Online Crowd Learning with Heterogeneous Workers via Majority Voting

Chao Huang, Haoran Yu, Jianwei Huang, and Randall A Berry

***Abstract*—Many platforms recruit workers through crowdsourcing to finish online tasks involving a huge amount of effort (e.g., image labeling and content moderation). These platforms aim to incentivize heterogeneous workers to exert effort finishing the tasks and truthfully report their solutions. When the verification for the workers' solutions is absent, the crowdsourcing problem is challenging and is known as *information elicitation without verification* (IEWV). Majority voting is a common approach to solve an IEWV problem, where a worker is rewarded based on whether his solution is consistent with the majority. However, most prior related work relies on a strong assumption that workers' solution accuracy levels are public knowledge. We relax such an assumption and propose an online learning mechanism based on majority voting, which allows the platform to learn the distribution of the workers' solution accuracy levels. In the mechanism, workers will be asked to report their private accuracy levels (which do not need to be the true values), in addition to deciding their effort levels and solution reporting strategies. The mechanism computes the workers' rewards based on their reported accuracy levels, and the workers obtain rewards if their reported solutions match the majority. We show that our mechanism induces workers to truthfully report their solution accuracy levels in the long run, in which the platform asymptotically achieves zero regret. Moreover, we show that our online mechanism converges faster when the workers are more capable of solving the tasks.**

I. Introduction

A. Motivations

The emerging applications of crowdsourcing have successfully harnessed the intelligence of an unprecedentedly wide-ranging population of workers for solving various tasks [1], [2]. For example, in Waze, one of the most successful crowd-powered start-ups, users report fairly accurate traffic jam information and are provided with automatically generated optimal route suggestions [3]. In OpenReview, an online paper review platform, world-wide researchers anonymously conduct academic reviews [4].

To encourage high-quality solutions from the crowdsourced workers, a platform needs to carefully design the reward mechanism. This is challenging, especially when the platform cannot verify the correctness or quality of workers' solutions [5], [6]. This may be because it is costly and time-consuming to obtain the ground truth. For example, in Waze, traffic jam information reported by the users may be difficult to verify in real-time, and the platform can only reward users based on their reported data. In OpenReview, it is also hard to verify the quality of reviews due to the intrinsic complexity of academic review and reviewer anonymity. When there is a lack of verification for the workers' solutions, the crowdsourcing problem is known as *information elicitation without verification* (IEWV) [7].

There is a large body of research on IEWV during the last decade. Much existing literature (e.g., [7], [8]) considers homogeneous workers, i.e., the capabilities of generating the correct (or high-quality) solutions are the same among workers. This often results in some symmetric equilibria (where all the workers adopt the same strategy), and it provides insights into applications where workers have similar capabilities (e.g., image labeling [9]). However, in many other practical scenarios, workers have quite different capabilities. For example, in mobile crowdsensing, accuracy of sensors from different mobile phones may vary tremendously [10]. In OpenReview, senior researchers are more likely to generate more reliable reviews than junior ones can. In this paper, we consider workers with heterogeneous solution accuracy levels. It is challenging to design a reward mechanism considering such heterogeneity, as accounting for the worker diversity makes the elicitation of effort and truthfulness more difficult.

There is another key difference between our work and the prior literature on IEWV. Most prior work relies on the strong assumption that workers' solution accuracy levels are public information, i.e., such information is known by both the platform and the workers [8], [11], [12]. This assumption may not hold in many practical scenarios. It is becoming more difficult for a platform to access and make use of personal data due to an increasing tendency of privacy protection [13]. Without the personal data, the platform can hardly estimate the workers' solution accuracy levels. For example, in peer grading, without the information of a student's education background, a platform cannot efficiently estimate the student's capability of homework evaluation. Several recent work (e.g., [6], [14]) considered heterogeneity of workers under this strong assumption that the workers' accuracy distribution is public information. This paper takes the first step to study the IEWV problem without requiring the platform to know even

Chao Huang is with the Department of Information Engineering, the Chinese University of Hong Kong. Email: hc017@ie.cuhk.edu.hk. Haoran Yu is with the School of Computer Science, Beijing Institute of Technology. Email: yhrhawk@gmail.com. Jianwei Huang is with the School of Science and Engineering at the Chinese University of Hong Kong, Shenzhen, the Shenzhen Institute of Artificial Intelligence and Robotics for Society, and the Department of Information Engineering at the Chinese University of Hong Kong. Email: jianweihuang@gmail.com. Randall A Berry is with Department of Electrical and Computer Engineering, Northwestern University. Email: rberry@ece.northwestern.edu. This work is supported by the Shenzhen Institute of Artificial Intelligence and Robotics for Society, the General Research Fund CUHK 14219016 from Hong Kong UGC, and the Presidential Fund from the Chinese University of Hong Kong, Shenzhen.

ISBN 978-3-903176-29-4 © 2020 IFIP

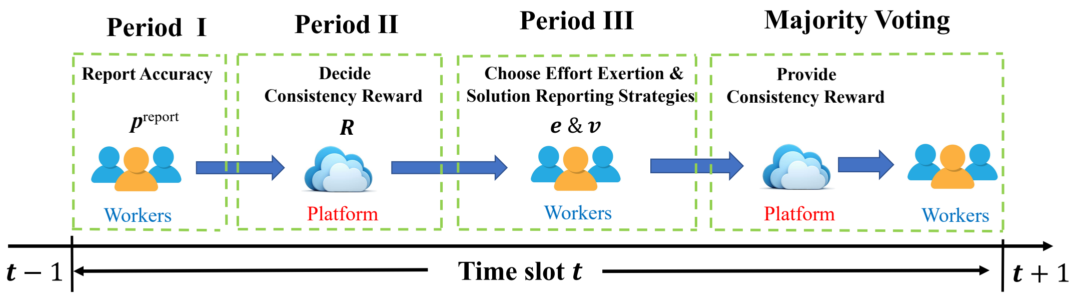


Fig. 1: Interactions between platform and workers in each time slot t.

the workers' accuracy distribution. Such information will be effectively learned and utilized by the platform through our proposed online learning mechanism.

When the workers' accuracy distribution is public information, the majority voting mechanism is a popular practical method for solving the IEWV problem [1], [2], [6]. In the majority voting mechanism, a worker obtains a reward if his solution matches the majority solution from the other workers. In this paper, we build our mechanism based on majority voting. As discussed, prior work lacks an analytical study of the mechanism design when the platform does not know workers' accuracy distribution. We address such a challenge by proposing an online learning framework, where workers' accuracy distribution is estimated based on their accuracy reports as tasks are assigned and finished over time. Hence, the online mechanism gradually learns to use proper rewards to incentivize workers to truthfully report their accuracy. Due to its online nature, our mechanism can be implemented in real time. This is a desirable property as the ability to finish rapidly incoming tasks on the fly can help reduce the dalay and save cost for the platform.

To design an online learning mechanism based on majority voting, we need to answer several questions:

- First, how should the platform design the reward mechanism without knowing the workers' accuracy distribution?
- Second, how will the heterogeneous workers behave in our proposed learning mechanism?

In this paper, we focus on addressing the above key questions. To answer the first question, we propose a majority-voting-based online learning mechanism, which encourages the workers to report their accuracy levels (in addition to deciding their effort levels and solution reporting strategies). To answer the second question, we analyze workers' equilibrium strategies on their accuracy reporting (as well as effort exertion and solution reporting). The platform will implement the learning algorithms over multiple time slots. In each time slot, the mechanism will compute rewards to workers based on their accuracy reports, and the workers will obtain rewards if their reported solutions match the majority. The interactions between the platform and the workers in each time slot are as follows (illustrated in Fig. 1):

1) *Period I*: The workers report their solution accuracy levels to the platform.
2) *Period II*: The platform decides the amount of *consistency rewards* based on the workers' reported accuracy levels.
3) *Period III*: The workers choose effort exertion and solution reporting strategies.

After finishing the task in Period III, the workers report their solutions, and *Majority Voting* is implemented by the platform. More specifically, the platform aggregates the workers' reported solutions and provides *consistency rewards* to the workers whose solutions are consistent with the majority. Note that the results of the majority voting phase are completely dependent on the previous three periods, hence we do not treat it as a separate period.

The online learning mechanism design is very challenging in our context. This is because there is no accessible ground truth to verify either the workers' task solutions or their accuracy reports. Moreover, without prior information of the workers' accuracy levels, the majority voting mechanism provides a natural incentive for the workers to under-report their accuracy so that the platform will set a larger reward (see Section III for more detailed discussions on this point). To address these challenges, we propose a randomized reward mechanism that incentivizes the workers to truthfully report their accuracy levels. Moreover, by carefully designing the reward, the platform can also anticipate the workers' decisions on effort exertion and solution reporting, based on their accuracy reports.

B. Key Contributions

The main contributions of this paper are as follows.

- *Mechanism design without knowing worker accuracy distribution for the IEWV problem*: To the best of our knowledge, this is the first attempt to study the mechanism design for the IEWV problem when the platform does not know the workers' accuracy distribution. The lack of such information makes the mechanism design very challenging, as neither the workers' reported accuracy nor their reported task solutions can be verified.
- *Proposing online crowd learning mechanism*: We design an online mechanism to learn the distribution of workers' solution accuracy levels without ground-truth verification. Our mechanism enables the platform to anticipate the workers' effort exertion and solution reporting decisions based on their past accuracy reports.

- *Characterizing online reward design in closed-form*: We compute in closed-form the online reward design that incentivizes workers to truthfully report their accuracy levels. The reward level also enables the platform to accurately learn the accuracy distribution in the long run.
- *Performance Evaluation*: We evaluate our online mechanism via numerical experiments. Interestingly, we show that workers achieve the highest payoffs by truthfully reporting their accuracy levels, when comparing with several benchmark reporting strategies. We also show that our mechanism converges faster when the workers are more capable of solving the tasks.

The rest of this paper is organized as follows. In Section II, we introduce the model. In Section III, we present the mechanism. In Sections IV and V, we analyze the workers' and the platform's optimal strategies, respectively. We show numerical results in Section VI and conclude in Section VII.

II. Model

In Section II-A, we introduce the workers' decisions and payoffs. In Section II-B, we introduce the interations between the workers and the platform, with an emphasis on the workers' accuracy reporting.

A. Workers' Decisions and Payoffs

In this subsection, we first define each worker's strategy, and then define each worker's payoff function.

1) Workers and Tasks: We consider a finite discrete time horizon $\mathcal{T} = \{1, 2, \cdots, T\}$ and a set $\mathcal{N} = \{1, 2, \cdots, N\}$ of workers. In each time slot t, all the workers are assigned an identical, binary-solution task and required to report their solutions after solving the task. The task can be, for example, judging whether the quality of an online article is *Good* or *Bad*. We use $\mathcal{X}_t = \{1, -1\}$ to denote the solution space of the task assigned in time slot t, where 1 means *Good* and -1 means *Bad*.[1] We use $x_t \in \mathcal{X}_t$ to denote the true solution of the task in time slot t. Each worker i can have a different estimated solution after solving the task, denoted by $x_{i,t}^{\text{estimate}} \in \mathcal{X}_t$, and he can choose to report a value $x_{i,t}^{\text{report}} \in \mathcal{X}_t$ to the platform that may or may not be the same as $x_{i,t}^{\text{estimate}}$. Although the task assigned to all workers in a particular slot is the same, the task can change over different time slots. Note that we consider the challenging case where the platform cannot verify the workers' solutions, hence the platform and the workers do not know x_t, for all $t \in \mathcal{T}$.

2) Worker Effort Exertion and Solution Reporting Strategy: In each time slot t, each worker needs to decide whether to exert effort to complete the task, where the accuracy (i.e., quality) of his solution depends on his effort level. Let $e_{i,t} \in \{0, 1\}$ be a binary variable denoting worker i's effort level [1]. Spending effort with a cost $c_{i,t} > 0$ will improve the accuracy of a worker's solution. Specifically, the probability that worker i's estimated solution in time slot t is the same as the true solution of the task is given by

$$P(x_{i,t}^{\text{estimate}} = x_t) = \begin{cases} p^{\text{low}}, & \text{if } e_{i,t} = 0 \text{ (with zero cost)}, \\ p_{i,t} \in \left[p^{\text{high}}, 1\right], & \text{if } e_{i,t} = 1 \text{ (with a cost } c_{i,t} \geq 0), \end{cases} \quad (1)$$

where $0.5 < p^{\text{low}} < p^{\text{high}} < 1$ and the values of p^{low} and p^{high} are public knowledge to both the platform and the workers. We assume that even without any effort a worker still has some information about the true solution, so his estimate is more accurate than random guessing, i.e., $p^{\text{low}} > 0.5$.[2] Here, $p_{i,t}$ is the solution accuracy level (after effort exertion) of worker i in time slot t, and it may change over time (depending on the specific task). Although $p_{i,t}$ is worker i's private information, the platform knows its range $\left[p^{\text{high}}, 1\right]$ that is common among all workers.[3]

Each worker also needs to decide whether to truthfully report his solution to the platform. Let $v_{i,t} \in \{1, -1\}$ denote the reporting strategy of worker i, where $v_{i,t} = 1$ indicates truthful reporting and $v_{i,t} = -1$ indicates untruthful reporting,[4] i.e.,

$$x_{i,t}^{\text{report}} = \begin{cases} x_{i,t}^{\text{estimate}}, & \text{if } v_{i,t} = 1, \\ -x_{i,t}^{\text{estimate}}, & \text{if } v_{i,t} = -1. \end{cases} \quad (2)$$

For notational convenience, we use $s_{i,t} \triangleq (e_{i,t}, v_{i,t})$ to denote each worker's effort exertion and solution reporting strategy, where $s_{i,t}$ belongs to the set $\mathcal{S} = \{(0, 1), (0, -1), (1, 1), (1, -1)\}$.

3) Worker Accuracy Reporting Strategy: In practice, the platform may have no prior knowledge on the distribution of workers' solution accuracy levels. We assume that all $p_{i,t}$ (i.e., worker's solution accuracy level after effort exertion) are independently and identically distributed with a general cumulative distribution function $F(p)$ on support $[p^{\text{high}}, 1]$.[5] The platform knows the support but not the specific form of $F(p)$, hence needs to design an online learning mechanism to discover $F(p)$. Our mechanism requires the workers to report their $p_{i,t}$. A worker i can choose to report the value of $p_{i,t}^{\text{report}} \in \left[p^{\text{high}}, 1\right]$ in time slot t, and such value may not

[1]Binary-solution tasks are widely considered in literature [1], [2]. We can extend our analysis to the scenario where a task has more than two possible solutions, i.e., by decomposing a multiple-solution task into several binary-solution tasks [1].

[2]We can extend our analysis to the case where $p^{\text{low}} = 0.5$, i.e., without exerting effort a worker has no information about the true solution, so the estimated solution is equally likely to be right or wrong. However, this case can complicate the platform's reward design. We leave the details for interested readers to the online appendix [15].

[3]Since the focus of this paper is to model the case where workers have heterogeneous accuracy levels, we assume that workers have homogeneous costs. Specifically, we assume that $c_{i,t} = c > 0$, for all i, t.

[4]In fact, workers can benefit by colluding to always report 1 (or -1) as the task solution. However, in many online crowdsourcing platforms, workers are temporally and spacially separated and have very limited communications. Hence, we assume that workers report their solutions indenpendently and restrict their solution reporting strategies to either 1 or -1 [2], [6], [14].

[5]Note that workers' solution accuracy levels over time are considered independent and may change due to different task assignments. However, we assume that the overall distribution of the workers' solution accuracy levels does not vary over time. With a fixed worker population, the accuracy distribution can be treated as fixed when the assigned tasks over time have similar levels of difficulty [2].

be his true solution accuracy level, i.e., $p_{i,t}^{\text{report}}$ can be different from $p_{i,t}$.

4) Consistency Reward for Majority Voting: In each time slot t, after workers report their solutions, the platform will distribute *consistency rewards* to workers accordingly. If worker i's reported solution $x_{i,t}^{\text{report}}$ is aligned with the majority solution from the rest of the workers (denoted by $x_{-i,t}^{\text{majority}}$), he will receive a *consistency reward* $R_{i,t} \geq 0$. Specifically, the majority solution from worker i's perspective is

$$x_{-i,t}^{\text{majority}} = \begin{cases} 1, & \text{if} \quad \sum_{j\in\mathcal{N},j\neq i} x_{j,t}^{\text{report}} > 0, \\ -1, & \text{if} \quad \sum_{j\in\mathcal{N},j\neq i} x_{j,t}^{\text{report}} < 0, \\ \text{tie}, & \text{if} \quad \sum_{j\in\mathcal{N},j\neq i} x_{j,t}^{\text{report}} = 0. \end{cases} \tag{3}$$

If $x_{i,t}^{\text{report}} = x_{-i,t}^{\text{majority}}$, worker i receives $R_{i,t}$ for matching the majority solution. If $x_{-i,t}^{\text{majority}} = \text{tie}$, worker i also receives $R_{i,t}$, because his reported solution decides the majority solution from all the workers. If $x_{i,t}^{\text{report}} \neq x_{-i,t}^{\text{majority}}$ and $x_{-i,t}^{\text{majority}} \neq \text{tie}$, worker i will receive no reward, i.e., $R_{i,t} = 0$. Note that the consistency reward $R_{i,t}$ is affected by the workers' accuracy reports, which will be explained in Section III. We denote the probability that worker i receives $R_{i,t}$ as $P_{i,t}(\boldsymbol{p}_t^{\text{report}}, \boldsymbol{s}_t)$, where $\boldsymbol{p}_t^{\text{report}} = (p_{i,t}^{\text{report}}, \forall i \in \mathcal{N})$, and $\boldsymbol{s}_t = ((e_{i,t}, v_{i,t}), \forall i \in \mathcal{N})$.

5) Worker Payoff: : We define worker i's payoff in slot t as

$$u_{i,t}(\boldsymbol{p}_t^{\text{report}}, \boldsymbol{s}_t, R_{i,t}) = R_{i,t} \cdot P_{i,t}(\boldsymbol{p}_t^{\text{report}}, \boldsymbol{s}_t) - e_{i,t} \cdot c, \tag{4}$$

where $R_{i,t} \cdot P_{i,t}(\boldsymbol{p}_t^{\text{report}}, \boldsymbol{s}_t)$ captures the expected consistency reward and $-e_{i,t} \cdot c$ captures the cost for effort exertion.

B. Interactions between Workers and Platform

In this subsection, we introduce the interactions between the workers and the platform in each period of time slot t, respectively.

1) Worker Effort and Solution Reporting Game in Period III: In Period III of each time slot t, after workers reported accuracy levels $\boldsymbol{p}_t^{\text{report}} = \left(p_{i,t}^{\text{report}}, \forall i \in \mathcal{N}\right)$ in Period I, and the platform announced reward bundle $(R_{i,t}, \forall i \in \mathcal{N})$ in Period II, each worker decides effort exertion and solution reporting strategies.[6] Recall that workers obtain rewards if their solutions match the majority, hence their decisions on effort exertion and solution reporting affect other workers' payoffs in a game theoretical fashion.[7]

2) Platform Reward Design in Period II: In Period II of time slot t, the platform computes the reward bundle $(R_{i,t}, \forall i \in \mathcal{N})$ based on the accuracy reports in Period I, and anticipates the workers' decisions in Period III. Note that the platform does not know the workers' accuracy distribution, and it wants to obtain a good estimate of the distribution by designing the reward to incentivize the workers to truthfully report their solution accuracy levels. In this case, the platform's objective is to minimize the regret defined as follows:

$$\lim_{T\to\infty} \left(\max_{p\in[p^{\text{high}},1]} \left| \tilde{F}_T(p) - F(p) \right| \right), \tag{5}$$

where for any $p \in \left[p^{\text{high}}, 1\right]$, the empirical distribution (characterized by the cumulative distribution function) of the workers' accuracy reports is given by[8]

$$\tilde{F}_T(p) = \frac{\sum_{t=1}^{T} \sum_{i\in\mathcal{N}} \mathbb{1}_{p_{i,t}^{\text{report}} < p}}{NT}. \tag{6}$$

The platform's objective is to minimize the regret, which indicates the difference between the empirical distribution and the actual distribution of the workers' solution accuracy levels.[9]

Next, we formulate the platform's regret minimiation problem. The platform chooses the consistency reward to minimize the asymptotic regret. The problem is formulated as follows:

Problem 1. *(Platform's Reward Design Problem in Period II)*

$$\begin{aligned} &\min \lim_{T\to\infty} \left(\max_{p\in[p^{\text{high}},1]} \left| \tilde{F}_T(p) - F(p) \right| \right) \\ &\text{var. } R_{i,t} \in [0, R^{\max}], \ \ \forall i \in \mathcal{N}, \ t \in \mathcal{T}. \end{aligned} \tag{7}$$

Here, the consistency reward is upper bounded by a finite value $R^{\max}$, modeling that the platform cannot provide an arbitrarily large reward (e.g., due to budget concerns [16]).

3) Worker Accuracy Reporting Game in Period I: In Period I of time slot t, workers independently report their accuracy levels, anticipating the platform's decision in Period II and the workers' own equilibrium decisions in Period III. The accuracy reporting game among the workers is formulated as follows.

Game 1. *(Workers' Accuracy Reporting Game in Period I of slot t) In Period I of each time slot t, the workers' accuracy reporting game is a Bayesian game, i.e., a tuple $\Omega = (\mathcal{N}, \mathcal{P}, \mathcal{Y}, \boldsymbol{u}_t^{arg}, F)$ that consists of*

- Players*: The set $\mathcal{N}$ of workers.*
- Strategies*: Each worker chooses his accuracy reporting strategy $p_{i,t}^{\text{report}} \in \mathcal{P}_i \triangleq \left[p^{\text{high}}, 1\right]$. The accuracy reporting strategy profile of all the workers is $\boldsymbol{p}_t^{\text{report}} = (p_{i,t}^{\text{report}}, \forall i \in \mathcal{N})$ and the set of feasible strategy profile of all the workers is $\mathcal{P} = \prod_{i\in\mathcal{N}} \mathcal{P}_i$.*
- Types*: Each worker i's type is his accuracy level $p_{i,t} \in \mathcal{Y} \triangleq \left[p^{\text{high}}, 1\right]$, which is his private information.*
- Payoffs*: The vector $\boldsymbol{u}_t^{arg} = (u_{i,t}^{arg}, \forall i \in \mathcal{N})$ contains all workers' payoffs as defined in (8).*
- Type Distribution*: Each worker i's type $p_{i,t}$ is drawn from a common distribution with cdf $F(\cdot)$ on support $[p^{\text{high}}, 1]$.*

[6]Here workers are informed of the announced reward bundle, not the private reported accuracy levels. Further as will be seen, from the announced rewards a worker cannot infer what accuracy levels other workers reported.

[7]The detailed game formulation is left to appendix [15] due to space limits. We assume that workers know the distribution $F(p)$. This models the scenario where workers know each other, e.g., in peer grading, students from the same class may know other students' background and capabilities well.

[8]Note that in (6), $\mathbb{1}$ is an indicator function, i.e., $\mathbb{1}_{p_{i,t}^{\text{report}} < p} = 1$ if $p_{i,t}^{\text{report}} < p$, and $\mathbb{1}_{p_{i,t}^{\text{report}} < p} = 0$ if $p_{i,t}^{\text{report}} \geq p$.

[9]Our model is also applicable to other objectives such as minimizing the regret between the platform's tradeoff under estimated accuracy distribution and that under actual accuracy distribution. The tradeoff here can be a balance between the quality of the workers' solutions and the total rewards [6].

Since the accuracy reporting game is a Bayesian game, we define the expected payoff for a worker i in time slot t as

$$u_{i,t}^{arg}(p_{i,t}^{\text{report}}; R_{i,t}, \boldsymbol{p}_{-i,t}^{\text{report}}, \boldsymbol{s}_t) = \int u_{i,t}^{\text{blf}}(p_{i,t}^{\text{report}}; R_{i,t}, \boldsymbol{p}_{-i,t}^{\text{report}}, s_{i,t}, \boldsymbol{s}_{-i,t}(\boldsymbol{p}_{-i,t}))dF(\boldsymbol{p}_{-i,t}|p_{i,t}), \tag{8}$$

where $u_{i,t}^{\text{blf}}(p_{i,t}^{\text{report}}; R_{i,t}, \boldsymbol{p}_{-i,t}^{\text{report}}, s_{i,t}, \boldsymbol{s}_{-i,t}(\boldsymbol{p}_{-i,t}))$ is worker i's payoff in time slot t, given any belief $\boldsymbol{p}_{-i,t}$. Due to space limits, its expression is left to the online appendix [15].

We apply ϵ-approximate Bayesian Nash equilibrium (ϵ-BNE) as the solution concept for Game 1 [17]. This equilibrium concept is defined as follows.

Definition 1. *(ϵ-BNE) A set of reporting strategy $\left(\boldsymbol{p}_i^{\text{report}*}, \forall i \in \mathcal{N}\right)$ with $\boldsymbol{p}_i^{\text{report}*} = \left(p_{i,t}^{\text{report}*}, \forall t \in \mathcal{T}\right)$ is an ϵ-BNE if for any i and $\boldsymbol{p}_i^{\text{report}'} \neq \boldsymbol{p}_i^{\text{report}*}$, we have*

$$\begin{aligned} &\frac{1}{T}\sum_{t=1}^{T} u_{i,t}\left(p_{i,t}^{\text{report}*}, \boldsymbol{p}_{-i,t}^{\text{report}*}; R_{i,t}, \boldsymbol{s}_t\right) \\ &\qquad \geq \frac{1}{T}\sum_{t=1}^{T} u_{i,t}\left(p_{i,t}^{\text{report}'}, \boldsymbol{p}_{-i,t}^{\text{report}*}; R_{i,t}, \boldsymbol{s}_t\right) - \epsilon. \end{aligned} \tag{9}$$

At an ϵ-BNE, it is not possible for a worker to increase its payoff by more than ϵ via unilaterally deviating from his equilibrium strategy. Note that ϵ-BNE is a generalized concept of BNE, where no worker can increase his payoff via unilaterally deviating from his equilibrium strategy. An ϵ-BNE with $\epsilon = 0$ is a BNE. As we will show, our mechanism induces an ϵ-BNE where ϵ approaches zero as T increases.

Recall that the platform's goal is to learn the workers' accuracy distribution by asking them to report their accuracy levels. The platform needs to carefully design the reward mechanism to incentivize the workers to truthfully report their accuracy levels, so that the learning is accurate. We present the reward mechanism design in the following section.

III. Online Crowd Learning Mechanism

In this section, we present our online crowd learning mechanism based on majority voting. The mechanism aims to encourage the workers to truthfully report their accuracy levels. However, without information about the workers' accuracy distribution, majority voting provides the (wrong) incentive for workers to under-report their accuracy levels. A lower accuracy (from under-reporting) makes the platform believe that this worker has a smaller chance of matching the majority. Hence, to compensate the cost of effort exertion, the platform needs to use a larger reward. To address this issue, we propose a randomized reward design. Moreover, after the platform estimates the rewards based on workers' accuracy reports, it needs to add a positive term to the estimated rewards in order to reduce the bias. The bias comes from the imperfect estimation of the accuracy distribution due to a finite number of accuracy reports, as well as the workers' potenial accuracy misreports.

Mechanism 1 provides the details of the online mechanism. In **Step 1** the workers submit their accuracy reports; in **Step 2** the platform randomly selects a threshold and uses this to set the reward in **Step 3**.

Mechanism 1 Online Crowd Learning Mechanism

initialization: set $t = 1$;
while $t \leq T$ **do**
  Assign an identical task to all the workers;
  Step 1 (Reporting Accuracy): Workers voluntarily submit accuracy reports $\left(p_{i,t}^{\text{report}}, \forall i \in \mathcal{N}\right)$;
  if worker i does not submit **then**
    $p_{i,t}^{\text{report}} \leftarrow p^{\text{low}}$;
  end if
  Step 2 (Selecting Threshold): The platform randomly selects $\bar{p}_t$ uniformly from support$\left[p^{\text{high}}, 1\right]$;
  Step 3 (Setting Reward):
  if $p_{i,t}^{\text{report}} < \bar{p}_t$ **then**
    $R_{i,t} \leftarrow 0$;
  else
    Estimate $\tilde{R}_{i,t}$ via (13) and compute δ_t via (11);
    $R_{i,t} \leftarrow \tilde{R}_{i,t} + \delta_t$;
  end if
  $t \leftarrow t + 1$;
end while

In **Step 2**, the reason to select a random threshold $\bar{p}_t$ is to incentivize the workers to truthfully report their accuracy levels. Such a random selction offsets the benefit a worker obtains from under-reporting his accuracy and compensates the loss a worker suffers from over-reporting his accuracy. Specifically, if a worker i under-reports his accuracy, he will have a larger chance of matching the majority solution. This is because the platform will use his accuracy report to calculate other workers' rewards, and his under-reporting will lead other workers' rewards to be larger. With larger rewards, other workers are more likely to exert effort (and truthfully report solutions). As a result, worker i will have a larger chance of matching the majority solution. However, when the platform randomly sets $\bar{p}_t$, under-reporting increases the probability that $p_{i,t}^{\text{report}} < \bar{p}_t$ and hence increases the probability of getting a zero reward. This reduces the benefit a worker obtains from under-reporting his accuracy level. Similarly, if a worker i over-reports his accuracy level, he will have a smaller chance of matching the majority solution. When the platform randomly sets $\bar{p}_t$, such a loss is compensated by decreasing the possibility of getting a zero reward.

In **Step 3**, for workers with $p_{i,t}^{\text{report}} \geq \bar{p}_t$, the platform will first estimate a reward $\tilde{R}_{i,t}$, using only the accuracy reports submitted by all the workers excluding i during all previous time slots (including current time slot). Then, the platform adds a positive term δ_t to determine the final reward, where δ_t helps reduce the bias discussed previously. Specifically, the calculation of the final reward is done as follows:

$$P_{N-1}^{\text{majority}}\left(\bar{p}_t\right)=1-\frac{N-1}{2N} -\frac{1}{N}\mathbb{E}\left\{\sum_{n=1}^{N-1}\left\{\frac{1-e^{-j\pi n(N-1)/N}}{1-e^{-2j\pi n/N}}\cdot\prod_{k=1}^{N-1}\left\{\left(p_k\mathbb{1}_{p_k\geq\bar{p}_t}+p^{\text{low}}\mathbb{1}_{p_k<\bar{p}_t}\right)e^{2j\pi n/N}+\left(1-\left(p_k\mathbb{1}_{p_k\geq\bar{p}_t}+p^{\text{low}}\mathbb{1}_{p_k<\bar{p}_t}\right)\right)\right\}\right\}\right\}. \tag{10}$$

$$\delta_t=\frac{2c\left(\bar{p}_t-0.5\right)\sqrt{\frac{2\ln t}{(N-1)t}}}{\left(\left[2P^{\text{low}}-1\right]\left(\bar{p}_t-0.5\right)-\gamma\right)^2}+\frac{c}{\left[2P_t^{\text{a}}\left(\bar{p}_t,\epsilon_t\right)-1\right]\left(\bar{p}_t-0.5\right)-\gamma}-\frac{c}{\left[2P_t^{\text{b}}\left(\bar{p}_t\right)-1\right]\left(\bar{p}_t-0.5\right)-\gamma} +\frac{c\epsilon_t}{\left(\left[2P^{\text{low}}-1\right]\left(\bar{p}_t-0.5-\epsilon_t\right)-\gamma\right)\left(\left[2P^{\text{low}}-1\right]\left(\bar{p}_t-0.5\right)-\gamma\right)}. \tag{11}$$

1) The platform calculates the empirical distribution $\tilde{F}_{i,t}(p)$ for each worker i, i.e., for all $p\in[0.5,1]$,

$$\tilde{F}_{i,t}(p)=\frac{\sum_{\tau=1}^{t}\sum_{j\in\mathcal{N},j\neq i}\mathbb{1}_{p_{j,\tau}^{\text{report}}\leq p}}{(N-1)t}. \tag{12}$$

2) The platform estimates $\tilde{R}_{i,t}$ for each worker i according to the following equality:

$$\tilde{R}_{i,t}=\frac{c}{\left(2P_{N-1}^{\text{majority}}\left(\bar{p}_t\right)-1\right)\left(\bar{p}_t-0.5\right)-\left(p^{\text{low}}-0.5\right)}, \tag{13}$$

where $P_{N-1}^{\text{majority}}\left(\bar{p}_t\right)$ is the probability that the majority among $N-1$ workers provide the correct solution. Its expression is given in (10), where $j=\sqrt{-1}$ and p_k is generated from the distribution with cdf $\tilde{F}_{i,t}(p)$ in (12). The choice of the estimated reward $\tilde{R}_{i,t}$ in (13) assumes that workers' accuracy reports are truthful and unbiased. Hence, we need to add a term δ_t (to be defined as follows) to cancel the effect from potential accuracy misreports as well as offset the bias.

3) The platform adds a uniform time-dependent term δ_t in the form of (11) to each estimated reward $\tilde{R}_{i,t}$, and announces the final reward $R_{i,t}=\tilde{R}_{i,t}+\delta_t$ to each worker i. In (11), the expressions of $\gamma,\epsilon_t,P^{\text{low}},P_t^a(\bar{p}_t,\epsilon_t)$, and $P_t^b(\bar{p}_t)$ are given in the appendix [15] due to space limits.

Adding δ_t ensures that the final reward can incentivize workers with $p_{i,t}^{\text{report}}\geq\bar{p}_t$ to exert effort and truthfully report his solution, i.e., $s_{i,t}=(1,1)$, while workers with $p_{i,t}^{\text{report}}<\bar{p}_t$ to exert zero effort and truthfully report his solution, i.e., $s_{i,t}=(0,1)$. This enables the platform to anticipate the workers' effort exertion and solution reporting strategies based on their accuracy reports.

As we will show in Theorem 1, with the mehcanism being repeatedly implemented, as time goes by, workers' accuracy reports will approach their actual accuracy levels. At the very beginning, a worker may benefit from misreporting accuracy, but over time, the effect of a worker's accuracy misreport on others' strategies (on effort exertion and solution reporting) becomes smaller.[10] This in turn provides less incentive for a worker to misreport his accuracy level. As a result, workers tend to truthfully report their accuracy levels in the long run.

IV. Worker Decisions

In this section, we will analyze the workers' decisions under **Mechanism 1**.

We first characterize workers' effort exertion and solution reporting strategies in Proposition 1.

Proposition 1. *In Period III of each time slot t, with a probability at least $1-\frac{1}{t^4}$, we have*

$$s_{i,t}^*=\begin{cases}(1,1), & \text{if} \quad p_{i,t}^{\text{report}}\geq\bar{p}_t,\\ (0,1), & \text{if} \quad p_{i,t}^{\text{report}}<\bar{p}_t.\end{cases} \tag{14}$$

As stated previously, adding δ_t ensures a worker with $p_{i,t}^{\text{report}}\geq\bar{p}_t$ will exert effort and truthfully report solutions, i.e., use $(1,1)$. Note that this happens with a probability at least $1-\frac{1}{t^4}$. The uncertainty here is due to the sampling bias of the workers' solution accuracy levels. For example, suppose that all workers truthfully report solution accuracy levels. With a small probability, the empirical accuracy distribution can still differ with the true distribution by a large amount. Note that as time goes by (i.e., t increases), workers are more likely to adopt strategies shown in (14), hence the platform can anticipate the workers' behaviors with a higher accuracy.

Next, we characterize the workers' accuracy reporting strategies. Recall that we adopt ϵ-BNE as the equilibrium concept. As ϵ decreases, the condition for a strategy profile to constitute an equilibrium becomes stronger. We focus on the case where ϵ asymptotically converges to zero. In this case, an ϵ-BNE converges to a BNE. We show the detials in Theorem 1.

[10] Note that a worker's reward is calculated by other workers' accuracy reports from all previous time slots.

Theorem 1. *There exists an* $O\left(\sqrt{\frac{\ln T}{T}}\right)$*-BNE for Game 1 in Period I of slot* t*, where each worker* i*'s reporting strategy* $p_{i,t}^{\text{report}}$ *satisfies*

$$\max\left\{p_{i,t}-\sigma_t, p^{\text{high}}\right\} \leq p_{i,t}^{\text{report}} \leq \min\{p_{i,t}+\sigma_t, 1\}, \quad (15)$$

where $\sigma_t = O\left(\sqrt{\frac{\ln t}{t}}\right)$.

The implication behind Theorem 1 is as follows. In each slot t, if each worker i's accuracy misreport is upper-bounded by $\sigma_t = O\left(\sqrt{\frac{\ln t}{t}}\right)$, then the allowed deviation of profit a worker gains over the entire time horizon T (i.e., ϵ) is upper bounded by $O\left(\sqrt{\frac{\ln T}{T}}\right)$, which converges to zero as T increases.

There are two things to note. First, the accuracy reporting strategy in the form of (15) can correspond to infinitely many possible cases with reporting deviation being upper bounded by the same order of $O\left(\sqrt{\frac{\ln t}{t}}\right)$. This shows that our mechanism is fairly general and is capable of characterzing a large set of equilibrium strategies. Second and more importantly, truthful reporting of the accuracy from the very beginning (i.e., $t = 1$) also belongs to such ϵ-BNE, since (15) holds with $p_{i,t}^{\text{report}}$ replaced by $p_{i,t}$, for all i, t. It means that our mechanism can induce truthful accuracy reporting, which converges to a BNE. This is a desirable property.

V. Platform's Reward Design

In this section, we solve the platform's problem in (7). The platform decides $(R_{i,t}, \forall i \in \mathcal{N}, t \in \mathcal{T})$ to minimize the asymptotic regret which indicates the difference between the empirical accuracy distribution and the actual distribution.

Suppose workers follow ϵ-BNE in the form of (15). We bound the platform's regret in Theorem 2.

Theorem 2. *If the platform sets* $(R_{i,t}, \forall i \in N, \forall t \in \mathcal{T})$ *according to* ***Mechanism 1****, then the platform's regret is asymptotically zero. That is, for any* $\psi \in (0, 1]$*, we have*

$$|\tilde{F}_T(p) - F(p)| \leq \sqrt{\frac{\ln \frac{2}{\psi}}{2NT}} + O\left(\sqrt{\frac{\ln T}{T}}\right), \forall p \in \left[p^{\text{high}}, 1\right], \quad (16)$$

with probability at least $1 - \psi$*, where*

$$\lim_{T\to\infty}\left(\sqrt{\frac{\ln \frac{2}{\psi}}{2NT}} + O\left(\sqrt{\frac{\ln T}{T}}\right)\right) = 0. \quad (17)$$

Theorem 2 implies that the choice of $R_{i,t}$ according to **Mechanism 1** is an optimal solution to Problem (7), as it achieves an asymptotic zero regret. Note that there may exist other values of $R_{i,t}$ that induce other different types of ϵ-BNEs, which may also yield zero regrets with potentially different convergence rates. We leave a more detailed discussion on the convergence rates to future work.

VI. Numerical Results

In this section, we provide simulation results to investigate the impact of workers' accuracy reporting strategies and workers' characteristics on the overall system performance. As will be shown, our mechanism renders workers to truthfully report accuracy levels, which achieves the largest worker payoff and smallest platform regret. Moreover, the mechanism performs better when the workers are more capable of solving tasks.

A. Simulation Setup

We consider that workers' accuracy levels follow a two-point distribution,[11] i.e.,

$$p_{i,t} = \begin{cases} p_l, & \text{w.p.} \quad z, \\ p_h, & \text{w.p.} \quad 1-z, \end{cases} \quad (18)$$

where $p^{\text{high}} \leq p_l < p_h \leq 1$ and $0 \leq z \leq 1$. To evaluate the impact of workers' reporting strategies, we consider three different types of strategies for workers, which are as follows:

- *Random Reporting (RR)*: all the workers randomly report their accuracy levels, i.e., $Pr(p_{i,t}^{\text{report}} = p_l) = Pr(p_{i,t}^{\text{report}} = p_h) = 0.5$, for all i, t. This reporting strategy serves as a benckmark.
- *Truthful Reporting (TR)*: all the workers truthfully report accuracy levels, i.e., $p_{i,t}^{\text{report}} = p_{i,t}$, for all i, t. This is a special case of workers' reporting in (15), i.e., an equilibrium that can be induced by our proposed mechanism.
- *Asymptotically Truthful Reporting (ATR)*: all the workers follow the reporting strategy specified in (15), and they are allowed to misreport. Since (15) contains infinitely many cases, we choose $\sigma_t = \sqrt{\frac{\ln t}{t}}$ and consider the scenario where workers randomly report their accuracy levels when $\sqrt{\frac{\ln t}{t}} \geq p_h - p_l$, and workers truthfully report when $\sqrt{\frac{\ln t}{t}} < p_h - p_l$, i.e.,

$$p_{i,t}^{\text{report}} = \begin{cases} p_l \text{ or } p_h \text{ with equal prob.,} & \text{if } \sqrt{\frac{\ln t}{t}} \geq p_h - p_l, \\ p_{i,t}, & \text{if } \sqrt{\frac{\ln t}{t}} < p_h - p_l. \end{cases} \quad (19)$$

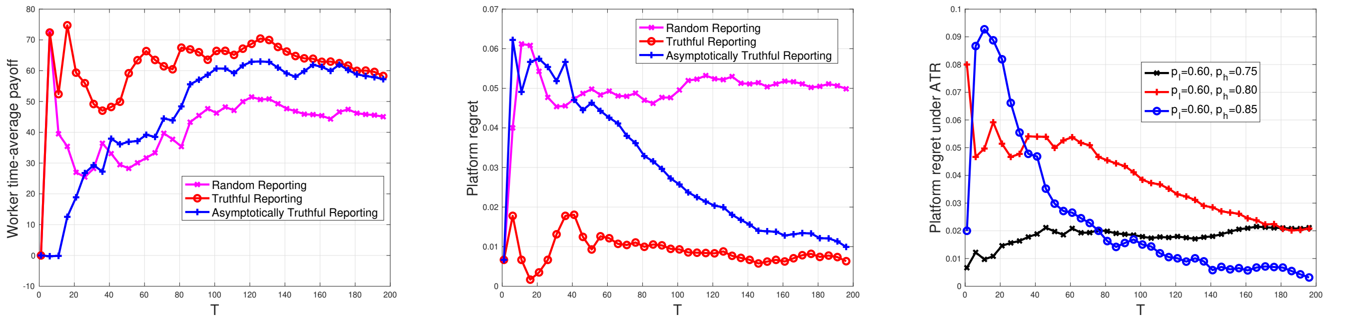
Under this strategy, a worker randomly reports initially and then keeps truthfully reporting. Notice that both TR and ATR satisfy (15), whereas ATR may be more reasonable in practice as it allows possible accuracy misreports.

Moreover, we set $c = 1$, $N = 100$, $p^{\text{low}} = 0.51$, $p^{\text{high}} = 0.55$, $z = 0.8$, $p_l = 0.6$, choose p_h from set $\{0.75, 0.80, 0.85\}$, and implement the mechanism for 200 times.

B. Impact of Worker Reporting Strategies and Worker Characteristics on Overall System Performance

To study the impact of workers' reporting strategies, we first examine how the time-average payoff of a randomly selected

[11] We choose a simple two-point distribution to facilitate the interpretation of insights. We leave a more comprehensive numerical study of our mechanism on various distributions to future work.



(a) Worker time-average payoff vs T.

(b) Platform regret vs T.

(c) Platform regret vs T under different p_h.

Fig. 2: Impact of worker reporting strategies and worker characteristics on the overall system performance.

worker i, $\frac{1}{T}\sum_{t\in\mathcal{T}} u_{i,t}(p_{i,t}^{\text{report}}, s_{i,t}; R_{i,t}, \boldsymbol{p}_{-i,t}^{\text{report}}, \boldsymbol{s}_{-i,t})$, is affected by different reporting strategies (see Fig. 2a). Moreover, we study how the platform's regret ($|\tilde{F}_T(p)-F(p)|$) is affected by different reporting strategies (see Fig. 2b). To investigate the impact of worker characteristics, we examine how the platform regret is affected by p_h, supposing that workers use asymptotically truthful reporting strategy (see Fig. 2c).

Impact of reporting strategies on worker payoff. In Fig. 2a, we observe that TR yields the highest average worker payoff, ATR leads to an intermediate payoff, and RR achieves the lowest worker payoff (e.g., $T = 200$). Our mechanism not only induces truthful reporting as an asymptotic BNE, but also the workers adopting the truthful reporting strategy achieve the highest average payoffs. This is desirable as the mechanism motivates the workers to truthfully report their accuracy levels.

Impact of reporting strategies on platform regret. In Fig. 2b, we observe that the regret under ATR converges to the regret under TR, and is smaller than that under RR. This is because the platform adopts an empirical estimate of the worker accuracy distribution via (6). As T increases, the workers under ATR are more likely to truthfully report their accuracy (the condition $\sqrt{\frac{\ln t}{t}} \geq p_h - p_l$ is harder to satisfy). Hence, the regret under ATR converges to that under TR.

Impact of worker characteristics on platform regret. In Fig. 2c, as p_h increases, the platform's regret demonstrates an earlier decreasing tendency (e.g., the decrease appears at $T = 10$ for the blue curve and $T = 60$ for the red curve) and achieves a smaller regret (e.g., $T = 200$). Given T, as p_h increases, the condition $\sqrt{\frac{\ln t}{t}} \geq p_h - p_l$ is easier to be violated. As a result, the workers are more likely to truthfully report their accuracy levels, which helps achieve a smaller regret. In other words, our mechanism has a better performance when the workers are more capable of solving the tasks.

VII. Conclusion

In this paper, we present a crowd learning mechanism to learn the workers' accuracy distribution, which helps solve an information elicitation without verification (IEWV) problem. We formulate the interactions between the workers and the platform as a sequential repeated game, and compute in close-form the reward design that achieves the best system performance. We show that our mechanism can incentivize workers to truthfully report their solution accuracy levels in the long term. Moreover, our mechanism converges faster when the workers are more capable of solving the tasks.

For the future work, we plan to study the mechanism design under multi-dimensional worker heterogeneity, where both the workers' costs and accuracy levels are heterogeneous.

References

[1] J. Xu, S. Wang, N. Zhang, F. Yang, and X. S. Shen, "Reward or penalty: Aligning incentives of stakeholders in crowdsourcing," *IEEE Transactions on Mobile Computing*, 2018.

[2] Y. Liu and Y. Chen, "Learning to incentivize: Eliciting effort via output agreement," *arXiv preprint arXiv:1604.04928*, 2016.

[3] https://www.waze.com.

[4] https://openreview.net.

[5] Y. Kong, K. Ligett, and G. Schoenebeck, "Putting peer prediction under the micro (economic) scope and making truth-telling focal," in *International Conference on Web and Internet Economics*. Springer, 2016, pp. 251–264.

[6] C. Huang, H. Yu, J. Huang, and R. Berry, "Crowdsourcing with heterogeneous workers in social networks," in *Proc. of IEEE GLOBECOM*, 2019.

[7] B. Waggoner and Y. Chen, "Output agreement mechanisms and common knowledge," in *Proc. of AAAI HCOMP*, 2014.

[8] N. Miller, P. Resnick, and R. Zeckhauser, "Eliciting informative feedback: The peer-prediction method," *Management Science*, vol. 51, no. 9, pp. 1359–1373, 2005.

[9] http://microworkers.com.

[10] Z. Li, H. Liu, and R. Wang, "Service benefit aware multi-task assignment strategy for mobile crowd sensing," *Sensors*, vol. 19, no. 21, p. 4666, 2019.

[11] J. Witkowski, B. Nebel, and D. C. Parkes, "Robust peer prediction mechanisms," Ph.D. dissertation, University of Freiburg, Freiburg im Breisgau, Germany, 2015.

[12] Y. Kong and G. Schoenebeck, "Equilibrium selection in information elicitation without verification via information monotonicity," *arXiv preprint arXiv:1603.07751*, 2016.

[13] G. Liao, X. Chen, and J. Huang, "Optimal privacy-preserving data collection: A prospect theory perspective," in *Proc. of IEEE GLOBECOM*, 2017, pp. 1–6.

[14] C. Huang, H. Yu, J. Huang, and R. Berry, "Incentivizing crowdsourced workers via truth detection," in *Proc. of IEEE GlobalSIP*, 2019.

[15] http://jianwei.ie.cuhk.edu.hk/publication/AppendixChaoWiOpt20.pdf.

[16] D. Zhao, X.-Y. Li, and H. Ma, "How to crowdsource tasks truthfully without sacrificing utility: Online incentive mechanisms with budget constraint," in *Proc. of IEEE INFOCOM*, 2014, pp. 1213–1221.

[17] V. Bosshard, B. Bünz, B. Lubin, and S. Seuken, "Computing Bayes-nash equilibria in combinatorial auctions with continuous value and action spaces." in *Proc. of ACM IJCAI*, 2017, pp. 119–127.