriate effort and cost, the providers of the data can improve the quality of their data — but of course they may lack incentive to do so. Crowdsourcing (9) can be seen as a popular and widespread instantiation of the phenomenon.

The situation is not unlike *Mechanism Design*, a well developed field in Mathematical Economics (32), and indeed related problems in connection to crowdsourcing have been recently treated within this framework; see the references in Appendix A. In Mechanism Design, in order for the interaction of the designer with several rational strategic agents to be as beneficial as possible, a game between the agents is created in which the pursuit by the agents of individual advantage leads to the optimum outcome for the designer. Perhaps the archetypical Mechanism Design problem, solved in Myerson’s celebrated work (34), is how to auction an item to a number of agents, whose values for the item are unknown but drawn from known prior distributions. Myerson’s is a power-

‡ Supported by NSERC Discovery Grant 06127-15.

‡ Supported by Microsoft Research faculty fellowship, NSF Awards CCF-0953960 (CAREER) and CCF-110149, and ONR grant N00014-12-1-0999.

‡ Supported by NSF award CCF-0964033 and a Google university research award.

ful, sophisticated, and clean result, and has had tremendous impact—a useful ideal to keep in mind when venturing into new areas.

Coming back to estimation, suppose that a Statistician has an algorithm which, given appropriate data points $X = \{(x_i, y_i)\}$, approximates an unknown function f that captures an important aspect of reality, and can be usefully employed in prediction. For concreteness, let us say that the unknown function is linear and the context is one-dimensional linear regression (noting that our results are far more general). The quality of the output of the algorithm, \hat{f} , depends crucially on the quality of the data X . If the data is poor, the Statistician’s prediction $\hat{f}(x^*)$ —where x^* is the unknown test point where a prediction will have to be made—will be off the mark. Assume that the loss suffered is $(f(x^*) - \hat{f}(x^*))^2$. How can the Statistician incentivize her data sources — suppose for simplicity that each of them provides one of the (x_i, y_i) points, for an x_i selected by the Statistician — to work hard so as to supply high-quality data?

Assume that each data provider is a “worker” who, by expending effort e , achieves a level of quality of the data; the larger e , the higher the quality. To fix ideas, assume that, by expending effort e , the i th worker is able to sample a distribution for his target data y_i that is centered at the true value $f(x_i)$ and has variance $\sigma_i(e)^2$. That is, the variance of the error of the datapoint is our measure of the worker’s quality (rather, lack thereof), and it is a monotonically decreasing function of e . Of course there are many workers, and they may have different functions $\sigma_i(e)$. *We assume that these functions are common knowledge* to the participants, including of course the Statistician. In fact, as it will become clear, each worker needs to know this information only to decide whether to participate in the mechanism, and not to decide on his optimum effort level. The question we explore is this: *Can the Statistician create a mechanism—or protocol, or contract—which incentivizes through appropriate payments the workers to supply high-quality data?* We assume, naturally enough, that each worker acts so as to maximize the expectation of payment received minus the effort exerted (By rescaling, we can assume that units of loss by the Statistician, effort by a worker, and currency all coincide). Naturally, this expectation must be nonnegative, because otherwise the worker will refuse to work.

The Statistician, of course, does not see the effort exerted by the i th worker, but only the result y_i of this effort. Evidently, the payment in the protocol, or contract, being designed should be determined by comparing each supplied y_i to the true value $f(x_i)$. But of course this is impossible *because we will never learn $f(\cdot)$* . The payments must somehow depend on what we know, which is the data, the algorithm, the quality functions of the workers, and the prior of the test point x^* , but importantly not f .

And what should such a mechanism aspire to achieve? Let us define an ideal performance which, a priori, may seem unreasonably optimistic. The quality of our estimation is going to depend on the following:

- the set W of workers that we select from the available workers;
- the set $\{x_i : i = 1, \dots, |W|\}$ of the regression points we choose to assign to each of them;
- on the effort e_i each worker decides to invest (which is out of our control).

Once these have been determined, then we also know the variance $\sigma_i(e_i)^2$ of each data point, and the expectation of the loss to the statistician is a known function $L(W, \{x_i\}, \{e_i\})$. Define OPT to be the minimum, over all possible $W, \{x_i\}, \{e_i\}$, of the quantity $L(W, \{x_i\}, \{e_i\}) + \sum_{i \in W} e_i$.¹

1. We can handle much more than this objective, as discussed later in this section.

That is, OPT is the optimum of loss plus effort, what is known in Economics as the social optimum (best possible sum of costs by all participants). But OPT is also a lower bound on the total cost (loss plus payments) to the Statistician: To achieve total cost OPT, the Statistician must convince each worker to supply the precise optimum effort, and to do so *at zero surplus*, that is, by being paid the smallest possible amount for this effort (recall that the worker will refuse to work if payment minus effort has negative expectation). But this is hard to achieve, given that the efforts exerted by the workers are not observable and out of the Statistician’s control.

So, *is there such a mechanism?* In the light of the difficulties outlined in the last two paragraphs, the situation may seem quite hopeless.

Very surprisingly, we show that such a mechanism does exist not only for linear regression, but also for a broad class of estimators satisfying a simple and intuitive condition, and includes, for example, polynomial regression, finite dimensional kernel regression and many other linear estimators. The solution engages workers in a game not with the Statistician, but *with each other*. The game is constructed in such a way that, in the end, the optimum outcome is achieved, in that loss plus payments is minimized, and all “surplus” of the workers is extracted (that is, no worker is paid more than his work).

One important consideration is what we mean above by “in the end,” that is, what are our assumptions on how workers will behave? *What is our solution concept?* After all, solution concepts are known to be delicate and fragile, the subject of endless discussions and controversy in Game Theory. It turns out that our design can afford a solution concept that is extremely robust and uncontroversial: Each worker’s decision is a *unique dominant strategy*, that is, it is the unique action that optimizes the worker’s objective (payment minus effort) *no matter what anybody else does*. Of course, this restricts severely the design space of mechanisms that we can employ, and despite this restriction we can still design a mechanism that attains the ideal total cost OPT.

One key idea of the mechanism is that, in the game as defined, the payment of each worker depends on the data supplied by *the other workers*. To some this may seem unreasonable, while to others it may seem a bit unsurprising in view of the VCG mechanism, among others, which has a similar structure (45; 6; 13). That this maneuver works in this instance is an entirely new phenomenon related to statistical estimation, as opposed to mechanism design, and comes with a different mathematical justification, even though in a sense it does have the same roots as VCG: Our design essentially creates a *race for accuracy*, in which workers compete knowing that they will fare badly if left behind.

Our mechanism applies to all statistical estimators that satisfy a certain intuitive property: The expectation of the estimator’s loss depends only on the data points $\{x_i\}$ and the distribution of the test point x^* , and depends on the y_i ’s *only through their variances*. There are two variants of the problem, depending on how the x_i ’s are determined. So far, we have assumed that the x_i ’s should be optimally chosen by the algorithm/mechanism, and the estimators we can handle already include, e.g., linear regression, polynomial regression, finite-dimensional kernel regression, and several other linear estimators. In a model where the x_i ’s are fixed and given in advance (but the algorithm should still optimally assign them to workers) the situation is even more favorable: our technique applies to an even broader class of estimators, including *ridge regression* (44). This is discussed in Section 3.2. In the same section, we also discuss several extensions of our results. Besides a weighted combination of loss plus payments, we can accommodate objectives that aim at minimizing loss subject to budget constraints, or loss plus a function of the payments, and can also go beyond mean square error to other loss functions.

Related work is discussed in Appendix A due to space constraints.

2. Estimation with Strategic Workers: the Model

In this section, we introduce the statistical estimation task that we solve in the next section. We start with some standard definitions.

Definition 1 (Estimator) *Let \mathcal{H} be a family of functions $f : \mathcal{D} \rightarrow \mathbb{R}$, where $\mathcal{D} \subseteq \mathbb{R}^n$. An estimator for \mathcal{H} takes as input a collection $(x_i, y_i)_{i=1}^k$ of examples $(x_i, y_i) \in \mathcal{D} \times \mathbb{R}$ and produces an estimated function $\hat{f}_{(x_i, y_i)_{i=1}^k} \in \mathcal{H}$.*

For example, \mathcal{H} may be the class of linear functions from \mathbb{R}^n to \mathbb{R} , in which case the estimator could be linear regression. If the input $(x_i, y_i)_{i=1}^k$ to the estimator is clear from context, we may omit it from the subscript of \hat{f} . Sometimes we use (\mathbf{x}, \mathbf{y}) as a shorthand for $(x_i, y_i)_{i=1}^k$, and use $\hat{f}_{(\mathbf{x}, \mathbf{y})}$ to denote the estimator. We may also use the shorthand \hat{f}_{-j} or $\hat{f}_{(\mathbf{x}, \mathbf{y})_{-j}}$ for the output of the estimator when given all examples except (x_j, y_j) ; i.e. $\hat{f}_{-j} \triangleq \hat{f}_{(x_i, y_i)_{i \in \{1, \dots, k\} \setminus \{j\}}} \triangleq \hat{f}_{(\mathbf{x}, \mathbf{y})_{-j}}$. This is assuming that our estimator is well-defined with one example omitted. Whenever we use this notation we assume that our estimator satisfies this property. We call \hat{f}_{-j} the *estimator \hat{f} with one example less*.

In estimation it is usually assumed that the examples are readily available. Here we study the scenario in which we choose a collection $x_1, \dots, x_k \in \mathcal{D}$ of points and assign them to experts, who then return estimates of the function f at those points. How good will these estimates be? In this paper we assume that the experts, or *workers*, are *strategic*; for example, they will put no effort into producing good estimates of the function value at their given point, unless they are provided monetary incentives to exert such effort. We capture the behavior of such strategic experts in the following definition.

Definition 2 (Strategic Worker) *Let $f \in \mathcal{H}$ be as in Definition 1. A worker for f is a strategic agent who, given some query $x \in \mathcal{D}$, will decide how much effort $e \in \mathcal{E} \subseteq \mathbb{R}_+$ to exert in order to produce an estimate $y(e)$ of $f(x)$.² The worker:*

- is characterized by some known strictly decreasing convex function $\sigma : \mathcal{E} \rightarrow \mathbb{R}_+$ such that, whenever effort e is exerted, the estimate produced satisfies:

$$y(e) = f(x) + \epsilon,$$

where ϵ is distributed according to some (potentially unknown) distribution with mean 0 and variance $\sigma(e)^2$.

- aims to minimize the amount of exerted effort to produce the estimate of $f(x)$, unless provided monetary incentives to do otherwise; in particular, if the worker is promised a payment function $p : \mathcal{D} \times \mathbb{R} \rightarrow \mathbb{R}$ that assigns to each pair (x, y) of a query point x and estimate y a dollar amount $p(x, y)$, then the worker will choose to exert an amount of effort

$$e^* \in \arg \max_{e \in \mathcal{E}} \mathbb{E}[p(x, y(e))] - e,$$

where the expectation is taken with respect to the randomness in y and the randomness in the payment function, if any.³

2. Note that we have omitted the dependence of y on x in our notation to ease notation.

3. Note again that we have omitted the dependence of e^* on x and the payment function $p(\cdot)$ in our notation.

The following definition formulates the problem of estimating an unknown function $f \in \mathcal{H}$, when one's only access to f is through strategic workers for f .

Definition 3 (Estimation with Strategic Workers (ESW)) *Suppose that we are given:*

- an estimator \hat{f} for a family of functions \mathcal{H} , as in Definition 1;
- access to a set \mathcal{W} of strategic workers for some unknown function $f \in \mathcal{H}$, as in Definition 2, where each worker $i \in \mathcal{W}$ has a known function σ_i mapping effort to accuracy; we also assume that all workers' estimations are independent;
- a distribution F over \mathcal{D} (the distribution of the test point $x^* \in \mathcal{D}$).

Our goal is to:

1. choose some subset $\mathcal{W}' \subseteq \mathcal{W}$ of workers
2. provide an input x_i to each worker $i \in \mathcal{W}'$, requesting an estimate y_i of $f(x_i)$ from i
3. commit to a payment function p_i to each $i \in \mathcal{W}'$, where p_i is a (potentially) randomized mapping $p_i : (x_i, y_i)_{i \in \mathcal{W}'} \mapsto \mathbb{R}$, which may depend not only on the estimate produced by worker i but also the estimates produced by the other workers.

Subject to our decisions in 1, 2 and 3, we are looking to minimize a weighted average of the mean-square error of our estimation \hat{f} and the expected payments made to the workers, namely:

$$\mathbb{E}_{x^*, \mathbf{y}(e^*)} \left[\left(\hat{f}_{(\mathbf{x}, \mathbf{y}(e^*))}(x^*) - f(x^*) \right)^2 + \eta \cdot \sum_{i \in \mathcal{W}'} p_i((x_j, y_j(e_j^*))_{j \in \mathcal{W}'}) \right], \quad (1)$$

for some $\eta > 0$, where the expectation is taken with respect to all the randomness in the setting: the randomness in $x^* \sim F$, the randomness in the outputs $\mathbf{y}(e^*) \triangleq \{y_i(e_i^*)\}_{i \in \mathcal{W}'}$ produced by the workers, and the randomness in the payment functions. For (1) to be a well-defined objective, we need to be able to predict the efforts $(e_i^*)_{i \in \mathcal{W}'}$ that the workers of our selected set will exert given our decisions for 1, 2 and 3. We discuss how this can be achieved in Section 2.1 below. At the very least, our prediction needs to satisfy the individual rationality constraint of Definition 6, i.e. that the expected payment to each worker i is at least as large as e_i^* , otherwise the worker would not participate.

Remark 4 (Variants)

1. In Definition 3, we assume that the designer optimizes over the selection of the points x_i . Alternatively, it could be that each expert i comes with a point x_i , which is “his expertise,” or that the x_i 's are predetermined but the designer can still decide how to assign them to experts. In Section 3.2, we briefly discuss this variant of our problem, which yields a much richer class of estimators for which our result, described in the next section, applies.
2. We also assumed that the objective is to minimize a weighted average of mean-square error and expected payments. While we stick to this objective for the development of our mechanism in the next section, our techniques go through with minimal modifications to much more general objectives. For example, we can accommodate the problem of minimizing mean-square error subject to a budget constraint, or minimizing the sum of mean-square error and an arbitrary increasing function of payments. We discuss these extensions in Section 3.2.

3. Finally, our technique is general enough to even eliminate the use of mean-square error from the objective, an extension that we also discuss in Section 3.2.

2.1. Incentives

We have already noted inside Definition 3 that for (1) to be a well-defined objective, we need to be able to predict the efforts $(e_i^*)_{i \in \mathcal{W}'}$ that the workers will exert as a result of our solution to ESW. It is important to note that the form of the payment functions (Decision 3 in Definition 3) couples the decision of each worker i about the amount of effort he exerts with the amounts of effort the other workers exert (since these influence \mathbf{y}_{-i}), which themselves depend on y_i and hence the effort that worker i exerts. This cyclical dependence is familiar in Game Theory. Indeed, any solution to ESW—comprising a subset \mathcal{W}' of workers, queries $(x_i)_{i \in \mathcal{W}'}$ to them, and payment commitments $(p_i)_{i \in \mathcal{W}'}$ —induces a *game* among the workers in \mathcal{W}' ; the effort levels $(e_j^*)_{j \in \mathcal{W}'}$ eventually chosen by the workers are the outcome of their strategic interaction in this game. Therefore, to be able to evaluate (1), we need to predict how agents will behave in this game.

There are numerous *solution concepts* in Game Theory, whose goal is to close into the possible behavior of rational players in a game. The prominent ones are Nash equilibrium and its several refinements. However, Nash equilibria may be randomized and, most problematically, they may be multiple, which would result into equilibrium selection issues in our setting. To avoid such issues and guarantee robustness of our solutions, we insist on the most compelling and uncontroversial solution concept, namely that of *Unique Dominant Strategy Equilibrium*, defined next.

Definition 5 (Unique Dominant Strategy Equilibrium) *A solution to ESW—comprising a subset \mathcal{W}' of workers, queries $(x_i)_{i \in \mathcal{W}'}$, and payment commitments $(p_i)_{i \in \mathcal{W}'}$ —induces a unique dominant strategy equilibrium $(e_i^*)_{i \in \mathcal{W}'}$ iff, for all $i \in \mathcal{W}'$ and all $(e_j)_{j \in \mathcal{W}'}$:*

$$\mathbb{E} \left[p_i \left((x_i, y_i(e_i^*)), (x_j, y_j(e_j))_{j \in \mathcal{W}' \setminus \{i\}} \right) \right] - e_i^* \geq \mathbb{E} \left[p_i \left((x_j, y_j(e_j))_{j \in \mathcal{W}'} \right) \right] - e_i,$$

where the expectation is with respect to everything that is random, with equality only if $e_i = e_i^*$. In words, no matter what effort levels the other workers choose, the unique optimal effort level of every worker i is e_i^* .

Note that, if a unique dominant strategy equilibrium exists in a game it is fairly trivial for agents to decide what strategy to play. We will insist that our solution to ESW should induce a game that has a unique dominant strategy equilibrium. Note that, in general, it is very rare for a game to have such an outcome, and consequently this poses a significant constraint on our design. While our main result (Theorem 9) satisfies this constraint, it nevertheless does not sacrifice any objective value, fairing as well as it would without this constraint present; see Remark 7 below and the discussion following the statement of Theorem 9.

Finally, as we have already noted in Definition 3, not all combinations of solutions to ESW and predictions of worker behavior are realistic. Since the workers are assumed strategic and their participation is voluntary, they should not be making a loss when participating. This is captured by the following definition, adding an additional requirement to our solutions to ESW.

Definition 6 (Individual Rationality) *Given a solution to ESW—comprising a subset \mathcal{W}' of workers, queries $(x_i)_{i \in \mathcal{W}'}$, and payment commitments $(p_i)_{i \in \mathcal{W}'}$, a collection of efforts $(e_i^*)_{i \in \mathcal{W}'}$ satisfies*

individual rationality *iff*, for all workers $i \in \mathcal{W}'$,

$$\mathbb{E} \left[p_i \left((x_j, y_j(e_j^*))_{j \in \mathcal{W}'} \right) \right] - e_i^* \geq 0.$$

Remark 7 (Robustness vs Optimality Non-Tradeoffs) *We have chosen to restrict our attention to solutions to ESW that induce a game among workers with a unique dominant strategy equilibrium that satisfies individual rationality. The worry might be that such a strong requirement might sacrifice too much objective value. We will show that it does not, in a very strong sense. In particular, while this requirement is handicapping our solutions to ESW, we still compare our solution's performance against any other solution evaluated at the most favorable for that solution collection $(e_i^*)_i$, with the minimum requirement that the $(e_i^*)_i$ satisfy the individual rationality constraint.*

3. Optimal Estimation with Strategic Workers: the Mechanism

Our main contribution in this paper is to establish the existence of an optimal solution to ESW, which induces a unique dominant strategy equilibrium and which has very strong optimality guarantees, as discussed later in this section, for a broad class of estimators \hat{f} , containing several familiar ones:

Definition 8 *An estimator \hat{f} for \mathcal{H} , as in Definition 1, is well-behaved *iff* there exists some function g ⁴ such that, for all distributions F over \mathcal{D} , functions $f \in \mathcal{H}$, and vectors $\mathbf{x} \in \mathcal{D}^*$ and $\boldsymbol{\sigma} \in \mathbb{R}_+^*$ (of the same dimension as \mathbf{x}).⁵*

$$\mathbb{E}_{\mathbf{y}, \mathbf{x}^*} \left[\left(\hat{f}_{(\mathbf{x}, \mathbf{y})}(x^*) - f(x^*) \right)^2 \right] = g(\mathbf{x}, F, \boldsymbol{\sigma}),$$

where for the purposes of the expectation on the left hand side $x^* \sim F$ and, independently for all i , $y_i = f(x_i) + \epsilon_i$, where each ϵ_i is sampled from an arbitrary distribution of mean 0 and variance σ_i^2 , (and when \mathbf{x} is such that $\hat{f}_{(\mathbf{x}, \mathbf{y})}$ is well-defined).

Note that several common estimators, such as linear regression, polynomial regression, finite-dimensional kernel estimation, are well-behaved according to our definition.⁶

Our main result is the existence of an optimal algorithm for ESW, whenever \hat{f} is well-behaved, according to Definition 8, and well-defined with one example less.⁷ We note that the first condition is a sufficient condition and discuss how to relax it in Section 3.2. The second condition is necessary for interesting solutions to ESW, and we discuss how it can be removed by broadening the set of allowable payment functions in Section 3.2. Our algorithm optimally solves ESW in a rather strong sense, as captured by Properties 2 and 3 in the following theorem.

4. Intuitively, an estimator \hat{f} is *well-behaved* *iff* its expected mean square error does not depend on the function f .

5. We use the shorthand $\mathcal{D}^* \triangleq \bigcup_{i=1}^{\infty} \mathcal{D}^i$, and similarly for \mathbb{R}_+^* .

6. For instance, if \hat{f} is linear regression then

$$g(\mathbf{x}, F, \boldsymbol{\sigma}) = \mathbb{E}_{\mathbf{x}^* \sim F} \left[[(x^*)^T, 1] \cdot (X^T X)^{-1} X^T \cdot \text{diag}(\boldsymbol{\sigma}^2) \cdot X (X^T X)^{-1} \cdot [(x^*)^T, 1]^T \right],$$

where X is the matrix whose rows are $[x_i^T, 1]$ and $\text{diag}(\boldsymbol{\sigma}^2)$ is the diagonal matrix whose (i, i) entry is σ_i^2 .

7. An estimator is not well-defined with one example less if omitting one example from its input makes the output of the estimator undefined. For example, if \hat{f} is linear regression (for linear functions from \mathbb{R}^n to \mathbb{R}) restricted to take as input exactly $n + 1$ examples (x_i, y_i) , then it is not well-defined with one example less, since n examples won't suffice to produce an estimate.

Theorem 9 *There exists an optimal algorithm for ESW for all well-behaved estimators \hat{f} that are well-defined with one example less. The algorithm:*

1. *produces a solution to ESW that induces a unique dominant strategy equilibrium that satisfies individual rationality;*
2. *under the unique dominant strategy equilibrium the solution achieves objective value (1) that is optimal; in fact, the achieved objective value matches the following quantity:*

$$\min_{\mathcal{W}', (x_i, e_i)_{i \in \mathcal{W}'}} \left(\mathbb{E}_{x^*, \mathbf{y}} \left[\left(\hat{f}_{(\mathbf{x}, \mathbf{y})}(x^*) - f(x^*) \right)^2 \right] + \eta \cdot \sum_i e_i \right), \quad (2)$$

where for the purposes of the expectation we assume that, for all $i \in \mathcal{W}'$, $y_i = f(x_i) + \epsilon_i$, where ϵ_i is sampled according to worker i 's distribution when s/he exerts effort e_i (which has mean 0 and variance $\sigma_i(e_i)^2$);

3. *extracts optimal worker surplus; in particular, the expected utility of every worker is 0 at the unique dominant strategy equilibrium.*

Notice that Quantity (2) clearly provides a lower bound to the objective value (1) of *any* solution to ESW (and not just those inducing a unique dominant strategy equilibrium), evaluated at *any* $(e_i^*)_{i \in \mathcal{W}'}$ that satisfies individual rationality. Indeed, by individual rationality for all workers combined,

$$\mathbb{E} \left[\sum_{i \in \mathcal{W}'} p_i \left((x_j, y_j(e_j^*))_{j \in \mathcal{W}'} \right) \right] \geq \sum_i e_i^*.$$

Hence, (1) \geq (2). In fact, (2) corresponds to the objective value that one would achieve, if one could dictate the effort level that each worker should exert and only paid workers exactly for the amount of effort they exerted and not a cent more. So, in fact, Property 2 implies Property 3 in our theorem above. What our theorem establishes is that there always exist solutions to ESW that induce a unique dominant strategy equilibrium satisfying individual rationality, and that these solutions achieve the same objective value that a dictator who could dictate the behavior of each worker would achieve. Even though we do not assume we have such power, we still achieve the same objective value that such a powerful dictator would.

Proof of Theorem 9: We design a solution to ESW whose unique dominant strategy equilibrium e^* satisfies individual rationality, and achieves objective value (1) that equals Quantity (2). We have already argued that if we do this, we immediately satisfy Properties 1, 2 and 3 in the statement of the theorem.

We define our solution to ESW in terms of an arbitrary optimal solution $\mathcal{W}', (x_i, e_i)_{i \in \mathcal{W}'}$ to the minimization problem (2). In terms of this solution:

- We choose the same set of workers \mathcal{W}' and assign to each $i \in \mathcal{W}'$ the point x_i .
- It remains to define our payment commitments to the workers. To each worker $i \in \mathcal{W}'$, we commit to the payment:⁸

$$p_i((\mathbf{x}, \mathbf{y})) = c_i - d_i \cdot \left(y_i - \hat{f}_{(\mathbf{x}, \mathbf{y})-i}(x_i) \right)^2, \quad (3)$$

8. For compactness, we denote by $(\mathbf{x}, \mathbf{y}) = (x_i, y_i)_{i \in \mathcal{W}'}$.

for some c_i, d_i to be chosen. Notice that the payment to worker i depends also on the reports of the other workers.

We now choose the constants $(c_i, d_i)_{i \in \mathcal{W}'}$ so that our solution induces a unique dominant strategy equilibrium e^* that satisfies individual rationality (with equality) and also $e^* \equiv e$, where $e = (e_i)_{i \in \mathcal{W}'}$ is as in the solution to (2) that we have fixed. What is the expected payment to worker i if the workers exert some arbitrary efforts e' ? Denoting $\mathbf{y}(e') = (y_i(e'_i))_{i \in \mathcal{W}'}$, we have:

$$\begin{aligned} & \mathbb{E}_{\mathbf{y}(e')} [p_i((\mathbf{x}, \mathbf{y}(e')))] \\ &= c_i - d_i \cdot \mathbb{E}_{\mathbf{y}(e')} \left[(y_i - f(x_i))^2 + \left(f(x_i) - \hat{f}_{(\mathbf{x}, \mathbf{y})_{-i}}(x_i) \right)^2 - 2(y_i - f(x_i)) \left(f(x_i) - \hat{f}_{(\mathbf{x}, \mathbf{y})_{-i}}(x_i) \right) \right] \\ &= c_i - d_i \cdot (\sigma_i(e'_i)^2 + g(\mathbf{x}_{-i}, \mathbf{1}_{x_i}, \boldsymbol{\sigma}_{-i}(e'_{-i}))), \end{aligned}$$

where we used that our estimator \hat{f} is well-behaved, according to Definition 8, and the independence of the estimation of worker i and the other workers. We denote by $\boldsymbol{\sigma}_{-i}(e'_{-i}) = (\sigma_j(e'_j))_{j \neq i}$, and by $\mathbf{1}_{x_i}$ the distribution that samples x_i with probability 1. g is a known function determined by the estimator according to Definition 8.

Since each worker $i \in \mathcal{W}'$ is assumed rational, aiming to maximize his expected payment minus exerted effort, if the other workers exert effort levels e'_{-i} , worker i 's best response is found by solving the maximization problem:

$$\max_{e'_i} (c_i - d_i \cdot (\sigma_i(e'_i)^2 + g(\mathbf{x}_{-i}, \mathbf{1}_{x_i}, \boldsymbol{\sigma}_{-i}(e'_{-i}))) - e'_i). \quad (4)$$

Taking derivative with respect to e'_i and setting to 0 gives the following condition for the optimum e_i^* :

$$2d_i \sigma_i(e_i^*) \cdot \sigma'_i(e_i^*) + 1 = 0.$$

In order to ensure that $e_i^* \equiv e_i$, where e_i was the effort level computed by solving (2), we set

$$d_i = \frac{-1}{2\sigma_i(e_i)\sigma'_i(e_i)}.$$

Given that $\sigma_i(\cdot)$ is convex decreasing, our setting of d_i ensures that e_i is the unique solution to (4).

So our choice of $(d_i)_{i \in \mathcal{W}'}$ has made sure that the unique dominant strategy equilibrium of the game among workers defined by our solution to ESW (regardless of the c_i 's) is $(e_i)_{i \in \mathcal{W}'}$. Now we set the c_i 's so that this equilibrium also satisfies individual rationality tightly. It suffices to choose, for each $i \in \mathcal{W}'$:

$$c_i = d_i \cdot (\sigma_i(e_i)^2 + g(\mathbf{x}_{-i}, \mathbf{1}_{x_i}, \boldsymbol{\sigma}_{-i}(e_{-i}))) + e_i.$$

This choice makes sure that the expected payment to each worker equals his effort. Hence, the unique dominant strategy equilibrium of the game among workers defined by our solution to ESW is e and it satisfies:

$$\mathbb{E} \left[\sum_{i \in \mathcal{W}'} p_i((x_j, y_j(e_j))_{j \in \mathcal{W}'}) \right] = \sum_i e_i.$$

Hence, the objective value (1) achieved by the unique dominant strategy equilibrium matches (2). This concludes the proof of the theorem. ■

3.1. The Computational Complexity of Our Algorithms

An important question is, of course, how the Statistician could arrive at the optimum mechanism in a computationally efficient manner. From the proof of Theorem 9, it is clear that the major computational overhead is finding an optimal solution to the minimization problem (2). All the other steps of our algorithm can be executed in polynomial time. In this section, we study the computational complexity of the minimization problem (2).

A closely related problem to ours, *optimal experiment design (ODE)*, has received lots of attention in the Statistics community (20; 23; 24; 21; 25; 22; 37). In ODE, the goal is also to find an optimal set of regression vectors $\{x_i\}$ for estimation of their corresponding $\{f(x_i)\}$ such that a certain objective is optimized. Depending on the objective, different criteria of optimality have been proposed e.g. A-optimality, C-optimality, D-optimality etc.. However, unlike our model, the estimation error of $f(x_i)$ for any regression vector x_i in ODE is assumed to come from a fixed distribution, independent of the worker's effort level. Hence, our problem can be viewed as a generalization of ODE. In particular, it is straightforward to show that the C-optimal ODE problem is a special case of ours. Unfortunately, even this special case is NP-hard to solve exactly (46). We propose two different approaches to address this computational intractability.

Robustness to Approximation. The first approach is to use an approximate solution to the minimization problem (2) in our mechanism. It is clear from the proof of Theorem 9 that our mechanism can set the unique dominant strategy equilibrium to be any feasible solution of the minimization problem (2). Moreover, the objective value (2) of our mechanism at this unique dominant strategy equilibrium equals the value of the minimization problem (2) on that particular feasible solution. In short, our mechanism provides *an approximation preserving reduction* from the ESW problem to the minimization problem (2). I.e. whenever an α -factor approximation algorithm exists for (2), our technique translates that (in a black-box manner) to an α -factor approximation algorithm for ESW.

Optimal Assignment Problem. We may also consider a special case of the minimization problem (2), where the set of regression vectors are predetermined but not assigned to the workers. Formally, we restrict (2) to the following minimization problem:

$$\min_{\mathcal{W}', (x_i := s_{\pi(i)}, e_i)_{i \in \mathcal{W}'}} \left(\mathbb{E}_{x^*, \mathbf{y}} \left[\left(\hat{f}_{(x, \mathbf{y})}(x^*) - f(x^*) \right)^2 \right] + \eta \cdot \sum_i e_i \right), \quad (5)$$

where $\mathbf{s} = (s_1, \dots, s_k)$ is a fixed set of regression vectors and π is a bijection from \mathcal{W}' to $[k]$. We show that if the well-behaved estimator \hat{f} is *separable*,⁹ the minimization problem (5) is solvable in polynomial time via min-cost bipartite matching.

Definition 10 A well-behaved estimator \hat{f} for \mathcal{H} , as in Definition 8, is separable iff there exists a function h , such that, for all distributions F over \mathcal{D} , functions $f \in \mathcal{H}$, and vectors $\mathbf{x} \in \mathcal{D}^*$ and $\boldsymbol{\sigma} \in \mathbb{R}_+^*$ (of the same dimension as \mathbf{x}):

$$\mathbb{E}_{\mathbf{y}, x^*} \left[\left(\hat{f}_{(x, \mathbf{y})}(x^*) - f(x^*) \right)^2 \right] = \sum_i h(x_i, X, F) \cdot \sigma_i^2,$$

where X is the set of all x_i 's, and for the purposes of the expectation on the left hand side $x^* \sim F$ and, independently for all i , $y_i = f(x_i) + \epsilon_i$, where each ϵ_i is sampled from an arbitrary distribution of mean 0 and variance σ_i^2 , (and when \mathbf{x} is such that $\hat{f}_{(x, \mathbf{y})}$ is well-defined).

9. Linear regression, polynomial regression and finite-dimensional kernel regression are all separable.

Theorem 11 *Given a set of regression vectors $\mathbf{s} = (s_1, \dots, s_k)$, and a separable estimator \hat{f} for \mathcal{H} , there is a polynomial time algorithm for finding an optimal solution to the minimization problem (5).*

Proof We reduce the minimization problem to the min-cost bipartite matching problem. Let S be the set of regression vectors in \mathbf{s} , (W, T) be a complete bipartite graph, such that W is the set of all workers and T is the set of all regression vectors in S plus $|W| - k$ dummy regression vectors. For any edge (i, j) with $j \in [k]$, the cost $c_{i,j}$ is $\min_{e_i} h(s_j, S, F) \cdot \sigma_i(e_i)^2 + \eta \cdot e_i$. Since $\sigma_i(e_i)$ is a convex function, this cost can be computed in polynomial time. For any edge incident to a dummy regression vector, the cost is 0.

Now consider (5). Since \hat{f} is a separable estimator, the minimization problem can be simplified as

$$\begin{aligned} & \min_{\mathcal{W}', (\pi_i, e_i)_{i \in \mathcal{W}'}} \left(\sum_i h(s_{\pi(i)}, S, F) \cdot \sigma_i(e_i)^2 + \eta \cdot \sum_i e_i \right) \\ &= \min_{\mathcal{W}', \pi} \sum_{i \in \mathcal{W}'} \min_{e_i} h(s_{\pi(i)}, S, F) \cdot \sigma_i(e_i)^2 + \eta \cdot e_i \\ &= \min_{\mathcal{W}', \pi} \sum_{i \in \mathcal{W}'} c_{i, \pi_i}. \end{aligned}$$

The formula above wants to find a subset of workers W' and match them to the regression vectors in S , such that the total cost is minimized. Notice that matching any worker to a dummy regression vector has cost 0, therefore finding an optimal solution to the formula above is equivalent as finding a min-cost bipartite matching in (W, T) . Given any min-cost bipartite matching M , we can construct an optimal solution to (5) by assigning regression vector s_j to j 's partner i in M . Since finding a min-cost bipartite matching takes only polynomial time, (5) can also be solved in polynomial time. \blacksquare

3.2. Discussion and Extensions

One of the conditions in Theorem 9 is that the estimator \hat{f} is well-defined with one example less. Without it, we cannot hope for any interesting solutions to ESW. As a trivial example, suppose that \hat{f} is the unbiased estimator of constant functions from \mathbb{R} to \mathbb{R} , which takes as input one example (x, y) and outputs y . In this case \hat{f} is not well-defined with one example less, and no interesting solutions to ESW exist, since the payments to a worker may only depend on his own report, and hence the worker will put the minimal effort in set \mathcal{E} , regardless of our payment, as long as our payment is at least that minimal effort. (Recall that, since we never learn the unknown function f we cannot use it to penalize the worker.) In this example, our constraint that \hat{f} is well-defined with one example less effectively says that we need to use at least two workers for estimating constant functions from \mathbb{R} to \mathbb{R} .

We point out that this requirement can be removed in settings where the test point x^* and the function value $f(x^*)$ can be observed by the Statistician and the experts. We can then treat $(x^*, f(x^*))$ as an additional example with zero variance, and modify (3) to include that example in \hat{f} .

Our condition that \hat{f} is well-behaved does not contain regularized estimators, such as ridge regression. As ridge regression is biased, its mean square error also includes a bias term that depends

on f . This could create two potential problems for our mechanism. (i) As we do not know f , we can not possibly solve the minimization problem (2) even if we are given infinite computational power. (ii) This term will also appear in every worker's expected payment (if we use payments as in (3)), and, since f is unknown, the workers can't evaluate their expected utilities exactly, and thus can't decide if it's beneficial for them to participate in the mechanism. We point out that, under the assumption that the x_i 's are fixed in advance (but the Statistician is still allowed to optimally assign them to workers) our mechanism can address both problems and accommodate ridge regression with only mild modifications. For problem (i), the mean square error of ridge regression can be separated into two parts – the bias term and the variance term. The bias term depends on f and the x_i 's but is independent of the σ_i 's. This means that no matter how the samples are assigned and how they are estimated, the bias term remains the same. So we only need to consider the variance term. Luckily the variance does not depend on f , thus we can still find the optimal solution of the minimization problem (2). For problem (ii), we can modify our payments. Roughly, we can use any unbiased estimator \tilde{f} and add another term $(\hat{f} - \tilde{f})^2$ to the payment function (both estimators applied to the examples from the other workers). The extra term will cancel the bias in expectation and will allow the workers to reason about their optimal behavior even without knowing f . Note that the new term will introduce some extra terms in the expected utility, but these only depend on known quantities such as the variances, and the worker will be able to reason about his optimal behavior.

Although our objective function is stated as the weighted sum of mean square error and the total payment in ESW, our mechanism can accommodate many variants of that objective function. In general, we can modify the minimization problem (2) to reflect the new objective function. We can then apply our same mechanism to enforce that the unique dominant strategy equilibrium is achieved at the optimal solution of the new minimization problem. Here we list a few variants of our objective function and show how to modify the minimization problem (2) accordingly.

Worker-specific scaling factor: An easy generalization is to replace the single scaling factor for the total payment to worker-specific scaling factors in the objective. To accommodate this objective, we only need to change the scaling factor for the sum of efforts to worker-specific scaling factors in (2).

Budget Constraints: A possible alternative objective is minimizing the mean square error subject to a budget constraint for the total payment. For this objective, we update (2) to minimize the mean square error subject to the constraint that the sum of all workers' efforts is no greater than the budget.

Replacing the total payment with any increasing function of the payments: Another possible generalization of our objective is to replace the total payment with any increasing function of the payments $q(p)$. In this case, we update (2) to minimize the sum of mean square error plus $q(e)$.

Budget on any increasing function of the payments: We can also impose budget constraints on any increasing function of the payments. For this objective, we modify (2) to minimize the mean square error subject to the constraint that the increasing function over the workers' efforts does not exceed the budgets.

Beyond Mean Square Error: Our techniques can be generalized to accommodate general error functions beyond mean-square. If the function combined with the estimator are well-behaved, that is, it is independent of the function f , then our techniques go through. We only need to modify (2) to minimize the weighted sum of the new error function and the total effort.

References

- [1] A. Badanidiyuru, R. Kleinberg, and A. Slivkins “Bandits with Knapsacks: Dynamic procurement for crowdsourcing,” in (31).
- [2] D. F. Bacon, Y. Chen, I. Kash, D. C. Parkes, M. Rao, and M. Sridharan “Predicting your own effort,” *International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, 2012.
- [3] S. Chawla, J. D. Hartline, and B. Sivan “Optimal crowdsourcing contests,” SODA 2012.
- [4] X. Chen, Q. Lin and D. Zhou “Optimistic Knowledge Gradient Policy for Optimal Budget Allocation in Crowdsourcing,” ICML 2013.
- [5] X. Chen, P. N. Bennett, and K. Collins-Thompson “Pairwise ranking aggregation in a crowdsourced setting,” WSDM, 2013.
- [6] E. Clarke “Multipart pricing of public goods,” *Public Choice*, 11(1):1733, 1971.
- [7] A. Dasgupta and A. Ghosh “Crowdsourced judgement elicitation with endogenous proficiency,” *International conference on World Wide Web (WWW)*, 2013.
- [8] O. Dekel, F. Fischer, and A. D. Procaccia “Incentive compatible regression learning,” *Symposium on Discrete algorithms (SODA)*, 2008.
- [9] A. Doan, R. Ramakrishnan, and A. Y. Halevy “Crowdsourcing systems on the world-wide web,” *CACM* 2011.
- [10] F. Fang, M. Stinchcombe, and A. Whinston “Putting Your Money Where Your Mouth Is - A Betting Platform for Better Prediction,” *Review of Network Economics*, 6(2), 2007.
- [11] RPH Fishe, RP McAfee “Nonlinear contracts, zero profits and moral hazard,” *Economica*, 1987.
- [12] G. Goel, A. Nikzad, and A. Singla “Matching Workers Expertise with Tasks: Incentives in Heterogeneous Crowdsourcing Markets,” in (31).
- [13] T. Groves “Incentives in teams,” *Econometrica*, 41(4):617631, 1973.
- [14] C. Ho, A. Slivkins, and J. Wortman-Vaughan “Adaptive Contract Design for Crowdsourcing,” in (31).
- [15] T. Hastie, R. Tibshirani, and J. Friedman “The Elements of Statistical Learning,” Springer-Verlag, New York, 2001
- [16] J. Horton and L. Chilton “The labor economics of paid crowdsourcing,” EC 2010.
- [17] ICML 2013 Worskhop: Machine Learning Meets Crowdsourcing, http://www.ics.uci.edu/~qliu1/MLcrowd_ICML_workshop/
- [18] N. Immorlica, G. Stoddard, and V. Syrgkanis “Social Status and the Design of Optimal Badges,” in (31).

- [19] S. Ioannidis and P. Loiseau “Linear regression as a non-cooperative game,” *Web and Internet Economics* (WINE), 2013.
- [20] J. Kieffer. “Optimal Experiment Designs, (with discussion),” *J. Roy. Statist. Soc. B.* 272-319, 1959.
- [21] J. Kieffer. “Optimum designs in regression problems,” *II. Ann. Math. Statist.* 32 298-325, 1961.
- [22] J.Kieffer. “General equivalence theory for optimum designs (approximate theory),” *Ann. Statist.* 2 849-879, 1974.
- [23] J. Kieffer and J. Wolfowitz. “Optimum designs in regression problems,” *Ann. Math. Statist.* 30 271-294, 1959
- [24] J. Kieffer and J. Wolfowitz. “The equivalence of two extremum problems, ” *Canad.J.Math.* 14 363-366, 1960.
- [25] J. Kieffer and J. Wolfowitz. “On a theorem of Hoel and Levine on extrapolation,” *Ann. Math.Statist*3.61627-1655, 1965.
- [26] A. Mas-Colell, M. D. Whinston, and J. R. Green. *Microeconomic Theory*. Oxford university press, 1995.
- [27] R. Meir, A. D. Procaccia, and J. S. Rosenschein “Strategyproof Classification under Constant Hypotheses: A Tale of Two Functions,” *AAAI*, 2008.
- [28] R. Meir, A. D. Procaccia, and J. S. Rosenschein “Strategyproof Classification with Shared Inputs,” *IJCAI*, 2009.
- [29] N. Miller, P. Resnick, and R. Zeckhauser. ”Eliciting informative feedback: The peer-prediction method.” *Management Science* 51.9 (2005): 1359-1373.
- [30] S. Nath, P. Dayama, D. Garg, and Y. Narahari “Mechanism design for time critical and cost critical task execution via crowdsourcing,” *Internet and Network Economics* (WINE), 2012.
- [31] NIPS 2013 Woskshop on Croudsourcing: Theory, Algorithms, and Applications, http://www.ics.uci.edu/~qliu1/nips13_workshop/.
- [32] N. Nisan, T. Roughgarden, E. Tardos, and V. Vazirani (eds.) “Algorithmic Game Theory,” Cambridge University Press, 2007.
- [33] R. Nix and M. Kantarcioglu “Incentive compatible privacy-preserving distributed classification,” *IEEE Transactions on Dependable and Secure Computing*, 9(4): 451–462, 2012.
- [34] R. Myerson “Optimal Auction Design,” *Mathematics of Operations Research*, 6(1):5873, 1981.
- [35] D. DiPalantino and M. Vojnovic “Crowdsourcing and all-pay auctions,” *EC* 2009.
- [36] G. Patterson, G. Van Horn, S. Belongie, P. Perona, and J. Hays “Bootstrapping Fine-Grained Classifiers: Active Learning with a Crowd in the Loop,” in (31).

- [37] F. Pukelsheim. “Optimal Design of Experiments,” Wiley & Sons, 1993.
- [38] V. Rajan, S. Bhattacharya, L. Celis, D. Chander, K. Dasgupta and S. Karanam “CrowdControl: An online learning approach for optimal task scheduling in a dynamic crowd platform,” in (17)
- [39] P. Ruvolo, J. Whitehill and J. Movellan “Exploiting Commonality and Interaction Effects in Crowdsourcing Tasks Using Latent Factor Models,” in (31).
- [40] S. Sabato and A. Kalai “Feature Multi-Selection among Subjective Features,” in (17).
- [41] Y. Singer and M. Mittal “Pricing mechanisms for crowdsourcing markets,” WWW 2013.
- [42] A. Singla and A. Krause “Truthful incentives in crowdsourcing tasks using regret minimization mechanisms,” WWW 2013.
- [43] A. Singla and A. Krause “Truthful Incentives for Privacy Tradeoff: Mechanisms for Data Gathering in Community Sensing,” in (17)
- [44] T. Jaakkola “Lecture Notes of Machine Learning”
<http://ocw.mit.edu/courses/electrical-engineering-and-computer-science/6-867-machine-learning-fall-2006/lecture-notes/lec5.pdf>
- [45] W. Vickrey “Counterspeculation, auctions, and competitive sealed tenders,” The Journal of Finance, 16(1):837, 1961.
- [46] M. Cerny, and M. Hladik. “Two complexity results on c-optimality in experimental design,” Computational Optimization and Applications 51, 13971408, 2012.

Appendix A. Related Work

For background in statistical analysis and estimation see, e.g., (15), and for mechanism design see, e.g., (32). In the past few years there have been several papers treating crowdsourcing in a framework that is at least superficially similar to ours. (16) gauge through experiments the elasticity of effort under pay in crowdsourcing, while (14) use learning algorithms to find the optimum crowdsourcing contract, and (39) add experts to the crowd so non-experts perform better. Scheduling mechanisms are used to manipulate the time behavior of the crowd (38; 30), whereas (12) use incentives to match workers according to their specialization. Several papers address strategies to optimize performance keeping within budget (40; 41; 4), while in (1) the online task assignment problem is treated as a multi-armed bandit under a budget. The optimal design of non-monetary “prestige” rewards to optimally incentivize the participants is addressed in (18), and in (33; 19; 43) privacy concerns in data gathering are treated through incentives; also in (42) regret minimization is used in a crowdsourcing context.

A little closer to our framework, Mechanism Design has been used in (35; 3) to analyze crowdsourcing *contests*—in which all workers submit their work and one is selected to be paid—as all-pay auctions. These mechanisms also use monetary incentives to create a competition between workers to enhance performance, but their nature and the problems they address are quite different from ours—for example, they are by design not individually rational. In (8; 27; 28), mechanism design is used for regression and classification with strategic data providers. In contrast to our setting, the

agents are not interested in being paid for producing data for the learning task and have no cost for exerting effort. Instead, they control a subset of the data for which they already have the correct labels, and want to bias the outcome of the learning process to perform well on their subset of the data. (2) and follow-up papers design proper scoring rules for principal-agent problems where agents are asked to predict an outcome, but can also influence the outcome by exerting different levels of effort. In contrast to our setting, agents only care about rewards and have no cost for effort. More importantly, the designer only cares about how to incentivize the agents to exert maximum effort, and not how much money is paid to the agents. Moreover, they use the final outcome to decide rewards (this is common in scoring rules), while we never learn $f(x^*)$. (7) design mechanisms for crowdsourced binary labeling, where a collection of agents are asked to provide binary labels to a collection of tasks, and the designer wants to incentivize them via monetary rewards to exert maximum effort for each task. The agents' strategy consists on how much effort to exert for each task they are allocated (where increased effort increases their cost but also the probability of correct labeling), and whether to truthfully report the labels of their allocated tasks. Besides being for a different learning task, the biggest difference to our paper lies in the fact that the designer does not care about how much money is paid to the agents, or the tradeoff between rewards and accuracy. (29) use scoring rules to incentivize agents whose private signals are correlated to reveal their private signals. In comparison to our work, the agents' decision is whether to report their true private signal or some other signal, but not how much effort to exert to get better quality signals. Finally, (10) look at a problem close to ours except they assume that each worker cannot affect their quality (it is sampled from a prior distribution), but only decide whether to participate in the mechanism; they also only consider the simple unbiased estimator for constant functions from \mathbb{R} to \mathbb{R} .

Relation to Optimal Contract Theory (Principal-agent Problem). Our mechanisms achieve in a unique dominant strategy equilibrium objective value that an omnipotent dictator who could dictate how much effort each worker exerts would achieve. This is a surprising result that is morally related to results in optimal contract theory; see, e.g., Chapter 14 of (26). The question, known as the *principal-agent problem*, is how to design a contract with a worker whose effort e affects your revenue π through some conditional density $f(\pi|e)$. The contract is a mapping $w(\cdot)$ from your revenue π and the worker's effort (if it is observable) to the worker's income $w(\pi)$. So your utility is $\pi - w(\pi)$, while the worker's utility is $w(\pi) - g(e)$, for some function $g(\cdot)$. The problem is straightforward to formulate when the worker's effort is observable. The surprising result, which is reminiscent of our result, is that under appropriate conditions one can design a contract $w(\cdot)$ resulting in the same utility when the worker's effort is *not observable* as if it were observable. While the two results are related, the setting and techniques are different. Moreover, there are important qualitative differences in the results. The aforementioned result applies to a single worker, while ours applies to any number. Also, the worker's payment, even when his/her effort is not observable, is still a function of the revenue π , which is exactly observable. Instead, in our setting we do not observe what mean squared error we achieve as a result of the efforts exerted by the workers. The use of convexity in the payments, that we employ, has been employed in the design of non-linear contracts (11). (Many thanks to Preston McAfee for bringing this paper to our attention.)

Comparison to VCG. The guarantees of our mechanism are also reminiscent to those of the celebrated VCG mechanism (45; 6; 13). VCG optimizes social welfare in a dominant strategy equilibrium, achieving as high a welfare as that of an omniscient algorithm (who knows the bidders' valuations and can therefore compute the exactly optimal outcome). While the guarantees of our

mechanism are similar to those of VCG, the technical reasons underlying the two results are different. Importantly, the VCG mechanism may allow multiple equilibria with worse guarantees. In contrast, our mechanism achieves optimality in a unique dominant strategy equilibrium. So, in particular, there are no bad equilibria. Moreover, VCG is very sensitive to the computational complexity of the underlying algorithmic problem. If the algorithmic problem is intractable, then so is running VCG, and the VCG mechanism is known to fail if the algorithmic problem can only be approximated. In contrast, our proposed mechanism is approximation preserving, as noted in Section 3.1.