Truthful Data Quality Elicitation for Quality-Aware Data Crowdsourcing

Xiaowen Gong, *Member, IEEE*, and Ness B. Shroff, *Fellow, IEEE*

***Abstract*—Data crowdsourcing has found a broad range of applications (e.g., environmental monitoring and image classification) by leveraging the "wisdom" of a potentially large crowd of "workers" (e.g., mobile users). A key metric of crowdsourcing is data accuracy, which relies on the *quality* of the participating workers' data (e.g., the probability that the data are equal to the ground truth). However, the data quality of a worker can be its own private information (which the worker learns, e.g., based on its location) that it may have incentive to misreport, which can, in turn, *mislead* the crowdsourcing requester about the accuracy of the data. This issue is further complicated by the fact that the worker can also manipulate its effort made in the crowdsourcing task and the data reported to the requester, which can also mislead the requester. In this paper, we devise truthful crowdsourcing mechanisms for *quality, effort, and data elicitation (QEDE)*, which incentivize strategic workers to truthfully report their private worker quality and data to the requester, and make truthful effort as desired by the requester. The truthful design of the QEDE mechanisms overcomes the lack of ground truth and the coupling in the joint elicitation of the worker quality, effort, and data. Under the QEDE mechanisms, we characterize the socially optimal and the requester's optimal (RO) task assignments, and analyze their performance. We show that the RO assignment is determined by the largest "virtual quality" rather than the highest quality among workers, which depends on the worker's quality and the quality's distribution. We evaluate the QEDE mechanisms using simulations that demonstrate the truthfulness of the mechanisms and the performance of the optimal task assignments.**

***Index Terms*—Crowdsourcing.**

I. Introduction

MOBILE data crowdsourcing (referred to as "crowdsourcing" for brevity) has found a wide range of applications. Typical applications involves physical sensing tasks (also known as "crowdsensing") such as spectrum sensing, traffic monitoring, and environmental monitoring. In principle, crowdsourcing leverages the "wisdom" of a potentially large crowd of workers (i.e., mobile users) for a crowdsourcing task. A key advantage of crowdsourcing lies in that it can exploit the diversity of inherently inaccurate data from many workers by aggregating the data obtained by the crowd, such that the data accuracy (also referred to as "data quality") after aggregation can be substantially enhanced. With enormous opportunities brought by big data, crowdsourcing serves as an important first step for data mining tools to harness the power of big data in many application domains.

To fully exploit the potential of crowdsourcing, it is important to assign crowdsourcing tasks to workers based on their quality. A worker's quality[1] captures the *intrinsic* accuracy of the worker's data relative to the ground truth of the interested variable, and it generally varies for different workers depending on a worker's characteristics (e.g., location and capabilities of sensors). For example, if the task is to detect whether a wireless device is transmitting or not (for dynamic spectrum access), then the quality of a worker's data is the probability of correct detection, which depends on the worker's location with respect to that device. Workers generally have *diverse* quality. A worker can learn its quality based on its characteristics or context, such as its location.[2] However, the quality of a worker can be its private information, which is unknown to and cannot be verified by the requester. As a result, a strategic worker may have incentive to misreport its quality to the requester so as to gain an advantage. For example, a worker of low quality may report high quality in the hope of receiving a high reward for contributing high quality data.

In addition to the worker quality, the data quality of a worker is also affected by its effort exerted in a crowdsourcing task. The data quality of a worker when it makes effort in the task is higher than when it makes no effort. For example, to detect whether a licensed frequency band is idle, a worker should measure the signal in that band to make an estimate, rather than making a guess without any measuring. However, a worker's effort can also be its hidden action that cannot be observed by the requester. Therefore, a strategic worker may make arbitrary

Manuscript received August 30, 2018; revised December 29, 2018; accepted February 17, 2019. Date of publication March 14, 2019; date of current version March 18, 2020. This paper was presented in part at the ACM International Symposium on Mobile Ad Hoc Networking and Computing 2018 [1]. The work of X. Gong was supported by the start-up fund provided by Auburn University. Recommended by Associate Editor Peng Cheng. *(Corresponding author: Xiaowen Gong.)*

X. Gong is with the Department of Electrical and Computer Engineering, Auburn University, Auburn, AL 36830 USA (e-mail: xgong@auburn.edu).

N. B. Shroff is with the Department of Electrical and Computer Engineering and the Department of Computer Science and Engineering, Ohio State University, Columbus, OH 43210 USA (e-mail: shroff.11@osu.edu).

Digital Object Identifier 10.1109/TCNS.2019.2905090

[1] We use "worker quality" and "quality" exchangeably in this paper. "Worker quality" should be distinguished from "data quality."

[2] Alternatively, a worker can report its characteristics (e.g., location) that determines its quality to the requester so that the requester can learn the worker's quality. In this case, reporting the worker's quality is equivalent to reporting its characteristics.

effort. Furthermore, the data itself obtained by a worker from the task could also be its private information that it can manipulate in favor of itself.

In the presence of strategic workers with private worker quality, hidden effort, and private data, our goal is to incentivize workers to truthfully reveal their worker quality and data, and make truthful effort as desired by the crowdsourcing requester. Such a truthful mechanism is desirable as it eliminates the possibility of manipulation, which would encourage workers to participate in crowdsourcing. More importantly, the joint truthful elicitation of quality, effort, and data ensures that the requester can *correctly know the data accuracy of the collected data*, which is a key metric of crowdsourcing. This is in contrast to the situation of crowdsourcing with private participating cost, where manipulating the cost does not mislead the requester about the data accuracy.

The joint elicitation of quality, effort, and data calls for new truthful design that is different from existing mechanisms. First, a worker's payoff as a function of its quality, effort, and data has a different structure from that of its private participating cost. As a result, the existing designs for cost elicitation cannot work for the problem here. Second, due to the statistical dependence of a worker's private data on its private quality and hidden effort, the joint elicitation of quality, effort, and data needs to overcome the coupling therein.

Given a truthful mechanism that can elicit quality, effort, and data from workers, an important question for the requester is to determine which worker(s) the task should be assigned to based on their quality, in order to maximize the social welfare or the requester's payoff. This involves the tradeoff between assigning the task to more workers to improve the data accuracy, and assigning it to fewer workers to reduce the total cost incurred or total reward paid to the workers.

The main contributions of this paper can be summarized as follows.

1) Under a quality-aware crowdsourcing framework, we devise truthful mechanisms for quality, effort, and data elicitation (QEDE) that incentivize strategic workers to truthfully report their private quality and data, and make truthful effort as desired by the requester. The truthful design of the QEDE mechanisms overcomes the lack of ground truth and the coupling in the joint elicitation of worker quality, effort, and data, by exploiting the statistical dependence of a worker's private data on its private worker quality and hidden effort.
2) Under the QEDE mechanisms, we characterize the socially optimal (SO) and the requester's optimal (RO) task assignments, and analyze their performance. We show that the RO assignment is determined by the largest virtual quality rather than the highest quality among workers, which depends on the worker's quality and the quality's distribution. We also show that, as the number of workers becomes large, the gap between the social welfare attained by the RO assignment and the SO assignment decreases and converges to 0.
3) We evaluate the QEDE mechanisms using simulation results, which demonstrate the truthfulness of the mechanisms and the performance of the RO and SO assignments.

The rest of this paper is organized as follows. Section 1II reviews related work. In Section III, we describe the system model of quality-aware crowdsourcing with private data quality and formulate the problem of the truthful mechanism design. In Section IV, we devise truthful mechanisms for QEDE, and explain the ideas of the design and the rationale behind. In Section V, we characterize the optimal assignments under the QEDE mechanisms and analyze their performance. Simulation results are presented in Section VII. Section VII concludes this paper and discusses future work.

II. Related Work

A. Quality-Aware Data Crowdsourcing

Crowdsourcing based on workers' data quality has been studied in a few works [2]–[7]. Some of them [6], [7] has studied truthful mechanisms that elicit workers' private participating costs. Some other works [2], [4], [5] have focused on learning the data quality of workers. Different from these works, this paper focuses on the situation where quality is a worker's private information that is unknown to the requester.

A recent work [8] has proposed a quality-aware crowdsourcing framework and devised truthful mechanisms for quality and effort elicitation. Compared to this paper, a key difference of [8] is that it considers continuous data and the quality is measured by the variance of the error. On the other hand, this paper focuses on *discrete* data and the quality measured by the *correct probability*. Moreover, the truthful mechanisms devised in this paper achieve joint elicitation of quality, effort, and data, which is stronger than in [8], which achieves quality and effort elicitation. As a result of these differences, the truthful mechanisms of in this paper and its analysis are nontrivially different from those in [8].

B. Truthful Mechanisms for Crowdsourcing

There have been a lot of recent research on incentive mechanisms for crowdsourcing [3], [9]–[14]. Most of these mechanisms incentivize workers to truthfully reveal their participating costs. Different from these works, we study the situation where the quality of a worker's data obtained from a crowdsourcing task is the worker's private information that it can manipulate. A worker's payoff as a function of its private quality has a different structure than that of its private cost. As a result, the existing designs for the cost elicitation (such as the classical Vickrey-Clarke-Groves (VCG) auction and the characterization of truthful mechanisms [15, Th. 9.36]) cannot work for quality elicitation, a new design is needed. Furthermore, this paper aims at the joint elicitation of quality, effort, and data. The statistical dependence of a worker's private data on its private quality and hidden effort leads to the coupling in the elicitation of quality, effort, and data, which needs to be addressed.

There have been many studies on the mechanism design for hidden actions in the economics literature [16]. A few recent works have studied this problem in the context of crowdsourcing

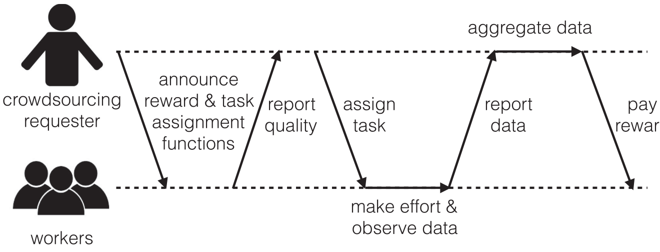

Fig. 1. Structure and procedure of the quality-aware crowdsourcing framework.

[12], [17]–[20]. For example, Cai *et al.* [18] have designed truthful mechanisms to incentivize workers to make effort as desired in statistical estimation; Luo *et al.* [12] have designed mechanisms that not only elicit desired effort from workers but also truthful revelation of their private costs and data. This paper is different from these works as we aim to jointly elicit workers' private quality, private data, and hidden effort, which cannot be achieved by the existing truthful design.

Mechanism design for the truthful elicitation of strategic agents' data (e.g., opinions) has been extensively studied in various applications (e.g., [21]), more recently, for crowdsourcing [12], [17], [19], [20], [22]. Different from the existing works, in this paper, we aim to design truthful mechanisms that jointly elicit workers' private data, private quality, and hidden effort, which calls for new truthful design.

III. Quality-Aware Data Crowdsourcing Framework

We consider a crowdsourcing requester recruiting a set of workers $\mathcal{N} \triangleq \{1, \ldots, N\}$ to work on a task. The structure and procedure of the crowdsourcing system is illustrated in Fig. 1 and described in detail as follows.

A. Data Crowdsourcing With Private Data Quality

1) Data Observation: The crowdsourcing task is to observe and estimate an unknown and random variable of interest X. The interested variable X takes discrete values (e.g., the answer of a multichoice question). For ease of exposition, we assume that $X \in \{0, 1\}$ (which can be generalized to the case of more than two values of X). After performing the task, each worker $i \in \mathcal{N}$ obtains random data D_i. The accuracy of the data D_i is quantified by the *correct probability* p_i, which is the probability that D_i is equal to X, given by

$$p_i \triangleq \Pr(D_i = X) = q_i e_i + \hat{q}(1 - e_i). \quad (1)$$

Here, p_i depends on the *quality* q_i of worker i and the *effort* e_i exerted by worker i in the task, and $\hat{q}$ denote the correct probability when the worker makes no effort $e_i = 0$ (e.g., using the prior distribution of X).

2) Worker Quality: Given that worker i makes effort in the task, the quality $q_i \in [0, 1]$ determines the correct probability p_i, which quantifies how accurate D_i is. The quality q_i is an *intrinsic* coefficient that captures worker i's capability for the task. Note that a larger q_i means higher quality. The quality generally varies for different workers. We assume that each worker $i \in \mathcal{N}$ knows its quality q_i (e.g., by learning the correct probability based on its location). However, the quality of each worker $i \in \mathcal{N}$ is unknown to the requester. For ease of exposition, we assume that each worker's quality q_i is within the range of $[\underline{q}, \bar{q}]$, which is known to the requester.

3) Worker Effort: The effort $e_i \in \{0, 1\}$ represents whether worker i makes effort in the task, where $e_i = 1$ and $e_i = 0$ indicate making and not making effort, respectively. If worker i makes effort, then the correct probability p_i of worker i is equal to the worker quality q_i; otherwise, p_i is equal to $\hat{q}$, which means that worker i simply makes a guess of X randomly according to the prior distribution. To ensure that making effort is meaningful, we assume that $q_i > \hat{q}$. Therefore, given the quality q_i, making effort $e_i = 1$ means a larger correct probability p_i, and thus, higher accuracy of D_i than not making effort. The binary effort model (i.e., either making effort or not) is reasonable (also used in, e.g., [17], [19], and [20]), as workers' behavior tend to be simple in practice. We assume that each worker i can control its effort e_i, but it cannot be observed by the requester. We assume that the requester itself always makes effort in the task (i.e., $e_0 = 1$).

4) Task Assignment: The requester assigns the crowdsourcing task to the workers by assigning effort e'_i to each worker i, which indicates whether it desires worker i to make effort in the task, based on the workers' quality. To this end, each worker i reports its quality q'_i to the requester.[3] Since the true quality q_i is worker i's private information, it may manipulate the reported quality q'_i to its own advantage such that $q'_i \neq q_i$. Based on the quality reported by all workers, the requester determines the effort e'_i assigned to each worker i according to a certain assignment function

$$e'_i(\boldsymbol{q}') \quad (2)$$

and notifies worker i of e'_i. The assignment function $e'_i(\boldsymbol{q}')$ is predefined by the requester and announced to all the workers before they report their quality. A worker's assignment generally varies for different workers due to the diversity of their quality. Intuitively, a worker of high quality is preferred to be assigned to the task. Note that, in general, the assignment e'_i is not only dependent on the quality q'_i reported by worker i but also on the quality $\boldsymbol{q}'_{-i}$ reported by the other workers. After being assigned effort e'_i to, each worker i works on the task by making actual effort e_i. Since e_i is a hidden action of worker i, it may manipulate it against the assignment e'_i to its own advantage such that $e_i \neq e'_i$. After obtaining data d_i from the task (which is a sample realization of the random data D_i), each worker i reports data d'_i to the requester. Since d_i is also private information of worker i, it may manipulate the reported data d'_i against the actual obtained data d_i to its own advantage such that $d_i \neq d'_i$.

5) Data Aggregation: After collecting all the data $\boldsymbol{d}$ reported by the workers, the requester aggregates the data $\boldsymbol{d}$ by making the optimal estimate x_0 of the interested variable X

[3] Workers should report their worker quality $\{q'_i\}$ rather than data quality $\{p'_i\}$, as it allows the requester to assign the task to workers based on their quality $\{q'_i\}$. This is desirable for achieving some particular task assignments, such as the SO assignment.

based on $\boldsymbol{d}$. The optimal estimate x_0 maximizes the posterior probability that x_0 is equal to the ground truth x, i.e.,

$$x_0(\boldsymbol{q}', \boldsymbol{e}', \boldsymbol{d}') \triangleq \arg\max_{d\in\{0,1\}} E_{X|\boldsymbol{d}'(\boldsymbol{q}',\boldsymbol{e}')}[\mathbf{1}_{X=d}]. \tag{3}$$

Note that the distribution of X conditioned on $\boldsymbol{d}'$ depends on workers' reported quality $\boldsymbol{q}'$ and assigned effort $\boldsymbol{e}'$. Then, the utility of crowdsourcing is represented by the correct probability p_c of the optimal estimate x_0, given by

$$p_c(\boldsymbol{q}', \boldsymbol{e}', \boldsymbol{d}') \triangleq E_{X|\boldsymbol{d}(\boldsymbol{q},\boldsymbol{e})}\left[\mathbf{1}_{X=x_0(\boldsymbol{q}',\boldsymbol{e}',\boldsymbol{d}')}\right]. \tag{4}$$

Note that the expectation is over the posteriori distribution $X|\boldsymbol{d}(\boldsymbol{q},\boldsymbol{e})$ conditioned on the true data $\boldsymbol{d}$ depending on the true quality $\boldsymbol{q}$ and actual effort $\boldsymbol{e}$. If the task is not assigned to any worker (i.e., $e_i' = 0\ \forall i$), then the correct probability p_c is defined to be 0.

6) Worker Reward: On the other hand, the requester pays a reward r_i to each worker i for working on the task, according to a certain reward function

$$r_i(\boldsymbol{q}', e_i', d_i', d_j'). \tag{5}$$

Note that the reward r_i depends on the *reference* data d_j obtained by another worker j where $j \neq i$. The reward function is also predefined by the requester, and announced to all the workers before they report their quality (together with the assignment function $e_i'(\boldsymbol{q}')$). Note that the reward function can only depend on the information that the requester knows, i.e., $\boldsymbol{q}'$, $\boldsymbol{e}'$, and $\boldsymbol{d}'$.

7) Worker Payoff: Each worker i's payoff u_i is the reward r_i paid by the requester minus its *cost* in the task, given by

$$u_i(\boldsymbol{q}', e_i, d_i') \triangleq r_i(\boldsymbol{q}', e_i', d_i', d_j') - c_i e_i. \tag{6}$$

Here, the cost c_i represents how much resource is consumed by worker i (e.g., how much time is spent by worker i) if it makes effort $e_i = 1$ in the task. If worker i make no effort $e_i = 0$, it incurs no cost. Note that the relative weight of the cost c_i with respect to the reward r_i in (6) can be captured by c_i. We assume that workers have the same cost[4] c (i.e., $c_i = c\ \forall i$), which is known to the requester. This assumption is reasonable when the cost c is determined by a uniform market price for working on a task.

8) Requester Payoff: The requester's payoff u_0 is the crowdsourcing utility (i.e., the correct probability p_c) minus the total reward paid to the workers, i.e.,

$$u_0(\boldsymbol{q}', \boldsymbol{e}', \boldsymbol{d}') \triangleq p_c(\boldsymbol{q}', \boldsymbol{e}', \boldsymbol{d}') - \sum_{i\in\mathcal{N}} r_i(\boldsymbol{q}', e_i', d_i', d_j'). \tag{7}$$

For the convenience of readers, we summarize the main notation used in this paper in Table I.

B. Mechanism Design Objective

As the workers have private quality and data and make hidden effort, if any worker manipulates its reported quality, reported data, or actual effort, then the estimate x_0 found by the requester would be different from the correct estimator, i.e., $x_0(\boldsymbol{q}', \boldsymbol{e}', \boldsymbol{d}') \neq x_0(\boldsymbol{q}, \boldsymbol{e}, \boldsymbol{d})$. More importantly, the correct probability p_c found by the requester would be different from the correct one, i.e., $p_c(\boldsymbol{q}', \boldsymbol{e}', \boldsymbol{d}') \neq p_c(\boldsymbol{q}, \boldsymbol{e}, \boldsymbol{d})$. This means that the requester has *incorrect information about the correct probability!* This is highly undesirable since the data accuracy is often a key performance metric (e.g., to meet some threshold requirement). Thus motivated, we aim to design a mechanism, which is a pair of an assignment function $e_i'(\boldsymbol{q}')$ and a reward function $r_i(\boldsymbol{q}', e_i', d_i', d_j')$ that can achieve the property of *truthfulness*. In particular, we are interested in a mechanism under which truthful behavior of all workers is a *Nash equilibrium (NE)*, defined as follows.

TABLE I
Summary of Main Notation

Symbol	Meaning
X	variable of interest (ground truth)
d_i	true data of worker i
d_i'	reported data of worker i
q_i	true quality of worker i
q_i'	reported quality of worker i
e_i	true effort of worker i
e_i'	effort assigned to worker i
x_0	optimal estimate of interested variable X
r_i	reward paid to worker i
c_i	cost of worker i
u_i	payoff of worker i
u_0	payoff of the requester

Definition 1: A mechanism achieves truthful strategies of all workers as an NE if, for each worker i, given that all other workers $j \neq i\ \forall j$ truthfully report their quality and data, and make the effort desired by the requester, the optimal strategy of worker i for maximizing its expected payoff is also to the truthful strategy, i.e.,

$$\begin{aligned} &E_{D_j|d_i(q_i,e_i)}\left[u_i(q_i, \boldsymbol{q}_{-i}, e_i, d_i, D_j)\right] \\ &\quad \geq E_{D_j|d_i(q_i,e_i)}\left[u_i(q_i', \boldsymbol{q}_{-i}, e_i', d_i', D_j)\right]\ \forall (q_i', e_i, d_i')\ \forall \boldsymbol{q}_{-i}. \end{aligned}$$

Another natural and desirable property we aim to achieve is that each worker's expected reward should at least compensate its cost (i.e., its expected payoff is nonnegative), since otherwise the worker would not participate in crowdsourcing for a payoff of 0. This property of *individual rationality (IR)* is stated as follows.

Definition 2: A mechanism is IR if for each worker i, given that it truthfully reports its quality and makes the effort desired by the requester, its expected payoff is nonnegative, i.e.,

$$E_{D_j|d_i(q_i,e_i)}\left[u_i(q_i, \boldsymbol{q}_{-i}', e_i', d_i')\right] \geq 0\ \forall \boldsymbol{q}_{-i}'.$$

IV. Truthful Mechanisms for Joint QEDE

In this section, we design truthful crowdsourcing mechanisms that achieve the truthful and IR properties.

We first present the QEDE mechanisms as follows.

Definition 3: A QEDE mechanism consists of any assignment function $e_i'(\boldsymbol{q}')$ that satisfies the condition in (8) and the reward function $r_i(\boldsymbol{q}', e_i', d_i', d_j')$ given by (9) based on that

[4] The truthful mechanisms still hold when workers have diverse costs c_i (i.e., $c_i \neq c_j\ \forall i \neq j$), which are known to the requester.

$e'_i(\boldsymbol{q}')$.

$$e'_i(q'_i, \boldsymbol{q}'_{-i}) \geq e'_i(q''_i, \boldsymbol{q}'_{-i}) \; \forall q'_i \leq q''_i \; \forall \boldsymbol{q}'_{-i} \tag{8}$$

$$r_i(\boldsymbol{q}', e'_i, d'_i, d'_j) = kq'_i e'_i(\boldsymbol{q}') \left[\frac{\mathbf{1}_{d'_j = d'_i} + q'_j - 1}{2q'_j - 1} \right] + ce'_i(\boldsymbol{q}')$$
$$+ \int_{\underline{q}}^{q'_i} kqe'_i(q, \boldsymbol{q}'_{-i})dq - kq'_i e'_i(\boldsymbol{q}') \left[\hat{q} + (q'_i - \hat{q}) e'_i(\boldsymbol{q}') \right] \tag{9}$$

where k is any constant that satisfies the condition

$$k \geq \frac{c}{\underline{q}(\underline{q} - \hat{q})} \tag{10}$$

and $\mathbf{1}_A$ is the indicator function that is equal to 1 if condition A is true and 0 otherwise.

The condition (8) is a general *monotonicity* property for the task assignment functions: given any quality of the other workers, if the task is assigned to a worker, then the task is still assigned to that worker when its quality improves. Intuitively, these assignment functions are natural and desirable for system efficiency. Next, we will explain the main ideas of the design of the QEDE mechanisms (8) and (9). In particular, we will show successively that, for each worker, it is optimal to report its true data (Lemma 1), it is optimal to make actual effort as desired (Lemma 2), and it is optimal to report the true quality (Lemma 3). As a result, the truthful property is achieved (Theorem 1). Then, we will explain the rationale behind the design in Remark 1.

In the following, we show how the QEDE mechanisms achieve the truthful property (with the proofs in Appendix). We first show that any worker's optimal reported data are to report the true data, independent of its reported quality and actual effort. Given the lack of the ground truth x, this is achieved by the peer prediction mechanism (see, e.g., [17], [19], [20], and [22]), which compares the reported data d_i with the reference data d_j from the requester.

Lemma 1: Under the QEDE mechanisms, given that any worker i reports any quality q'_i and makes any effort e'_i, its optimal reported data are its true data $d'_i = d_i$.

Using Lemma 1, given that worker i reports the optimal data $d'_i = d_i$, we can express its expected payoff as

$$E_{D_j | d_i(q_i, e_i)}[u_i(\boldsymbol{q}', e'_i, d_i, D_j)] = kq'_i e'_i(\boldsymbol{q}') \left[\hat{q} + (q_i - \hat{q}) e_i \right]$$
$$+ \int_{\underline{q}}^{q'_i} kqe'_i(q, \boldsymbol{q}'_{-i})dq - kq'_i e'_i(\boldsymbol{q}') \left[\hat{q} + (q'_i - \hat{q}) e'_i(\boldsymbol{q}') \right]$$
$$+ ce'_i(\boldsymbol{q}') - ce_i. \tag{11}$$

Then, we have

$$E_{D_j | D_i(q_i, e_i)}[u_i(\boldsymbol{q}', e'_i, D_i, D_j)]$$
$$= E_{D_j | d_i(q_i, e_i)}[u_i(\boldsymbol{q}', e'_i, d_i, D_j)] \; \forall d'_i$$

since the right-hand side of (11) is independent of d_i. For convenience, we can define

$$\bar{u}_i(\boldsymbol{q}', q_i, e'_i, e_i) \triangleq E_{D_j | D_i(q_i, e_i)}[u_i(\boldsymbol{q}', e'_i, D_i, D_j)] \tag{12}$$

according to (11). Then, we show that, as worker i can only affect its payoff in (12) via its reported quality q'_i and actual effort e_i, its optimal actual effort is the desired effort e'_i, independent of q_i and true quality q_i.

Lemma 2: Under the QEDE mechanisms, given that any worker i reports any quality q'_i and truthfully report its data d_i, its optimal actual effort is the desired effort $e_i = e'_i$.

Using Lemma 2, given that worker i reports the optimal data $d'_i = d_i$ and makes the optimal effort $e_i = e'_i$, we can express its expected payoff using (11) and (12) as

$$\bar{u}_i(\boldsymbol{q}', q_i, e'_i, e_i) = kq'_i e'_i(\boldsymbol{q}')(q_i - q'_i) + \int_{\underline{q}}^{q'_i} kqe'_i(q, \boldsymbol{q}'_{-i})dq. \tag{13}$$

For convenience, we can define

$$\hat{u}_i(\boldsymbol{q}', q_i, e'_i) \triangleq \bar{u}_i(\boldsymbol{q}', q_i, e'_i, e_i) \tag{14}$$

according to (13).

Next, we show that, as worker i can only affect its payoff in (14) via its reported quality q'_i, its optimal reported quality is its true quality q_i, under the general condition (8) on the assignment function $e'_i(\boldsymbol{q}')$.

Lemma 3: Under the QEDE mechanisms, given that any worker i truthfully reports its data d_i and makes the desired effort e'_i, its optimal reported quality is its true quality $q'_i = q_i$.

Using Lemmas 1–3, we can show that the truthful property is achieved as in the next theorem. Using (13) and (14), given that worker i reports the optimal data $d'_i = d_i$, makes the optimal effort $e_i = e'_i$, and reports the optimal quality $q'_i = q_i$, its payoff is given by

$$\hat{u}_i(q_i, \boldsymbol{q}'_{-i}, q_i, e'_i) = k \int_{\underline{q}}^{q_i} qe'_i(q, \boldsymbol{q}'_{-i})dq. \tag{15}$$

It follows that the IR property is also achieved since $\hat{u}_i(q_i, \boldsymbol{q}'_{-i}, q_i, e'_i) \geq 0$ due to that $e'_i(\boldsymbol{q}') \geq 0 \; \forall \boldsymbol{q}'$.

Theorem 1: The QEDE mechanisms are truthful in the NE and are IR.

Remark 1: We explain the design rationale of the QEDE mechanisms as follows. We first observe that the optimal reported data d'_i that maximizes $E_{D_j | d_i(q_i, e_i)}[\mathbf{1}_{D_j = d'_i}]$, which is the probability that d'_i is equal to d_j is always the true data d_i (as in Lemma 1). Then, we can design the expected reward as a function of $\hat{q} + (q_i - \hat{q})e_i$, which is the probability that d'_i is equal to the ground truth x, such that worker i's expected payoff depends on the true quality q_i and the actual effort e_i, and is independent of the true data d_i [as in (11)]. Now the expected payoff only depends on q'_i, e'_i, q_i, and e_i [as in (12)]. Then, we can design the reward function such that the optimal actual effort e_i that maximizes the payoff is always the desired effort e'_i and independent of q'_i and q_i (as in Lemma 2). As a result, worker i's payoff now only depends on q'_i, e'_i, and q_i [as in (13) and (14)]. Next, we further design the reward function such that, under the monotonicity condition (8) on e'_i, the optimal reported quality q'_i is always the true quality q_i and independent of e'_i (as in Lemma 3).

It follows from (15) that when all workers behave truthfully (i.e., $q_i' = q_i$, $e_i = e_i'$, and $d_i' = d_i$), the total expected reward paid by the requester is

$$\sum_{i \in \mathcal{N}} \left(k \int_{\underline{q}}^{q_i} q e_i'(q, \boldsymbol{q}_{-i}') dq + c e_i'(\boldsymbol{q}') \right). \tag{16}$$

It can be seen from (16) that, to minimize the requester's payoff, k should be minimized such that the condition (10) is satisfied with equality. We assume that this equality holds in the rest of this paper.

Remark 2: We can see from (16) that the requester's payment for each worker consists of two parts: while the second part $c e_i'(\boldsymbol{q}')$ is to compensate the worker's cost, the first part (i.e., the integral multiplied by k) is to elicit the worker's truthful behavior. This shows that the requester pays more than needed to cover the cost by the truth-eliciting payment (also known as "information rent" [23]), which is due to the requester's uncertainty of workers' quality. We can also observe from (16) that, as the multiplier k [determined by (10)] and the integral are both decreasing in the lower bound $\underline{q}$ of workers' quality, the truth-eliciting payment is also decreasing in $\underline{q}$ (as illustrated by Fig. 5 in Section VII). Intuitively, this is because the requester knows more information (i.e., less uncertainty) of workers' quality with a larger $\underline{q}$.

V. Optimal Task Assignment Under Truthful Mechanisms

In Section IV, we have shown that the truthful and IR properties can be achieved by all the QEDE mechanisms, which have general assignment functions that satisfy condition (8). In this section, we will find the optimal assignment under the QEDE mechanisms that maximizes the social welfare and the requester's payoff, respectively. Because of the truthful property, in this section, we assume that $\boldsymbol{q}' = \boldsymbol{q}$, $\boldsymbol{e} = \boldsymbol{e}'$, and $\boldsymbol{d}' = \boldsymbol{d}$. Therefore, for brevity, we use $\boldsymbol{q}$, $\boldsymbol{e}$, and $\boldsymbol{d}$ instead of $\boldsymbol{q}'$, $\boldsymbol{e}'$, and $\boldsymbol{d}'$, respectively.

A. SO Assignment

An important metric for the assignment $\boldsymbol{e}(\boldsymbol{q})$ is system efficiency, which is measured by the social welfare (the requester may be also interested in this objective). The social welfare v is the crowdsourcing utility (i.e., the correct probability p_c) minus the total cost of all workers, i.e.,

$$v(\boldsymbol{q}, \boldsymbol{e}(\boldsymbol{q})) \triangleq E_{\boldsymbol{D}(\boldsymbol{q},\boldsymbol{e})}[p_c(\boldsymbol{q}, \boldsymbol{e}, \boldsymbol{D})] - \sum_{i \in \mathcal{N}} c e_i. \tag{17}$$

Definition 4: The *SO assignment* $\boldsymbol{e}^{\text{so}}(\boldsymbol{q})$ for the QEDE mechanisms is the assignment function $\boldsymbol{e}(\boldsymbol{q})$ satisfying condition (8) that maximizes the social welfare, i.e.,

$$\{\boldsymbol{e}^{so}(\boldsymbol{q}) \, \forall \boldsymbol{q}\} \triangleq \arg \max_{\{\boldsymbol{e}(\boldsymbol{q}) \, \forall \boldsymbol{q}\} \text{ s.t. } (8)} E_{\boldsymbol{Q}}[v(\boldsymbol{Q}, \boldsymbol{e}(\boldsymbol{q}))]. \tag{18}$$

We first consider a single-worker assignment, which consists of the assignment functions that assign the task to at most one worker. The advantage of the single-worker assignment is that it simplifies the implementation of crowdsourcing: the requester needs to collect data from only one worker rather than potentially many workers. We should note that single-worker assignment still *exploits the diversity of potentially many available workers* in crowdsourcing, as the worker is selected based on the quality of all the workers. Under the single-worker assignment, the RO estimate x_0 of the interested variable X is just equal to the data d_i reported by the worker i who works on the task, and the correct probability p_c is equal to the quality q_i of that worker i.

We can find the SO assignment for the single-worker assignment as follows.

Proposition 1: For the single-worker assignment, the SO assignment is given by

$$e_i^{so}(\boldsymbol{q}) = \begin{cases} 1, & i = \arg\max_j q_j \text{ and } q_i \geq c \\ 0, & \text{otherwise.} \end{cases} \tag{19}$$

Proposition 1 shows that the task is assigned to the "best" worker i that has the highest quality q_i if and only if the cost c is less than the quality q_i. This is clearly because the best worker maximizes the correct probability p_c, and thus, the social welfare v. It is also clear that the SO assignment (19) satisfies the monotonicity condition (8) of the QEDE mechanisms (thus, the proof of Proposition 1 follows and is omitted). We should note that although the single-worker assignment involves only one worker to work on the task, it still exploits the diversity of multiple available workers, as it selects the worker of the highest quality.

Next, we consider general assignment functions that can assign the task to multiple workers. It can been shown (e.g., see [5]) that the optimal estimate given in (3) is equivalent to that $x_0 = 1$ if and only if

$$\prod_{i:e_i=1,d_i=1} \log \frac{q_i}{1-q_i} \geq \prod_{j:e_j=1,d_j=1} \log \frac{q_j}{1-q_j}$$

and $x_0 = 0$ otherwise. It has been shown in [5] that, without imposing the condition (8) of the QEDE mechanisms, the optimal assignment that maximizes the social welfare satisfies an intuitive property: there exists some k such that the task is assigned to only the top k workers that have the highest quality. As a result, it can be found by an efficient exhaustive search algorithm with linear complexity as described in Algorithm 1. In the following, we show that the solution found by Algorithm 1 is also the SO assignment for the QEDE mechanisms.

Proposition 2: For the multiworker assignment, the SO assignment is found by Algorithm 1.

The main idea of the proof of Proposition 2 is to show that the output of Algorithm 1 satisfies the monotonicity condition (8) of the QEDE mechanisms.

B. RO Assignment

A desirable objective for the requester is to find the optimal assignment that maximizes its expected payoff.

Definition 5: The *crowdsourcing RO assignment* $\mathbf{e}^*(\boldsymbol{q})$ for the QEDE mechanism is the assignment function $\boldsymbol{e}(\boldsymbol{q})$ satisfying condition (8) that maximizes the requester's expected payoff (7),

Algorithm 1: Find the SO assignment for multiworker assignment.

```
Index workers in the descending order of their quality, i.e., $q_1 \geq q_2 \geq \cdots q_N$;
$e_j \leftarrow 0 \, \forall j, t \leftarrow v(\boldsymbol{q}, \boldsymbol{e}), i = 1$;
while $i \leq N$;
do
    $e_i \leftarrow 1$;
    if $v(\boldsymbol{q}, \boldsymbol{e}) > t$ then
        $\boldsymbol{e}^* \leftarrow \boldsymbol{e}$;
    end
    $i \leftarrow i + 1$;
end
return $\boldsymbol{e}^{so}$;
```

i.e.,

$$\{\boldsymbol{e}^*(\boldsymbol{q}), \forall \boldsymbol{q}\} \triangleq \arg \max_{\{\boldsymbol{e}(\boldsymbol{q}), \forall \boldsymbol{q}\} \text{ s.t. } (8)} E_{\boldsymbol{D}(\boldsymbol{Q}, \boldsymbol{e})}[u_0(\boldsymbol{Q}, \boldsymbol{e}, \boldsymbol{D})]. \quad (20)$$

For ease of analysis, in the rest of this subsection, we focus on the single-worker assignment. One reason is that the characterization of the RO assignment for the single-worker assignment and the corresponding performance analysis provide useful insights. We further assume that each worker's quality follows an independent and identical distribution over an interval $[\underline{q}, \bar{q}]$, which is known to the requester.

Proposition 3: For the single-worker assignment, when

$$\alpha(q) \triangleq q + kq\frac{F(q) - 1}{f(q)}$$

is an increasing function of q, the RO assignment is given by

$$e_i^*(\boldsymbol{q}) = \begin{cases} 1, & i = \arg\max_j \alpha(q_j) \text{ and } \alpha(q_i) \geq c \\ 0, & \text{otherwise} \end{cases} \quad (21)$$

where $f(q)$ and $F(q)$ denote the probability density function (PDF) and cumulative density function (CDF) of each worker's quality, respectively.

We should note that the property that $\alpha(q)$ is an increasing function of q holds under mild conditions, e.g., when $F(q)$ and $f(q)$ follow a uniform distribution. We assume that this property holds in the rest of this section.

Remark 3: Proposition 3 shows that the task is assigned to the "best"[5] worker i that has the largest "virtual quality" $\alpha(q_i)$ (if the cost c is less than $\alpha(q_i)$). Note that each worker i's virtual quality depends on its quality q_i and also the quality's distribution $F(q_i)$ and $f(q_i)$. This implies that the range of a worker's possible quality, represented by $\Delta q \triangleq \bar{q} - \underline{q}$, affects the task assignment. For ease of analysis, suppose $F(q)$ and $f(q)$ follow a uniform distribution such that $(F(q) - 1)/f(q) = q - \bar{q}$. Given workers' quality, when the upper bound $\bar{q}$ of workers' quality decreases so that Δq decreases, each worker's virtual quality $\alpha(q_i)$ increases, and thus, the condition $\alpha(q_i) \geq c$ for assigning the task to the best worker i is more likely to hold. Intuitively, this is because a smaller quality range incurs a lower truth-eliciting payment in (16) by the requester in order to achieve truthful elicitation, which increases the requester's payoff. In the special case of $\Delta q = 0$, a worker's virtual quality is equal to its quality. The concept of virtual valuation was introduced by Myerson [15] and is in the same spirit as the result here.

Remark 4: Comparing (19) and (21), we can see that the SO assignment is similar to the RO assignment in that the task can be assigned only to the best worker i that has the highest quality q_i. This is because the virtual quality $\alpha(q)$ is increasing in q, and thus, $\arg\max_j q_j = \arg\max_j \alpha(q_j)$. The difference is that the RO assignment assigns the task to the best worker i based on the condition $\alpha(q_i) \geq c$ rather than the condition $q_i \geq c$ for the SO assignment. Since it can be easily seen that $\alpha(q_i) \leq q_i$ always holds, there exist some values of q_i such that $\alpha(q_i) < c$, while $q_i \geq c$. In this case, the RO assignment $e_i^*(\boldsymbol{q})$ is different from the SO assignment $e_i^{\text{so}}(\boldsymbol{q})$ and attains lower social welfare than $e_i^{\text{so}}(\boldsymbol{q})$. Intuitively, this is because, although assigning the task increases the crowdsourcing utility and also the social welfare, it incurs a too high truth-eliciting payment. As a result, the RO assignment is not SO, and the gap is essentially due to the asymmetry of workers' quality information between the workers and the requester.

C. Performance Analysis

Next, we analyze the impact of system parameters on the performance of the SO and RO assignments.

Proposition 4: The expected RO payoff $E_{\boldsymbol{Q}}[u_0(\boldsymbol{e}^*(\boldsymbol{Q}))]$ attained by the RO assignment, the expected SO social welfare $E_{\boldsymbol{Q}}[v(e_1^{\text{so}}(\boldsymbol{Q}))]$, and the expected social welfare $E_{\boldsymbol{Q}}[v(e_1^*(\boldsymbol{Q}))]$ attained by the RO assignment, all increase as the number of workers N increases, or the cost c decreases.

Remark 5: Proposition 4 shows that the RO payoff and social welfare benefit from a greater diversity in workers' quality. This is because when there are more workers, the quality of the best worker is likely to be higher, which improves the crowdsourcing utility. On the other hand, a larger c increases the cost incurred to workers as well as the truth-eliciting payment [i.e., the first term in (16)], and thus, reduces the RO payoff and social welfare.

Proposition 5: The gap between the expected social welfare of the SO assignment and the RO assignment $E_{\boldsymbol{Q}}[v(e_1^{\text{so}}(\boldsymbol{Q}))] - E_{\boldsymbol{Q}}[v(e_1^*(\boldsymbol{Q}))]$ converges to 0 as the number of workers N goes to infinity.

Remark 6: Proposition 5 shows that the performance gap between the RO assignment and the SO assignment decreases to 0 asymptotically as the number of workers increases. This is because when there are more workers, the quality of the best worker improves so that the gap between the RO and SO assignments decreases to 0 (i.e., they are more often the same), and thus, the gap between their social welfare also decreases to 0.

VI. Discussions on Truthful Reference Data

In the previous sections, we have assumed that each worker i's reward r_i depends on the reference data d_j' from another worker j. In some situations, the requester itself (or a trustworthy

[5] If there are multiple "best" workers, only one of them is selected by breaking the tie randomly.

worker, such as a social friend of the requester) can work on the task and obtains data d_0 with quality q_0 and effort $e_0 = 1$, which are known by the requester. The truthful reference data d_0 and its quality q_0 can be used in the QEDE mechanism to achieve a stronger property of truthfulness. In particular, we modify the reward function of the QEDE mechanism given in (9) by replacing d'_j with the truthful reference data d_0 and replacing q'_j with the quality q_0. The conditions (8) and (10) of the QEDE mechanism remain the same. We can show that the modified QEDE mechanism can achieve the property of dominant truthfulness as stated below, which is a truthful property that is stronger than Definition 1. In addition, they also satisfy the IR property.

Definition 6: A mechanism is *dominant incentive compatible* if, given any quality reported by the other workers, the optimal strategy of each worker i for maximizing its expected payoff is to truthfully report its quality and data, and make the effort desired by the requester, i.e.,

$$E_{D_j|d_i(q_i,e_i)}\left[u_i(q_i,\boldsymbol{q}'_{-i},e_i,d_i)\right] \geq E_{D_j|d_i(q_i,e_i)}\left[u_i(q'_i,\boldsymbol{q}'_{-i},e'_i,d'_i)\right]\ \forall(q'_i,e_i,d'_i)\ \forall\boldsymbol{q}'_{-i}.$$

The proofs of the dominant truthful and IR properties follow from the same argument as that of Theorem 1.

To guarantee that each worker i working on the task (i.e., $e_i = 1$) has a reference worker $j \neq i$ also working on the task (i.e., $e_j = 1$), we need to restrict the assignment function $\boldsymbol{e}'$ such that there are either at least two workers or no worker working on the task, i.e.,

$$\sum_{i\in\boldsymbol{N}} e'_i(\boldsymbol{q}') \neq 1\ \forall\boldsymbol{q}'. \qquad (22)$$

VII. Simulation Results

In this section, we evaluate the properties of the QEDE mechanisms and its performance with the RO assignment using simulations.

A. Worker's Payoff

To illustrate the truthfulness of the QEDE mechanisms, we compare a worker's expected payoff when it truthfully reports its quality and data and makes its effort with when it untruthfully reports its quality and/or data and/or makes its effort. We use the SO assignment $e_i^{\text{so}}(\boldsymbol{q})$ in (19) for the QEDE mechanisms. We set the default parameters as follows:[6] $n = 2$, $c = 0.3$, $\mu_q \triangleq (\bar{q} + \underline{q})/2 = 0.75$, $\Delta q = 0.4$, $q_1 = 0.7$, and $q_2 = 0.6$.

Figs. 2 and 3 illustrate worker 1's expected payoff as it reports varying quality q'_1 (see Fig. 2) or varying true quality q_1 (see Fig. 3) while making desired effort $e'_1 = e^*_1(q'_1, q_2)$ or undesired effort $e'_1 \neq e^*_1(q'_1, q_2)$, and reporting true data $d'_1 = d_1$ or untrue data $d'_1 \neq d_1$, compared to when it truthfully reports its quality and data and makes its effort. We can see that the worker's payoff when its behavior is untruthful is always less than when truthful.

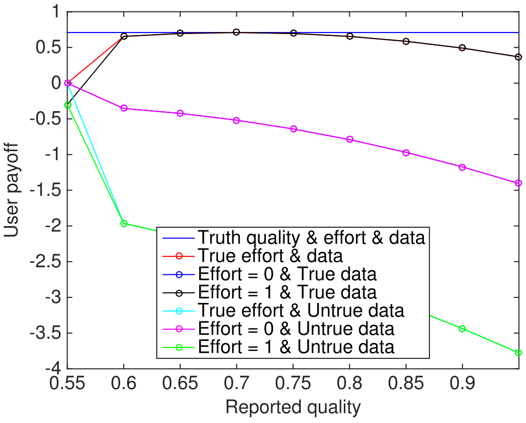

Fig. 2. Impact of the reported quality q'_1.

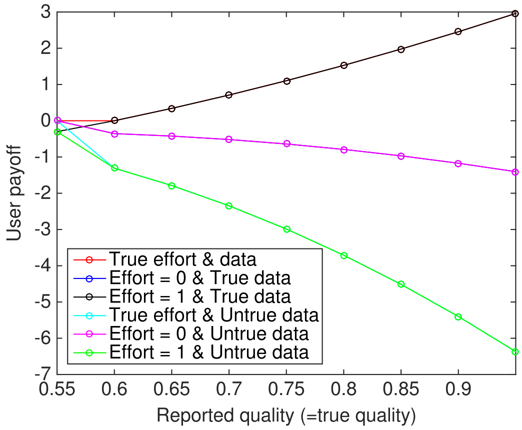

Fig. 3. Impact of the reported quality q'_1 when it is truthful.

B. Requester's Payoff

To illustrate the system efficiency of the RO assignment, we compare the expected requester's payoff (RP), workers' total payoff (UP), and social welfare (SW) attained by the RO assignment (RP-RO, UP-RO, SW-RO) with the expected social welfare (SW) attained by the SO assignment (SW-SO). Note that by definition, UP-RO is always equal to SW-RO minus RP-RO. We set the default parameters as follows: $N = 5$, $c = 0.04$, $\mu_q \triangleq (\bar{q} + \underline{q})/2 = 0.75$, and $\Delta q = 0.4$. We assume that each worker's quality follows an independent identically distributed uniform distribution over $[\underline{q}, \bar{q}]$.

Fig. 4 illustrates the impact of the number of workers N on the performance. We observe that all the curves are increasing in N, which is because they benefit from a greater diversity in workers' quality when there are more workers. We also observe that the gap between SW-RO and SW-SO converges to 0 as N increases.

Fig. 5 illustrates the impact of the quality range Δq on the performance. We observe that SW-SO is increasing in Δq. This is because the social welfare benefits from a greater diversity of workers' quality. We also observe that RP-RO is decreasing in Δq. Intuitively, this is due to that a larger quality range requires a higher truth-eliciting payment in (16).

[6] It suffices to consider two workers only as the RO assignment only depends on the best worker's quality.

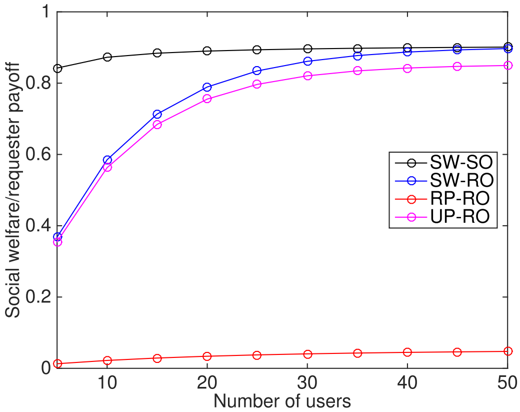

Fig. 4. Impact of the number of workers n.

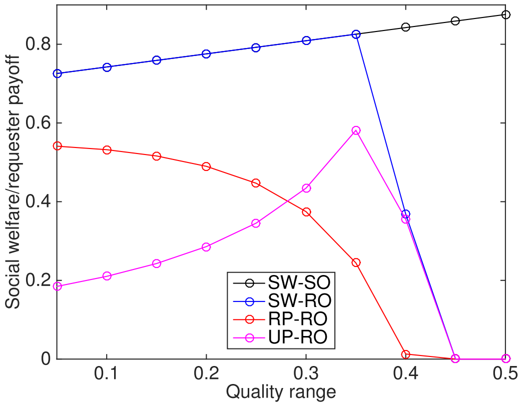

Fig. 5. Impact of quality range Δq.

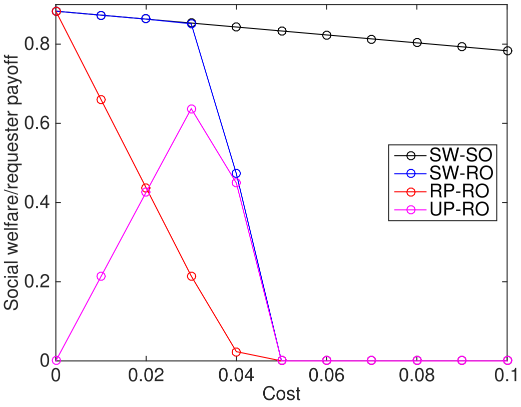

Fig. 6. Impact of cost c.

Fig. 6 illustrates the impact of the cost c on the performance. We observe that all the curves except for UP-RO are decreasing in c, which is because a higher cost results in lower social welfare or the requester's payoff. We also observe that RP-RO decreases faster than SW-RO as c increases. This is because a larger c not only results in a higher payment for compensating workers' cost, but also a higher truth-eliciting payment in (16) (as discussed in Remark 2).

VIII. Conclusion and Future Work

In this paper, we have devised the QEDE mechanisms for quality-aware crowdsourcing, to incentivize workers to truthfully report their private quality and data, and make truthful effort as desired by the requester. The QEDE mechanisms have achieved the truthful design by exploiting the statistical dependency of a worker's private data on its private worker quality and hidden effort, while overcoming the coupling in the joint elicitation of quality, effort, and data. Under the QEQE mechanisms, we have analyzed the SO and RO assignments, which provides useful insights.

For future work, one interesting direction is to consider workers with no knowledge of their quality. In this case, we need to learn the quality of workers from their data. We will also investigate the truthful mechanism design when workers' costs (besides quality, effort, and data) are also their private information, which is still an open problem.

Appendix

A. Proof of Lemma 1

Let $\bar{d}_i$ be the complementary value of d_i, i.e., $\bar{d}_i \neq d_i$. For convenience, let $P_{X|d_i(q_i,e_i)}(d)$ denote the probability that X is equal to d conditioned on data d_i given quality q_i and effort e_i. We observe that when $e_i = 1$, we have

$$P_{X|d_i(q_i,1)}(d_i) = q_i \geq 1 - q_i = P_{X|d_i(q_i,1)}(\bar{d}_i)$$

and when $e_i = 0$, we have

$$P_{X|d_i(q_i,0)}(d_i) = \hat{q} \geq 1 - \hat{q} = P_{X|d_i(q_i,0)}(\bar{d}_i).$$

Since

$$\begin{aligned}
E_{D_j|d_i(q_i,e_i)}\left[\mathbf{1}_{D_j=d_i}\right] &= P_{D_j|d_i(q_i,e_i)}(d_i)\\
&= q_j P_{X|d_i(q_i,e_i)}(d_i) + (1-q_j)(1-P_{X|d_i(q_i,e_i)}(d_i))\\
&= (2q_j - 1)P_{X|d_i(q_i,e_i)}(d_i) + 1 - q_j
\end{aligned}$$

we have

$$E_{D_j|d_i(q_i,e_i)}\left[\frac{\mathbf{1}_{D_j=d_i} + q_j - 1}{2q_j - 1}\right] = P_{X|d_i(q_i,e_i)}(d_i).$$

Similarly, we can show that

$$E_{D_j|d_i(q_i,e_i)}\left[\frac{\mathbf{1}_{D_j=\bar{d}_i} + q_j - 1}{2q_j - 1}\right] = P_{X|d_i(q_i,e_i)}(\bar{d}_i).$$

Then, it follows from (9) that, for any reported quality q_i' and any actual effort e_i, the optimal reported data are given by

$$\begin{aligned}
d_i' &= \arg\max_{d\in\{0,1\}} E_{D_j|d_i(q_i,e_i)}\left[\frac{\mathbf{1}_{D_j=d} + q_j - 1}{2q_j - 1}\right]\\
&= \arg\max_{d\in\{0,1\}} P_{X|d_i(q_i,e_i)}(d)\\
&= \arg\max_{d\in\{0,1\}} \left[P_{X|d_i(q_i,0)}(d) + (P_{X|d_i(q_i,1)}(d)\right.\\
&\qquad \left. - P_{X|d_i(q_i,0)}(d))e_i\right]\\
&= d_i.
\end{aligned}$$

B. Proof of Lemma 2

Using (12), when $e'_i = 1$, we have

$$\bar{u}_i(q'_i, q_i, 1, 1) - \bar{u}_i(q'_i, q_i, 1, 0) = kq'_i(q_i - \hat{q}) - c \geq 0$$

where the inequality follows from (10), and when $e'_i = 0$, we have

$$\bar{u}_i(q'_i, q_i, 0, 0) - \bar{u}_i(q'_i, q_i, 0, 1) = c \geq 0.$$

Hence, the optimal effort to make is $e_i = e'_i$.

C. Proof of Lemma 3

For convenience, we write $\hat{u}_i(\boldsymbol{q}', q_i, e'_i)$ as $\hat{u}_i(q'_i, \boldsymbol{q}'_{-i}, q_i, e'_i)$. It suffices to show that $\hat{u}_i(q_i, \boldsymbol{q}'_{-i}, q_i, e'_i) \geq \hat{u}_i(q, \boldsymbol{q}'_{-i}, q_i, e'_i)$ $\forall q \neq q_i$. Let $q'_i > q_i$. Using (14), we have

$$\begin{aligned}
&\hat{u}_i(q_i, \boldsymbol{q}'_{-i}, q_i, e'_i) - \hat{u}_i(q'_i, \boldsymbol{q}'_{-i}, q_i, e'_i)\\
&\quad = kq_i e'_i(q_i, \boldsymbol{q}'_{-i})(q_i - q_i) + \int_{\underline{q}}^{q_i} kqe'_i(q, \boldsymbol{q}'_{-i})dq\\
&\qquad - \left(kq'_i e'_i(q_i, \boldsymbol{q}'_{-i})(q_i - q'_i) + \int_{\underline{q}}^{q'_i} kqe'_i(q, \boldsymbol{q}'_{-i})dq\right)\\
&\quad = kq'_i e'_i(q_i, \boldsymbol{q}'_{-i})(q'_i - q_i) - \int_{q_i}^{q'_i} kqe'_i(q, \boldsymbol{q}'_{-i})dq\\
&\quad \geq kq'_i e'_i(q_i, \boldsymbol{q}'_{-i})(q'_i - q_i) - kq'_i e'_i(q_i, \boldsymbol{q}'_{-i})(q'_i - q_i) = 0
\end{aligned}$$

where the inequality follows from (8). Now let $q'_i < q_i$. Using (14), we have

$$\begin{aligned}
&\hat{u}_i(q_i, \boldsymbol{q}'_{-i}, q_i, e'_i) - \hat{u}_i(a, \boldsymbol{q}'_{-i}, q_i, e'_i)\\
&\quad = kq_i e'_i(q_i, \boldsymbol{q}'_{-i})(q_i - q_i) + \int_{\underline{q}}^{q_i} kqe'_i(q, \boldsymbol{q}'_{-i})dq\\
&\qquad - \left(kq'_i e'_i(q_i, \boldsymbol{q}'_{-i})(q_i - q'_i) + \int_{\underline{q}}^{q'_i} kqe'_i(q, \boldsymbol{q}'_{-i})dq\right)\\
&\quad = kq'_i e'_i(q_i, \boldsymbol{q}'_{-i})(q'_i - q_i) + \int_{q'_i}^{q_i} kqe'_i(q, \boldsymbol{q}'_{-i})dq\\
&\quad \geq kq'_i e'_i(q_i, \boldsymbol{q}'_{-i})(q'_i - q_i) + kq'_i e'_i(q_i, \boldsymbol{q}'_{-i})(q_i - q'_i) = 0
\end{aligned}$$

where the inequality follows from (8).

D. Proof of Theorem 1

As the IR property has been proved using (15), we only show that the truthful property is achieved. Choose and fix any (q'_i, e_i, d'_i). It follows from Lemma 1 that

$$E_{D_j}\left[u_i(q'_i, \boldsymbol{q}'_{-i}, e_i, d_i, D_j)\right] \geq E_{D_j}\left[u_i(q'_i, \boldsymbol{q}'_{-i}, e_i, d'_i, D_j)\right].$$

Using (11) and (12), it follows from Lemma 2 that

$$\bar{u}_i(\boldsymbol{q}', q'_i, e'_i, e'_i) \geq \bar{u}_i(\boldsymbol{q}', q_i, e'_i, e_i).$$

Using (13) and (14), it follows from Lemma 3 that

$$\hat{u}_i(\boldsymbol{q}', q_i, e'_i) \geq \hat{u}_i(\boldsymbol{q}', q'_i, e'_i).$$

Therefore, we have

$$E_{D_j}\left[u_i(q_i, \boldsymbol{q}'_{-i}, e'_i, d_i, D_j)\right] \geq E_{D_j}\left[u_i(q'_i, \boldsymbol{q}'_{-i}, e_i, d'_i, D_j)\right].$$

E. Proof of Proposition 2

It suffices to show that the output of Algorithm 1 satisfies the monotonicity condition (8), i.e., $e_i^{\mathrm{so}}(q_i, \boldsymbol{q}_{-i}) \geq e_i^{\mathrm{so}}(q'_i, \boldsymbol{q}_{-i})$ $\forall q_i \geq q'_i$ $\forall \boldsymbol{q}_{-i}$. This is equivalent to show that $e_i^{\mathrm{so}}(q_i, \boldsymbol{q}_{-i}) = 1$ if $e_i^{\mathrm{so}}(q'_i, \boldsymbol{q}_{-i}) = 1$. Consider two cases as follows.

1) Suppose $\sum_{i\in\mathcal{N}} e_i^{\mathrm{so}}(q_i, \boldsymbol{q}_{-i}) \geq \sum_{i\in\mathcal{N}} e_i^{\mathrm{so}}(q'_i, \boldsymbol{q}_{-i})$. Since $\{i \in \mathcal{N} | e_i^{\mathrm{so}}(q_i, \boldsymbol{q}_{-i}) = 1\}$ and $\{i \in \mathcal{N} | e_i^{\mathrm{so}}(q'_i, \boldsymbol{q}_{-i}) = 1\}$ consist of top workers that have the highest quality in $\mathcal{N}$, we must have $e_i^{\mathrm{so}}(q_i, \boldsymbol{q}_{-i}) = 1$ if $e_i^{\mathrm{so}}(q'_i, \boldsymbol{q}_{-i}) = 1$.
2) Suppose $\sum_{i\in\mathcal{N}} e_i^{\mathrm{so}}(q_i, \boldsymbol{q}_{-i}) < \sum_{i\in\mathcal{N}} e_i^{\mathrm{so}}(q'_i, \boldsymbol{q}_{-i})$. Also suppose $e_i^{\mathrm{so}}(q'_i, \boldsymbol{q}_{-i}) = 1$ but $e_i^{\mathrm{so}}(q_i, \boldsymbol{q}_{-i}) = 0$. By the definition of $\boldsymbol{e}^{\mathrm{so}}$, we have $v(q_i, \boldsymbol{q}_{-i}, \boldsymbol{e}^{\mathrm{so}}(q_i, \boldsymbol{q}_{-i})) \geq v(q_i, \boldsymbol{q}_{-i}, \boldsymbol{e}^{\mathrm{so}}(q'_i, \boldsymbol{q}_{-i}))$. Assume without loss of generality that $v(q_i, \boldsymbol{q}_{-i}, \boldsymbol{e}^*(q_i, \boldsymbol{q}_{-i})) > v(q_i, \boldsymbol{q}_{-i}, \boldsymbol{e}^*(q'_i, \boldsymbol{q}_{-i}))$. Then, we have

$$\begin{aligned}
v(q_i, \boldsymbol{q}_{-i}, \boldsymbol{e}^{\mathrm{so}}(q_i, \boldsymbol{q}_{-i})) &> v(q_i, \boldsymbol{q}_{-i}, \boldsymbol{e}^{\mathrm{so}}(q'_i, \boldsymbol{q}_{-i}))\\
&\geq v(q'_i, \boldsymbol{q}_{-i}, \boldsymbol{e}^{\mathrm{so}}(q'_i, \boldsymbol{q}_{-i}))\\
&\geq v(q'_i, \boldsymbol{q}_{-i}, \boldsymbol{e}^{\mathrm{so}}(q_i, \boldsymbol{q}_{-i}))\\
&= v(q_i, \boldsymbol{q}_{-i}, \boldsymbol{e}^{\mathrm{so}}(q_i, \boldsymbol{q}_{-i}))
\end{aligned}$$

where the second inequality follows from the following lemma, the third inequality follows from the definition of $\boldsymbol{e}^{\mathrm{so}}$, and the last inequality follows from the fact that $e_i^{\mathrm{so}}(q_i, \boldsymbol{q}_{-i}) = 0$. Thus, we have a contradiction. Therefore, we must have $e_i^{\mathrm{so}}(q_i, \boldsymbol{q}_{-i}) = 1$.

Lemma 4: The expected correct probability $E_D[p_c(\boldsymbol{q}, \boldsymbol{e}, \boldsymbol{D})]$ is an increasing function of q_i $\forall i$.

Proof: Choose and fix any $i \in \mathcal{N}$. For brevity, let

$$h(x, \boldsymbol{d}) \triangleq P(X = x) \prod_{j:d_j = x} q_j \prod_{j:d_j \neq x} (1 - q_j). \tag{23}$$

We can see that

$$E_{X|\boldsymbol{d}(\boldsymbol{q})}\left[\mathbf{1}_{X = x_0(\boldsymbol{q}, \boldsymbol{e}, \boldsymbol{d})}\right] = \frac{\max\left(h(1, \boldsymbol{d}), h(0, \boldsymbol{d})\right)}{P(\boldsymbol{D} = \boldsymbol{d})}.$$

Then, we have

$$\frac{\partial E_D[p_c(\boldsymbol{q}, \boldsymbol{e}, \boldsymbol{D})]}{\partial q_i} = \sum_{\boldsymbol{d}:h(1,\boldsymbol{d}) \geq h(0,\boldsymbol{d})} h'(1, \boldsymbol{d}) + \sum_{\boldsymbol{d}:h(1,\boldsymbol{d}) < h(0,\boldsymbol{d})} h'(0, \boldsymbol{d}) \tag{24}$$

where $h'(x, \boldsymbol{d}) \triangleq \frac{\partial h(x, \boldsymbol{d})}{\partial q_i}$. Choose and fix any x and $\boldsymbol{d}$ such that $h(x, \boldsymbol{d}) > h(\bar{x}, \boldsymbol{d})$ and $x \neq d_i$, and thus, $h'(x, \boldsymbol{d}) < 0$ and $h'(x, \bar{\boldsymbol{d}}) > 0$. Then, we have

$$\frac{h'(x, \bar{\boldsymbol{d}})}{|h'(x, \boldsymbol{d})|} < \frac{1 - q_i}{q_i}. \tag{25}$$

We observe that

$$\frac{h'(\bar{x}, d_i, \bar{\boldsymbol{d}}_{-i})}{|h'(x, d_i, \bar{\boldsymbol{d}}_{-i})|} = \frac{|h'(x, d_i, \boldsymbol{d}_{-i})|}{h'(\bar{x}, d_i, \boldsymbol{d}_{-i})} > \frac{q_i}{1 - q_i} > \frac{1 - q_i}{q_i} \tag{26}$$

where the equality follows from (23) and the fact that $P(X = 0) = P(X = 1)$, the first inequality follows from (25), and the last inequality follows from the fact that $q_i > \hat{q}$. Then, it follows from (26) that for $\bar{x}$ and $(d_i, \bar{\boldsymbol{d}}_{-i})$, we have $h(\bar{x}, d_i, \bar{\boldsymbol{d}}_{-i}) > h(x, d_i, \bar{\boldsymbol{d}}_{-i})$ where $h'(\bar{x}, d_i, \bar{\boldsymbol{d}}_{-i}) > 0$ and $h'(x, d_i, \bar{\boldsymbol{d}}_{-i}) < 0$. Since $h'(\bar{x}, d_i, \bar{\boldsymbol{d}}_{-i}) = |h'(x, d_i, \boldsymbol{d}_{-i})|$, we have $h'(\bar{x}, d_i, \bar{\boldsymbol{d}}_{-i}) + h'(x, d_i, \boldsymbol{d}_{-i}) = 0$. Then, we can see that we must have (24) ≥ 0.

F. Proof of Proposition 3

The expected requester's payoff defined in (7) is given by

$$E_{X|\boldsymbol{D}(\boldsymbol{q},\boldsymbol{e})}[u_0(X, \boldsymbol{D}, \boldsymbol{q}, \boldsymbol{e}(\boldsymbol{q}))] = \sum_{i \in \mathcal{N}} q_i e_i(\boldsymbol{q}) - \sum_{i \in \mathcal{N}} \left(k \int_{\underline{q}}^{q_i} q e_i(q, \boldsymbol{q}_{-i}) dq + c e_i(\boldsymbol{q}) \right). \tag{27}$$

For brevity, define

$$\bar{u}_0(\boldsymbol{e}(\boldsymbol{q})) \triangleq E_{X|\boldsymbol{D}(\boldsymbol{q},\boldsymbol{e})}[u_0(X, \boldsymbol{D}, \boldsymbol{q}, \boldsymbol{e}(\boldsymbol{q}))].$$

Hence, the optimal assignment defined in (20) is given by

$$\{\boldsymbol{e}^*(\boldsymbol{q})\, \forall \boldsymbol{q}\} = \arg \max_{\{\boldsymbol{e}(\boldsymbol{q})\, \forall \boldsymbol{q}\}} E_{\boldsymbol{Q}}[\bar{u}_0(\boldsymbol{e}(\boldsymbol{Q}))]. \tag{28}$$

Since we observe that

$$\begin{aligned}
E_{Q_i}\left[k \int_{\underline{q}}^{Q_i} q e_i(q, \boldsymbol{q}_{-i}) dq\right] &= \int_{\underline{q}}^{\bar{q}} f(q') k \int_{\underline{q}}^{q'} q e_i(q, \boldsymbol{q}_{-i}) dq dq' \\
&= \left[F(q') k \int_{\underline{q}}^{q'} q e_i(q, \boldsymbol{q}_{-i}) dq\right]_{q'=\underline{q}}^{q'=\bar{q}} - \int_{\underline{q}}^{\bar{q}} F(q') k q' e_i(q', \boldsymbol{q}_{-i}) dq' \\
&= k \int_{\underline{q}}^{\bar{q}} q e_i(q, \boldsymbol{q}_{-i}) dq - k \int_{\underline{q}}^{\bar{q}} F(q) q e_i(q, \boldsymbol{q}_{-i}) dq \\
&= k \int_{\underline{q}}^{\bar{q}} f(q) \frac{1 - F(q)}{f(q)} q e_i(q, \boldsymbol{q}_{-i}) dq \\
&= E_{Q_i}\left[k \frac{1 - F(Q_i)}{f(Q_i)} Q_i e_i(Q_i, \boldsymbol{q}_{-i})\right]
\end{aligned} \tag{29}$$

where the second equality follows from integration by parts, then using (27), we have

$$\begin{aligned}
E_{\boldsymbol{Q}}[\bar{u}_0(\boldsymbol{e}(\boldsymbol{Q}))] &= E_{\boldsymbol{Q}}\left[\sum_{i \in \mathcal{N}} Q_i e_i(\boldsymbol{Q})\right] \\
&\quad - \sum_{i \in \mathcal{N}} E_{\boldsymbol{Q}}\left[k \int_{\underline{q}}^{Q_i} q e_i(q, \boldsymbol{Q}_{-i}) dq + c e_i(\boldsymbol{Q})\right] \\
&= E_{\boldsymbol{Q}}\left[\sum_{i \in \mathcal{N}} \left(Q_i e_i(\boldsymbol{Q}) + k \frac{F(Q_i) - 1}{f(Q_i)} Q_i e_i(\boldsymbol{Q}) - c e_i(\boldsymbol{Q})\right)\right]
\end{aligned} \tag{30}$$

where the second equality follows from (29). Hence, finding $\{\boldsymbol{e}^*(\boldsymbol{q})\, \forall \boldsymbol{q}\}$ in (28) is equivalent to solving the following problem for each $\boldsymbol{q}$ independently:

$$\max_{\boldsymbol{e}(\boldsymbol{q})} \sum_{i \in \mathcal{N}} \left(q_i e_i(\boldsymbol{q}) + k \frac{F(q_i) - 1}{f(q_i)} q_i e_i(\boldsymbol{q}) - c e_i(\boldsymbol{q}) \right). \tag{31}$$

Then, we can see that the optimal solution of (31) must be (21).

G. Proof of Proposition 4

Let q_1^* be the best worker's quality and $e_1^*(\boldsymbol{q})$ be the task assignment for the best worker. It can be shown that the best quality $Q_1^*(N)$ for N workers stochastically dominates the best quality $Q_1^*(N')$ for N' workers for any $N' > N$, i.e.,

$$Q_1^*(N) \geq_{st} Q_1^*(N') \quad \forall N' > N. \tag{32}$$

Substituting (21) into (30), we have

$$E_{\boldsymbol{Q}}[u_0(\boldsymbol{e}^*(\boldsymbol{Q}))] = E_{Q_1^*}\left[\max(\alpha(Q_1^*) - c, 0)\right].$$

Since $\alpha(q_1^*)$ is increasing in q_1^*, it follows from (32) that $E_{\boldsymbol{Q}}[u_0(e^*(\boldsymbol{Q}))]$ is increasing in N.

Substituting (19) into (17), we have

$$E_{\boldsymbol{Q}}[v(\boldsymbol{e}^{\text{so}}(\boldsymbol{Q}))] = E_{Q_1^*}\left[\max(Q_1^* - c, 0)\right].$$

Since $q_1^* - c$ is increasing in q_1^*, it follows from (32) that $E_{\boldsymbol{Q}}[v(\boldsymbol{e}^{\text{so}}(\boldsymbol{Q}))]$ is increasing in N.

Substituting (21) into (17), we have

$$E_{\boldsymbol{Q}}[v(\boldsymbol{e}^*(\boldsymbol{Q}))] = E_{Q_1^*}\left[(Q_1^* - c)\mathbf{1}_{\alpha(Q_1^*) \geq c}\right].$$

Since $q_1^* - c$ and $\alpha(q_1^*)$ are increasing in q_1^*, it follows from (32) that $E_{\boldsymbol{Q}}[v(\boldsymbol{e}^*(\boldsymbol{Q}))]$ is increasing in N.

H. Proof of Proposition 5

We observe that

$$\begin{aligned}
&\lim_{N \to \infty} E_{\boldsymbol{Q}}[v(\boldsymbol{e}^{\text{so}}(\boldsymbol{Q}))] - E_{\boldsymbol{Q}}[v(\boldsymbol{e}^*(\boldsymbol{Q}))] \\
&= \lim_{N \to \infty} E_{Q_1^*}\left[\max(Q_1^* - c, 0) - (Q_1^* - c)\mathbf{1}_{\alpha(Q_1^*) \geq c}\right] \\
&= \lim_{q_1^* \to \bar{q}} \left(\max(q_1^* - c, 0) - (q_1^* - c)\mathbf{1}_{\alpha(q_1^*) \geq c}\right) \\
&= \max(\bar{q} - c, 0) - (\bar{q} - c)\mathbf{1}_{\alpha(\bar{q}) \geq c} = 0
\end{aligned}$$

where the second equality follows from that

$$\lim_{N \to \infty} f_{Q_1^*(N)}(\underline{q}) = \infty$$

and

$$\lim_{N \to \infty} f_{Q_1^*(N)}(q) = 0\ \forall q \neq \underline{q}$$

and the last equality follows from the fact that $\bar{q} = \alpha(\bar{q})$.

References

[1] X. Gong and N. Shroff, "Incentivizing truthful data quality for quality-aware mobile data crowdsourcing," in *Proc. ACM Int. Symp. Mobile Ad Hoc Netw. Comput.*, 2018.

[2] D. R. Karger, S. Oh, and D. Shah, "Budget-optimal crowdsourcing using low-rank matrix approximations," in *Proc. IEEE Annu. Allerton Conf. Commun., Control, Comput.*, 2011.

[3] I. Koutsopoulos, "Optimal incentive-driven design of participatory sensing systems," in *Proc. IEEE Int. Conf. Comput. Commun.*, 2013.

[4] D. Lee, J. Kim, H. Lee, and K. Jung, "Reliable multiple-choice iterative algorithm for crowdsourcing systems," in *Proc. ACM Int. Conf. Meas. Model. Comput. Syst.*, 2015.

[5] Y. Liu and M. Liu, "An online learning approach to improving the quality of crowd-sourcing," in *Proc. ACM Int. Conf. Meas. Model. Comput. Syst.*, 2015.

[6] H. Jin, L. Su, D. Chen, K. Nahrstedt, and J. Xu, "Quality of information aware incentive mechanisms for mobile crowd sensing systems," in *Proc. ACM Int. Symp. Mobile Ad Hoc Netw. Comput.*, 2015.

[7] H. Jin, L. Su, and K. Nahrstedt, "CENTURION: Incentivizing multi-requester mobile crowd sensing," in *Proc. IEEE Int. Conf. Comput. Commun.*, 2017.

[8] X. Gong and N. Shroff, "Truthful mobile crowdsensing for strategic users with private qualities," in *Proc. Int. Symp. Model. Optim. Mobile, Ad Hoc, Wireless Netw.*, 2017.

[9] L. Duan, T. Kubo, K. Sugiyama, J. Huang, T. Hasegawa, and J. Walrand, "Incentive mechanisms for smartphone collaboration in data acquisition and distributed computing," in *Proc. IEEE Int. Conf. Comput. Commun.*, 2012.

[10] D. Yang, G. Xue, X. Fang, and J. Tang, "Crowdsourcing to smartphones: Incentive mechanism design for mobile phone sensing," in *Proc. ACM Annu. Int. Conf. Mobile Comput. Netw.*, 2012.

[11] A. Tarable, A. Nordio, E. Leonardi, and M. A. Marsan, "The importance of being earnest in crowdsourcing systems," in *Proc. IEEE Int. Conf. Comput. Commun.*, 2015.

[12] Y. Luo, N. B. Shah, J. Huang, and J. Walrand, "Parametric prediction from parametric agents," in *Proc. 10th Workshop Econ. Netw., Syst. Comput.*, 2015.

[13] W. Wang, L. Ying, and J. Zhang, "The value of privacy: Strategic data subjects, incentive mechanisms and fundamental limits," in *Proc. ACM Int. Conf. Meas. Model. Comput. Syst.*, 2016.

[14] H. Zhang, B. Liu, H. Susanto, G. Xue, and T. Sun, "Incentive mechanism for proximity-based mobile crowd service systems," in *Proc. IEEE Int. Conf. Comput. Commun.*, 2016.

[15] N. Nisan, T. Roughgarden, E. Tardos, and V. V. Vazirani, *Algorithmic Game Theory*, vol. 1. New York, NY, USA: Cambridge Univ. Press, 2007.

[16] P. Bolton and M. Dewatripont, *Contract Theory*. Cambridge, MA, USA: MIT Press, 2005.

[17] A. Dasgupta and A. Ghosh, "Crowdsourced judgement elicitation with endogenous proficiency," in *Proc. Int. World Wide Web Conf.*, 2013.

[18] Y. Cai, C. Daskalakis, and C. H. Papadimitriou, "Optimum statistical estimation with strategic data sources," in *Proc. Conf. Learn. Theory*, 2015.

[19] Y. Liu and Y. Chen, "Learning to incentivize: Eliciting effort via output agreement," in *Proc. Int. Joint Conf. Artif. Intell.*, 2016.

[20] Y. Liu and Y. Chen, "Sequential peer prediction: Learning to elicit effort using posted prices," in *Proc. AAAI Conf. Artif. Intell.*, 2017.

[21] D. Prelec, "A Bayesian truth serum for subjective data," *Science*, vol. 306, no. 5695, pp. 462–466, 2004.

[22] H. Jin, L. Su, and K. Nahrstedt, "Theseus: Incentivizing truth discovery in mobile crowd sensing systems," in *Proc. ACM Int. Symp. Mobile Ad Hoc Netw. Comput.*, 2017.

[23] V. Krishna, *Auction Theory*. San Francisco, CA, USA: Academic, 2009.

Xiaowen Gong (S'13–M'17) received the B.Eng. degree in electronics and information engineering from the Huazhong University of Science and Technology, Wuhan, China, in 2008, the M.Sc. degree in communications from the University of Alberta, Edmonton, AB, Canada, in 2010, and the Ph.D. degree in electrical engineering from the Arizona State University, Tempe, AZ, USA, in 2015.

From 2015 to 2016, he was a Postdoctoral Researcher with the Department of Electrical and Computer Engineering, the Ohio State University, Columbus, OH, USA. He is currently an Assistant Professor with the Department of Electrical and Computer Engineering, Auburn University, Auburn, AL, USA. His current research interests include data crowdsourcing, edge computing, and mobile privacy.

Dr. Gong was the recipient of the Runner-up Best Paper Award at the IEEE International Conference on Computer Communications 2014.

Ness B. Shroff (S'91–M'93–SM'01–F'07) received the Ph.D. degree in electrical engineering from Columbia University, New York, NY, USA, in 1994.

He joined Purdue University, West Lafayette, IN, USA, as an Assistant Professor with the School of Electrical and Computer Engineering, and became a Full Professor in electrical and computer engineering (ECE) and the Director of a university-wide center on wireless systems and applications in 2004. In 2007, he joined the Ohio State University, Columbus, OH, USA, where he holds the Ohio Eminent Scholar Endowed Chair in networking and communications, with the Departments of ECE and Computer Science and Engineering. He holds or has held Visiting (chaired) Professor positions with Tsinghua University, Beijing, China; Shanghai Jiaotong University, Shanghai, China; and Indian Institute of Technology Bombay, Mumbai, India.

Dr. Shroff was the recipient of numerous best paper awards for his research and is listed in Thomson Reuters' on The World's Most Influential Scientific Minds, and is noted as a Highly Cited Researcher by Thomson Reuters. He was also the recipient of the IEEE International Conference on Computer Communications Achievement Award for seminal contributions to scheduling and resource allocation in wireless networks. He currently serves as the Steering Committee Chair for the ACM International Symposium on Mobile Ad Hoc Networking and Computing and the Editor-at-Large for the IEEE/ACM Transactions on Networking.