Abstract

The artificial intelligence (AI) industry has created new jobs that are essential to the real world deployment of intelligent systems. Part of the job focuses on labelling data for machine learning models or having workers complete tasks that AI alone cannot do. These workers are usually known as ‘crowd workers’—they are part of a large distributed crowd that is jointly (but separately) working on the tasks although they are often invisible to end-users, leading to workers often being paid below minimum wage and having limited career growth. In this chapter, we draw upon the field of human–computer interaction to provide research methods for studying and empowering crowd workers. We present our Computational Worker Leagues which enable workers to work towards their desired professional goals and also supply quantitative information about crowdsourcing markets. This chapter demonstrates the benefits of this approach and highlights important factors to consider when researching the experiences of crowd workers.

Keywords

crowdsourcing, crowd work, crowd labour, Amazon Mechanical Turk, turker, crowd market, fair labour, fair wages, skill development, plug-ins

9

Research Methods to Study and Empower Crowd Workers

Saiph Savage, Carlos Toxtli, Eber Betanzos

Introduction

Individuals are beginning to discover that digital platforms offer work opportunities and to find that they can earn a living from this work (Kuhn and Maleki 2017). Pew Research Center reported that by mid-2016 around 20 million Americans said that they had earned money through labour they completed on digital platforms in the previous year (Smith 2016). Jobs on digital platforms are expected to add $2.7 trillion to the global GDP by 2025. These digital platforms are typically coined ‘crowd marketplaces’ because employers offer temporary jobs in crowdsourcing platforms to workers on the platform (who are thus coined ‘crowd workers’). For a growing number of individuals, crowd markets have become an important new source of income (Abraham et al. 2018). An especially popular crowd market is Amazon Mechanical Turk (MTurk) (Amazon 2020). The labour that is posted on crowd markets like MTurk is usually known as ‘human-intelligence-tasks’ (Eickhoff and de Vries 2011). These tasks are tasks that artificial intelligence (AI) by itself cannot do and hence human intelligence is necessary. Together, crowd markets not only help to power our AI industry, but these markets also inject millions of new jobs into the economy (Smith 2019).

The new jobs that the AI industry has created on crowd markets are essential to enable the real-world deployment of intelligent systems. Part of these new jobs typically focuses on labelling data for machine learning models (Sorokin and Forsyth 2008). For instance, a number of tasks on MTurk focus on having human workers label where pedestrians appear in blurry videos (Hall and Perona 2015), which can be difficult for machines to do on their own. The labelled videos are then fed into machine learning models that use the data to learn to detect pedestrians from non-pedestrians. These machine learning models can then be inserted into autonomous vehicles. Indeed, the human labour behind our AI has powered self-driving cars, voice-based virtual assistants and search results with minimum hate speech (Young et al. 2018). However, the crowd workers powering the AI industry are often invisible to end-users. Their invisibility has exacerbated power imbalances where workers are often paid below the minimum wage (Hara et al. 2019) and have limited career growth (Deng and Joshi 2013). Part of the problem emerges because much of the algorithmic designs and platform choices of these crowd markets have focused on privileging employers and have not considered how they might harm workers. These types of designs can lead to unintended consequences that involve cruelty towards workers that is perpetrated by algorithms (henceforth referred to as 'algorithmic cruelty') for example platforms that automatically terminate workers (Jagabathula et al. 2014), resulting in workers losing their livelihoods and experiencing stress (Gray and Suri 2019).

In this chapter, we outline how human–computer interaction can re-envision the reality of crowd workers to improve their labour conditions. In particular, we present design criteria for tools that can serve to transform crowd markets and drive positive social change. Our design criteria are based on social theory that highlights how humans flourish when connected with the social essence of work (Yeoman 2014). For this purpose, we develop systems and computational methods that create on-demand 'professional leagues' for crowd workers. Our leagues focus on orchestrating workers to produce collective action while having social conversations. This collective social action enables: (1) increasing workers' wages (Savage et al. 2020); (2) enabling workers' skill development (Chiang et al. 2018b; Toxtli and Savage 2020; Hanrahan et al. 2020); and (3) driving justice in employers' evaluations of workers (Gaikwad et al. 2015). Unlike prior work that concentrated on primarily providing better communication and transparency (Huang and Fu 2013), the design criteria we present here open a new area of research focused on computationally orchestrating crowd workers to actively drive positive change in their professional lives.

Prior Research Methods to Investigate Crowd Work

Prior methods used for investigating crowd work can be broadly divided into two main strands: (1) qualitative studies for understanding crowd work (Hilton and Azzam 2019); (2) tools and platforms for improving crowd work (Kaplan et al. 2018). In the following section we present an overview of these methods and discuss where our research methods fall within this schema.

Qualitative studies for understanding crowd work

To understand the new work dynamics that are emerging in these crowd markets, some research has focused on conducting surveys and interviews with crowd workers and their employers to start to understand: how these crowd markets are functioning (Berg 2015; Hara et al. 2019; Kasunic et al. 2019); the type of dynamics that are emerging between the stakeholders of these markets (workers, employers, and platform owners) (Toxtli et al. 2019); and the challenges and opportunities that emerge in these markets (Slivkins and Vaughan 2014). Several of these qualitative research studies have been critical to understanding the precarious conditions in which crowd workers operate as well as helping researchers and practitioners to understand the unjust power dynamics and algorithmic cruelty that takes place within crowd markets (Gray and Suri 2019, Toxtli et al. 2021). We build on these previous studies to influence and improve crowd work. Additionally, the studies serve to identify critical points we can improve through systems design.

However, it is important to note that these qualitative studies based on self-reports from workers, while they are rich in information about workers' direct experiences in crowd markets, generally lack quantitative data based on direct measurement of crowd markets. Data from direct measurement or observation are important because they can help us to understand crowd markets from another perspective. For instance, crowd workers might state via surveys or interviews that they have low wages, and that the tasks they do within the crowd market do not facilitate their career growth. Without explicit log data, it can be difficult to understand just how low wages are. It is also difficult to understand the exact characteristics of the labour that crowd workers are exposed to, and that might not help their growth. This limited view makes it difficult to design adequate technologies and socio-technical interventions that can create and provide transformative change within crowd markets.

Tools and platforms for improving crowd work

Other research has focused on designing tools and platforms that aim to directly improve crowd work (Jarrahi and Sutherland 2019). The vast majority of research in this space has focused on studying the platform and system designs that are most appropriate for producing higher quality labour in these crowd markets while reducing costs (time, wages, etc.) (Allahbakhsh et al. 2013). These investigations are very much focused on the interests of employers and the owners of crowd markets (Singer and Mittal 2013). Crowd workers under these settings are sometimes viewed as ‘clogs’ in a pipeline whose operation needs to be optimized (Bernstein et al. 2012). These designs have led to circumstances where workers are earning less than minimum wage and have little opportunity to develop themselves (Saito et al. 2019b).

In part inspired by the qualitative research methods that identified the difficult living situations of crowd workers, a portion of investigations have also focused on devising tools and platforms that are more worker-centric. Part of the narrative around the worker-centric research is that crowd markets entail algorithmic cruelty in part because these markets have not been designed to provide transparency to workers. Economists consider that a market is transparent when all actors can access information about the market, such as products, services, or capital assets (Strathern 2000). The problem is that crowd markets have, for the most part, limited the amount of information that workers have about employers. Usually, employers (the individuals who post the labour they want workers to do) are granted access to information concerning the events in the marketplace, while workers have a much more limited perspective (Irani and Silberman 2013, 2016). For example, Amazon Mechanical Turk allows employers to view the previous performances and interactions that workers have had on the platform (Hara et al. 2018); while workers can discern very little about what employers have done previously (e.g. it is not possible to easily gauge whether a certain employer is paying unfair wages to workers, or whether the employer is actually a fraud looking to steal workers' data or obtain free labour; Irani and Silberman 2013; Gadiraju and Demartini 2019). This lack of transparency for workers can lead them to invest significant time in completing certain labour but receive anywhere from inadequate to no compensation. Six Silberman discusses how the lack of transparency on crowd markets affects workers earnings: 'A wide range of processes that shape platform-based workers' ability to find work and receive payment for work completed are, on many platforms, opaque' (Metall 2016).

To begin addressing the issue of transparency, Irani and colleagues (Irani and Silberman 2013) explored creating computational tools and forums through which workers could share information about crowd markets. Practitioners also started following similar efforts with the goal of empowering workers to share concrete and useful information about the crowd markets in which they worked (ChrisTurk 2018). The goal was that this information could help workers better navigate the crowd market and ultimately lead workers to have better working conditions. These tools and forums provide crowd workers with otherwise unavailable information about the employers, tasks, and expected payments within a crowd market. For instance, the computational tool of Turkopticon[1] (a popular tool used to bring

1 Turkopticon is an online tool that helps workers to see the reputation of employers (requesters) on crowdsourcing markets to help workers decide if they should work with those employers based on how much they typically pay, if they reject work, as well as other metrics. Turkopticon is built as a plug-in to be directly embded inside the crowdsourcing market. The use of plug-ins facilitates repurposing crowdsourcing markets to better suit the needs of workers. For instance, in this case the crowdsourcing market of Amazon Mechanical Turk did not officially

workers transparency) allows workers to obtain an overview of the expected hourly wage they would receive if they worked for a particular employer. This value is calculated based on what other workers have reported receiving when completing tasks for that employer.

However, while an ever-increasing number of workers are using these tools for transparency (Kaplan et al. 2018), only a fraction of workers' earnings are well above the minimum wage (Hara et al. 2018). Additionally, despite the tools, crowd workers are still stuck without clear ways to develop themselves and grow within the marketplace. Perhaps part of the problem is that utilizing transparency tools to grow professionally is not straightforward? Each tool displays several different metrics that provide workers with transparency information about different aspects of the market (e.g. how much is a given employer expected to pay? How often does the employer reject workers' labour?). This leaves it unclear which metrics a worker could use to help them to build a career pathway within a crowd market. This complexity has likely led most workers to employ transparency tools ineffectively (Kaplan et al. 2018; Saito et al. 2019b).

For workers alone, it can be hard to learn how to navigate crowd markets to ensure fair wages and professional skill development. The design criteria that we present build on the prior work described above to present computational mechanisms and tools that create 'on-demand leagues' that inform workers on how to effectively and collectively use transparency information to grow professionally in crowd markets. The leagues focus on helping workers to use transparency information effectively through directed social conversations in order to construct career pathways on crowd markets. We showcase how we can use these types of design criteria to: (1) increase workers' wages (Kasunic et al. 2019; Savage et al. 2020); (2) enable workers' skill development (Chiang 2018b; Toxtli and Savage 2020); and (3) drive justice in employers' evaluations on workers (Toxtli et al. 2020). Unlike prior work that

share any reputation information about the employers. The plug-in helps workers to now be able to access such information even without the official support of Amazon.

concentrated on providing more transparency for crowd workers (Huang and Fu 2013), the design criteria that we present here open a new area of research focused on computationally orchestrating workers to actively drive positive change in their professional lives through the use of transparency.

Research Framework: Computational Worker Leagues

In this chapter, we present our overarching research framework for studying crowd workers: 'Computational Worker Leagues'. Our Computational Worker Leagues are sets of tools that allow crowd workers to collaborate with other workers in an on-demand manner, to address and pursue any professional goals they set forward. These tools also offer the additional benefit of collecting quantitative log information about the conditions of the crowd market and present researchers with detailed information about workers' current markets and the challenges they face. Through this quantitative log data, researchers are empowered to design improved tools for crowd workers. Prior work had to infer the conditions from survey studies or interviews. Our framework, in contrast, offers ways for researchers to be able to study crowd markets 'in the wild' from a quantitative perspective.

Our Computational Worker Leagues tool uses a crowdsourcing technique called 'data brokers'. Our approach draws on the assistance of crowd workers who have become efficient in interpreting the information embedded in crowd markets to pursue professional goals. For instance, they may have become very effective at identifying what information to use to earn higher wages. We recruit workers ('data brokers') that have been able to achieve particular professional goals and enable them to share advice with other workers. Our approach also incorporates techniques from machine learning to learn the type of advice from the data brokers that is the most effective for enabling workers to achieve their desired goals. The result is that workers can define a specific professional goal (e.g. raising one's wages or developing skills) and locate concrete guidelines on how to navigate the crowd market to reach this goal.

We offer the data brokers different incentives for participating. For example, workers can be paid to provide the advice. Workers use our tool to provide the advice and then, at the end of the day, our tool measures how much advice a worker provided and pays the worker accordingly directly into the worker's bank account (just as if they were doing any other job on the platform). Another incentive we use is offering workers new career opportunities by becoming brokers. You can imagine that workers who provide the advice are in a way acting as managers for new workers. Workers who want to earn experience of becoming managers can participate and use our tool to gain that much needed expertise of guiding others to succeed. Workers acting as data brokers also do this based on their intrinsic motivation, with the purpose of helping their fellow workers have a better experience on the marketplace and to improve labour conditions for everyone. Our Computational Worker Leagues are composed of a group of data brokers who support workers to go after their goals with the help of the collective of workers. Figure 9.1 presents an overview of our Computational Worker Leagues.

<COMP: INSERT Figure 9.1 NEAR HERE>

Design principles of the Computational Worker Leagues

While there are many ways in which we could computationally organize workers to create these leagues, we focus on collective help while on the job. In our design, we took into account that it was critical to reduce the amount of time that crowd workers spent outside crowd markets, as this was a time when they would not be receiving wages. Consequently, we utilized a web plug-in that would enable them to continue working on the crowd market and earning money at the same time as participating in the league. Web plug-ins are pieces of software that act as an add-on to a web browser and gives the browser additional functionality. Plug-ins can allow a web browser to display additional content that it was not originally designed to display. Most crowd markets allow for web plug-ins. This means we do not have to depend on crowd market owners (usually large technology companies, such as Amazon) for support. Through the plug-ins, we can add the functionality we desire to any crowd market. By being directly embedded where workers are working, it makes it easier for workers to participate in providing and following advice. Web plug-ins allow us to add three main functionalities into crowd markets: (1) an interface through which workers can act as 'data brokers' and provide advice on how to pursue a wide range of professional goals; (2) a means of directly displaying on the crowd market the advice that the data brokers are providing; (3) mechanisms through which researchers can collect information about the crowd market to conduct quantitative analysis.

Additionally, as stated above, it was important for us to design solutions that would allow for the data brokers to operate in a manner that would not distract them from their main job. Our design is based on ideas from 'Micro-Volunteering"' Savage et al. 2016), where people do micro-tasks as a side activity that does not disturb their main task. For this purpose, we frame the design of our intervention around: (1) *availability*: workers should be able to engage in collectively helping each other with a click; (2) *low cognitive load*: workers should be able to collectively help each other without the task being a distraction from the main work; (3) *paid training*: given the economically harsh labour conditions that crowd workers face, our design focuses on enabling workers to receive advice from the data brokers while they are earning money.

To enable these points, our data brokers utilize three components:

Peer Help Collector. The collector lives as a plug-in that connects with the given crowd market in which the worker is operating. In contrast to prior work where workers have to provide lengthy assistance to others (Doroudi et al. 2016), we focus on asking workers to provide micro-assistance. Our data brokers' interface has a small 'provide tip' button. Upon clicking the button, workers see a small pop-up window where they can provide their micro-advice that will help other workers navigate a crowd market to reach a particular goal. We allow workers to input advice anywhere they are within the crowd market. We store the context so as to display the information to other workers at the same point.

This setup enables our design principle of 'availability'. To limit the cognitive load, we limited the length of the micro-assistance that workers gave to each other to 100 characters (length established through trial and error). In the pop-up window, workers just have to select the type of goals for which their micro-advice is relevant (other workers first input the different professional goals they have and for which they would appreciate having assistance from others who have been able to work out how to achieve this goal). Workers select the goal and then type their advice. This allows us to match the advice to the particular goals for which advice is relevant in a simple and direct manner. Figure 9.2 presents how workers can provide advice to others to help them achieve certain goals.

<COMP: INSERT Figure 9.2 NEAR HERE>

Intelligent Selector. For each of the different tasks that workers have on crowd markets, the Peer Help Collector returns a long list of micro-advice. However, not all of this advice will necessarily be helpful for workers to achieve their particular goal. To overcome this issue, we have an Intelligent Selector that focuses on learning what type of advice is best for reaching a given professional goal.

We use a reinforcement learning algorithm which focuses on maximizing the number of workers who consider that the micro-advice that is presented to them is useful. For this purpose, we first ask workers to micro-assess a particular piece of micro-advice for a given goal via upvotes or downvotes. These assessments are fed into our reinforcement learning algorithm that aims to maximize the number of upvotes it obtains from workers. Through this process, our data brokers start to learn the most suitable micro-advice.

Collective Help Display. This component focuses on presenting the micro-advice. The Collective Help Display presents workers with four different examples of micro-advice that the reinforcement learning algorithm ranked highest on the list. If workers want to read more advice, they can click the left or right button to view more. To ensure that new advice has a chance of being evaluated, our tool intermixes new advice that needs micro-assessments into the list of high-ranking advice.

Quantitative data collection

Given that our Computational Worker Leagues offer ways in which crowd workers can pursue goals, we had to establish mechanisms through which we could measure and study whether crowd workers were indeed able to reach their goals. For instance, if the goal was to earn higher wages, we had to develop mechanisms that would show whether we were actually able to increase workers' wages. Specifically, we needed methods for: (1) collecting and quantifying workers' behaviours and the labour they performed on the crowd market (this is necessary to start identifying how much the data brokers might help workers reach their desired goals); (2) flagging when workers followed the strategies from the data brokers; (3) measuring how much workers' behaviour and outcomes changed when following the strategies from the data broker. For this purpose, we developed tools that allowed us to collect information about crowd workers' behaviour, as well as information about the crowd market itself. The tool we built collected:

- Labour information, such as the title of the task posted on the crowd market, how much the task paid, timestamps (when a worker accepted to do the labour/when a worker submitted the labour/when a worker returned a task), employer IDs, and IDs of tasks.

- Worker information, such as daily earnings, tools they use to help their labour, approval rate (how much of their labour gets approved by employers), and worker IDs.
- Employer reputation information: this is part of the transparency information we collected from previous tools that focused on bringing more transparency to the crowd market. Such information can come from tools such as Turkopticon.

We use this information to study the effectiveness of our leagues. For instance, we have used our tools to identify the hourly wage of workers on Amazon Mechanical Turk (Saito et al. 2019b), as well as average wages (Hara et al. 2018). In this way, we offer researchers a more in-depth perspective on the marketplace and allow them to measure the quality of labour conditions for workers. Knowing the exact conditions facilitates taking action to be able to change and improve (Whiting et al. 2019). We believe it is especially valuable to connect these types of quantitative tools with interviews and surveys to have a much richer understanding of the reality of crowd markets. For readers interested in learning more about our tools, from the computational side, please refer to our research papers, Hara et al. (2018); Saito et al. (2019b).

Application of Our Research Framework

In this section, we present an overview of how our Framework has been used to help crowd workers attain two important professional goals: earning higher wages and developing their skills. We present how we deployed our tools in the wild and offer insights about how researchers could benefit from the quantitative data we collected.

Computational Worker Leagues for increasing wages

We conducted a field experiment to investigate how the hourly wage of workers changed when using our Computational Worker Leagues. It was not a simple task. Most crowd markets do not provide any information about the hourly wage for a particular task nor how much time it would take workers to complete the actual labour. It is, therefore, not straightforward to calculate the change in workers' wages over time (Hara et al. 2018). To overcome this challenge, we utilized the quantitative tools within our Computational Worker Leagues. These tools calculate how much time workers spend on each task and estimate each worker's hourly wage per task based on this. We are thus able to estimate how workers' wages varied over time.

Equipped with our Computational Worker Leagues and the associated tools for quantifying worker behaviour, we ran a two-week field experiment. We had real-world novice workers from Amazon Mechanical Turk complete over 25,000 tasks posted from 1,394 employers, with the experimental group of workers utilizing our Computational Worker Leagues and the control group operating as normal without being exposed to our Worker Leagues. Details about how we recruited workers for the study and how much we paid them for their participation can be found in our research papers (Saito et al. 2019a; Savage and Jarrahi 2020; Savage et al. 2020; Hanrahan et al. 2021). Our study concluded that having workers utilize the Computational Worker Leagues empowered them to increase their income.

Computational Worker Leagues for skill development

In our second application of our Computational Worker Leagues, we hypothesized that receiving advice from data brokers can help workers to build their skills in particular areas. Similar to Doroudi et al. (2016), we measured skills growth in terms of an increase in workers' speed and labour quality. To test this hypothesis and to understand the type of work that is well or poorly supported by the Computational Worker Leagues in the wild, we conducted: (1) a controlled field experiment; and a (2) real-world deployment of our design.

Controlled field experiment

The goal of our field experiment was to compare our Computational Worker Leagues with other approaches to evaluate the effectiveness of our proposed framework in helping workers develop their own skills. We considered three conditions: (1) workers do tasks without receiving any type of advice from the data brokers (control condition); (2) workers do tasks while receiving random advice from data brokers (random condition); (3) workers do tasks while receiving advice from data brokers that our machine learning models have identified are the best for helping workers to develop their skills (Computational Worker Leagues condition).

Given that we needed to measure participants' work quality, we focused on skill development for labour that was not open-ended and whose quality we could more easily measure. We focused on audio transcription tasks whose quality can be directly measured by transcription accuracy. It is important to note that audio transcription tasks are not only a popular task on crowd markets (Difallah et al. 2015) but, in addition, becoming proficient at audio transcription can substantially increase a person's wages. Transcribers typically earn US$0.01–0.02 per sentence (Novotney and Callison-Burch 2010), which could potentially translate to high wages if a worker is fast (and accurate) enough. Specializing in audio transcription can allow crowd workers to command higher wages as audio transcription is in high demand. Written records of court proceedings and captions for live television events, such as the news, sports, and political speeches, all require real-time audio transcription. Audio transcription skills are thereby highly specialized, highly valued, and well paid, earning up to $300 per hour outside MTurk. Building audio transcription skills on MTurk could thereby help crowd workers expand their horizons and increase their earnings.

We considered that novices were the ones who could benefit the most from our tool as it can be difficult to learn the ropes of a crowd market while also developing new skills. Novice workers were defined as both workers who were new to Amazon Mechanical Turk (i.e. to a particular crowd market) and those inexperienced in audio transcription tasks. Our field experiment therefore focused on investigating whether our design helps novice crowd workers improve their audio transcription skills.

We studied novices' completion time and work quality for three different audio transcription tasks under one of our three study conditions. We recruited a total of 90 novice MTurk workers and randomly divided them into three experimental conditions (30 in each condition). Participants in each condition were assigned the same audio transcription tasks with the same order. Tasks were sourced from real-world audio transcription tasks on MTurk and had similar difficulty: participants had to transcribe around 28 seconds of audio with similar levels of background noise, and with an average speaking rate of 165 words-per-second. We designed a tool that recorded workers' retention rate, completion time, and accuracy for each task. To calculate the time to complete a task, we measured the time when a worker first accessed the task as the start time and the time when workers submitted their labour (transcription) as their finish time. To study accuracy, we calculated the word error rate (WER) produced by each worker for each transcription, a commonly used metric to assess performance in audio transcription (Bigham et al. 2017). During the study period, novice workers completed a total of 253 tasks across all three conditions. Overall, our field experiment indicated that workers exposed to the Computational Worker Leagues were faster without sacrificing accuracy than workers without the advice from our data brokers. See our research paper Chiang et al. (2018b) for more details on this study.

Real-world deployment

We launched our Computational Worker Leagues for skill development and studied their use by real-world workers on Amazon Mechanical Turk. Our tool was installed by 179 workers; 86 per cent of these were active users of the system (i.e. they either became data brokers or used the advice from the data brokers to develop their skills); the rest used our system more passively (i.e. they just installed our tool). A total of 96 workers provided 363 snippets of advice while they acted as data brokers, and 146 workers provided 1401 micro-assessments of the advice from the data brokers.

From our real-world deployment, we found that workers who decided to become data brokers tended to create different types of advice each time. This is important, as it was unclear whether there would be an infinite set of advice that the data brokers could provide or whether there might be a vast but bounded set of advice that could be constantly shared to improve the experiences of workers. Our study highlights that a flow of advice can be constantly arriving from the data brokers. Crowd work is continuously evolving (Hara et al. 2018). We, therefore, believe that data brokers will likely always have new advice to provide. The popular worker tool of Turkopticon was published in 2013, and it is still active with new reviews of requesters (Irani and Silberman 2016). Therefore, we do see our data brokers being used long term. Through our real-world deployment, we identified that workers used our tool for all of the different types of tasks that are available on Amazon Mechanical Turk. This thus also highlights the viability of our tool. Our approach was able to get workers to act as data brokers for all of the different types of labour available on the platform. This is promising as there were no tasks where workers did not feel they could not help other workers become more efficient and develop their skills.

Broader Challenges

The research methods we present here focused on evaluating how our Computational Worker Leagues help workers to achieve the professional goals they set forward. This entails conducting real-world experiments. However, this is not simple given the variability and randomness that can exist around tasks that are available on the crowd market—it might be that in any one week some tasks that could help a worker achieve certain goals are not available and hence it is not so much that the Computational Worker Leagues were not effective, but rather what was available on the crowd market did not facilitate the achievement of the goal (Hara et al. 2018).

Another potential problem is that to recruit participants for our studies, we usually post tasks on crowd markets and use that platform for recruitment. But this means that we only reach workers who are willing to engage with our tasks in the first place. Future work could explore other ways of recruiting workers and eliciting information from them (e.g. via video recordings or interviews). Such studies could explore how using different mechanisms for eliciting information from workers shapes the type of information that is obtained. In other words, we believe there is significant value in exploring different setups of our data brokers. Additionally, we presented our systems primarily within the context of Amazon Mechanical Turk. Future work could also explore how our data brokers might operate in other crowd platforms (e.g. Uber, Upwork, or Citizen Science platforms). Nonetheless, given that our goal was to start to understand how our computational methods played out in the wild, we consider our approach to be appropriate and representative.

We also believe there are technical and educational challenges that need to be addressed. In particular, we think there is a gap between qualitative and quantitative researchers who are investigating crowd work. Qualitative researchers might not feel as comfortable deploying our Computational Worker Leagues to conduct studies, perhaps because significant technical knowledge is required to do such investigations in the wild. For instance, researchers need to have knowledge about databases to instal a database that can collect the worker data to conduct the quantitative data analysis; researchers also need to know some javascript in order to configure the plug-in and connect the plug-in to their database. Additionally, researchers need to have some data science skills in order to take the data collected from the plug-in and start to find patterns. However, despite these challenges, we have seen that qualitative researchers can make use of the quantitative data our studies collect (notice that this step involves simply having more data science skills). Our hope is that through the data collection that our systems offer, we can encourage and enable more qualitative researchers to study other aspects of crowd work that they might not have had access to in the past.

Ethical challenges

Our research approach involves collecting quantitative data about workers and the crowd market (including information about employers). While the data are used to benefit workers, ethical questions can emerge about close monitoring of workers even if it is for their benefit. Our research has always anonymized worker data, as well as conducting group analysis instead of studying individual behaviour. However, there are questions to ask about the best strategy for managing the data that are collected. Should it be data that are owned by the companies or universities which run the studies with our Computational Worker Leagues? Or should they always be data that are owned by a collective of workers? We believe that the best option is to develop approaches that involve all stakeholders, and especially include populations that are typically ignored and not given power over their data (e.g. workers). There is also value in exploring approaches that have been utilized in opensource collaboration projects to enable all stakeholders of the crowd market to use the anonymized data for the different goals they might have. For instance, employers might be interested in utilizing the data to improve the quality of the work that they obtain on the crowd market;

however, workers might benefit from using the data to identify the best strategies for developing their skills; while researchers, on the other hand, might want to use the data to better understand the crowd market.

However, part of the problem is that state of the art tools that collect crowd market information are limited and tend to be focused on particular tasks. For instance, Turkopticon (as above) focuses primarily on collecting data about requesters' ratings (Irani and Silberman 2013). To address these challenges, we propose 'The Opensource Storehouse For Multiple Stakeholders'. The storehouse would function in the same way as a traditional repository that, for a given crowd market, collects different types of information related to the market. However, the storehouse would also request that metadata are uploaded that can help the different stakeholders to achieve their desired goals, for example:

- *Stakeholder and goals*: information about the goal for which the data were collected, and the stakeholder who was interested in the goal;
- *Data collected*: This relates to all the crowd market information that is collected, such as worker characteristics, types of tasks, employer information, etc.
- *General crowd market characteristics*: There is value in understanding the nature of the crowd market in which the data were collected. Was it location-based crowd work? Was it volunteer labour? Paid labour? In what national context(s) was the work carried out?
- *Feedback*: A space for the different stakeholders to provide input on the data that were collected for a particular crowd market and a particular goal.
- *List of to-dos:* List of things that can be done to enhance the data collection (e.g. perhaps data from more populations are needed).

Workers were paid to participate in using our Computational Worker Leagues. We believe there is value in considering setups where workers are paid to participate in research, especially given the harsh labour conditions they face. We also believe there is value in identifying the best setups to make tools, such as our Computational Worker Leagues, into something that is sustainable long term. Research has identified that the emergence of private tools to help crowd workers is creating further social divisions among workers (Williams et al. 2019). It is, therefore, important to find ways in which these types of tools could be accessible for all. Turkopticon has recently turned from a space that was primarily run by academics, into a space run by workers.[2] To continue its operation, Turkopticon has become a type of NGO that can receive funding from different parties. A similar setup could be explored with our tools. However, a committee that analyses from which parties it is acceptable to receive funding would also be required, in particular to ensure the tools remain appropriate for all of the different stakeholders.

Conclusion

Crowd markets offer a wide range of readily available labour (Alkhatib et al. 2017). Unfortunately, it is often difficult for workers to know how best to navigate these crowd markets in order to find labour that pays well and might actually be useful for workers' career growth. A number of tools have emerged to help workers better navigate crowd markets by bringing transparency (Irani and Silberman 2013). We argue that transparency is not enough.

2 Turkopticon was created by academics Lilly Irani and Six Silberman while they were both PhD students at the University of California, Irvine. However, they decided to turn Turkopticon into a worker-owned NGO to provide workers with more agency in how they wanted the tool and future tools to evolve and the type of governance associated. You can read more about this decision here: https://blog.turkopticon.info/?page_id=474

We need to provide workers with effective tools for using transparency for the different professional goals that workers might have. In this chapter, we presented a brief overview of how these tools can be used to help workers aim for different professional goals, in particular developing their skills and increasing their wages. We began with an overview of the type of labour markets that are feeding our artificial intelligence industry. We explained the types of problems that workers in these platforms face and discussed how we might design tools to empower workers to address these problems and change their labour conditions. We also presented ways other researchers can use this approach to not only create interventions in crowd markets but also conduct quantitative analysis of what is happening inside such marketplaces.

Acknowledgments. Special thanks to all the anonymous reviewers who helped us to strengthen the paper. This work was partially supported by NSF grant FW-HTF-19541.

References

Abraham, K. G., Haltiwanger, J. C., Sandusky, K., and Spletzer, J. R. 2018. Measuring the gig economy: Current knowledge and open issues, Technical report, National Bureau of Economic Research.

Alkhatib, A., Bernstein, M. S., and Levi, M. 2017. Examining crowd work and gig work through the historical lens of piecework. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (pp. 4599–616), ACM.

Allahbakhsh, M., Benatallah, B., Ignjatovic, A., Motahari-Nezhad, H. R., Bertino, E., and Dustdar, S. 2013. Quality control in crowdsourcing systems: Issues and directions, *IEEE Internet Computing*, 17(2): pp. 76–81.

Amazon. 2020. Amazon Mechanical Turk, Amazon, https://www.mturk.com/ (accessed 28 September2020).

Berg, J. 2015. Income security in the on-demand economy: Findings and policy lessons from a survey of crowdworkers. *Comparative Labour Law and Policy Journal*, 37 (p. 543).

Berg, J., Furrer, M., Harmon, E., Rani, U., and Silberman, M. S. 2018. Digital labour platforms and the future of work, Towards Decent Work in the Online World. *Rapport de l'OIT,* Vol. 1

Bernstein, M. S., Karger, D. R., Miller, R. C., and Brandt, J. 2012. Analytic methods for optimizing realtime crowdsourcing. *arXiv*: 1204.2995.

Bigham, J. P., Kushalnagar, R., Huang, T.-H. K., Flores, J. P., and Savage, S. 2017. On how deaf people might use speech to control devices. In *Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility* (pp. 383–84), ACM.

Chiang, C.-W., Betanzos, E., and Savage, S. 2018a. Exploring blockchain for trustful collaborations between immigrants and governments. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems* (p. LBW 531), ACM.

Chiang, C.-W., Kasunic, A., and Savage, S. 2018b. Crowd coach: Peer coaching for crowd workers' skill growth. In *Proceedings of the ACM on Human-Computer Interaction* 2(CSCW), 37.

ChrisTurk. 2018. TurkerView Worker Review Platform for Amazon Mechanical Turk, Turkerview, https://turkerview.com/ (accessed 28 September 2020).

Deng, X. N. and Joshi, K. 2013. Is crowdsourcing a source of worker empowerment or exploitation? Understanding crowd workers' perceptions of crowdsourcing career. In International Conference on Information Systems 2013. Association for Information Systems.

Difallah, D. E., Catasta, M., Demartini, G., Ipeirotis, P. G., and Cudre-mauroux, P. 2015. The dynamics of micro-task crowdsourcing: The case of Amazon MTurk. In *Proceedings of the 24th International Conference on World Wide Web* (pp. 238–47). International World Wide Web Conferences Steering Committee.

Doroudi, S., Kamar, E., Brunskill, E., and Horvitz, E. 2016. Toward a learning science for complex crowdsourcing tasks. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (pp. 2623–34). ACM.

Eickhoff, C., and De Vries, A. 2011. How crowdsourcable is your task. In *Proceedings of the workshop on crowdsourcing for search and data mining (CSDM)* (pp. 11–14) at the fourth ACM international conference on web search and data mining (WSDM).

Gadiraju, U., and Demartini, G. 2019. Understanding worker moods and reactions to rejection in crowdsourcing. In *Proceedings of the 30th ACM Conference on Hypertext and Social Media*, (pp. 211–20). ACM.

Gaikwad, S., Morina, D., Nistala, R., Agarwal, M., Cossette, A., Bhanu, R., Savage, S., Narwal, V., Rajpal, K., Regino, J., and Mithal, A. 2015, November. Daemo: A self-governed crowdsourcing marketplace. In *Adjunct proceedings of the 28th annual ACM symposium on user interface software & technology* (pp. 101–2). ACM.

Gray, M., and Suri, S. 2019. *Ghost work: How to stop Silicon Valley from building a new global underclass.* Eamon Dolan Books.

Hall, D., and Perona, P. 2015. Fine-grained classification of pedestrians in video: Benchmark and state of the art. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 5482–91). IEEE Computer Society.

Hanrahan, B. V., Ma, N. F., Betanzos, E., and Savage, S. 2020. Reciprocal research: Providing value in design research from the outset in the rural United States. In

Proceedings of the 2020 International Conference on Information and Communication Technologies and Development (pp. 1–5). ACM.

Hanrahan, B. V., Chen, A., Ma, J., Ma, N. F., Squicciarini, A., & Savage, S. (2021). The Expertise Involved in Deciding which HITs are Worth Doing on Amazon Mechanical Turk. *Proceedings of the ACM on Human-Computer Interaction*, *5*(CSCW1), 1-23.

Hara, K., Adams, A., Milland, K., Savage, S., Callison-Burch, C., and Bigham, J. P. 2018. A data-driven analysis of workers' earnings on Amazon Mechanical Turk. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, (p. 449). ACM.

Hara, K., Adams, A., Milland, K., Savage, S., Hanrahan, B. V., Bigham, J. P., and CallisonBurch, C. 2019. Worker demographics and earnings on Amazon Mechanical Turk: An exploratory analysis. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, (pp. 1–6). ACM.

Hilton, L. G., and Azzam, T. 2019. Crowdsourcing qualitative thematic analysis. *American Journal of Evaluation*, 40(4): pp. 575–89.

Huang, S.-W., and Fu, W.-T. 2013. Don't hide in the crowd! Increasing social transparency between peer workers improves crowdsourcing outcomes. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, (pp. 621–30). ACM.

Irani, L. C., and Silberman, M. 2013. Turkopticon: Interrupting worker invisibility in Amazon Mechanical Turk. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 611–20). ACM.

Irani, L. C., and Silberman, M. 2016. Stories we tell about labour: Turkopticon and the trouble with design. In *Proceedings of the 2016 CHI conference on human factors in computing systems* (pp. 4573–86). ACM.

Jagabathula, S., Subramanian, L., and Venkataraman, A. 2014. Reputation-based worker filtering in crowdsourcing. In *Advances in Neural Information Processing Systems*: Proceedings of the 2014 Conference (Vol. 27), (pp. 2492–500). MIT Press.

Jarrahi, M. H., and Sutherland, W. 2019. Algorithmic management and algorithmic competencies: Understanding and appropriating algorithms in gig work. In *International Conference on Information*, (pp. 578–89). Springer.

Kaplan, T., Saito, S., Hara, K., and Bigham, J. P. 2018. Striving to earn more: A survey of work strategies and tool use among crowd workers.. In Proceedings of the AAAI Conference on Human Computation and Crowdsourcing (Vol. 6, No. 1, pp. 70-78). PKP.

Kasunic, A., Chiang, C.-W., Kaufman, G., and Savage, S. 2019. Crowd work on a CV? understanding how AMT fits into turkers' career goals and professional profiles, arXiv:1902.05361.

Kuhn, K. M., and Maleki, A. 2017. Micro-entrepreneurs, dependent contractors, and instaserfs: Understanding online labour platform workforces, *Academy of Management Perspectives*, 31(3): pp. 183–200.

Lampinen, A., and Cheshire, C. 2016. Hosting via Airbnb: Motivations and financial assurances in monetized network hospitality. In *Proceedings of the 2016 CHI conference on human factors in computing systems* (pp. 1669–80). ACM.

Li, T., Luther, K., and North, C. 2018. Crowdia: Solving mysteries with crowdsourced sensemaking. In *Proceedings of the ACM on Human-Computer Interaction 2(CSCW)*, (pp. 1–29). ACM.

Metall, I., 2016. *Frankfurt paper on platform-based work—proposals for platform operators, clients, policy makers, workers, and worker organizations*, IG Metall, Frankfurt.

Novotney, S. and Callison-Burch, C., 2010. Cheap, fast and good enough: Automatic speech recognition with non-expert transcription. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics* (pp. 207–15). Association for Computational Linguistics.

Radford, J., Pilny, A., Reichelmann, A., Keegan, B., Foucault Welles, B., Hoye, J., Ognyanova, K., Meleis, W., and Lazer, D. 2016. Volunteer science as an online platform for experiments in organizations. In *Academy of Management Proceedings*, Vol. 2016 (p. 17008). Academy of Management.

Saito, S., Chiang, C.-W., Savage, S., Nakano, T., Kobayashi, T., and Bigham, J. 2019a. Predicting the working time of microtasks based on workers' perception of prediction errors. *Human Computation,* 6(1): pp. 192–219.

Saito, S., Chiang, C.-W., Savage, S., Nakano, T., Kobayashi, T. and Bigham, J. 2019b. Turkscanner: Predicting the hourly wage of microtasks. arXiv:1903.07032.

Savage, S., and Jarrahi, M. H. 2020. Solidarity and AI for transitioning to crowd work during covid-19. In *The New Future of Work Symposium 2020*. Microsoft.

Savage, S., Monroy-Hernandez, A., and Höllerer, T. 2016. Botivist: Calling volunteers to action using online bots. In *Proceedings of the 19th ACM Conference on Computer Supported Cooperative Work & Social Computing* (pp. 813–822). ACM.

Savage, S., Chiang, C., Susumu, S., Toxtli, C., and Bigham, J. 2020. Becoming the super turker: Increasing wages via a strategy from high earning workers. In *Proceedings of the 29th International Conference on World Wide Web, WWW '20,* International World Wide Web Conferences Steering Committee.

Singer, Y., and Mittal, M., 2013. Pricing mechanisms for crowdsourcing markets. In *Proceedings of the 22nd International Conference on World Wide Web* (pp. 1157–66). ACM.

Slivkins, A., and Vaughan, J. W. 2014. Online decision making in crowdsourcing markets: Theoretical challenges, *ACM SIGecom Exchanges*, 12(2): pp. 4–23.

Smith, A. 2016. Gig work, online selling and home sharing, *Pew Research Center* 17..

Smith, A. 2019. Gig work, online selling, and home sharing. *Pew Research Center* 17 November 2016.

Sorokin, A., and Forsyth, D. 2008. Utility data annotation with Amazon Mechanical Turk. In *2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops* (pp. 1–8). IEEE.

Strathern, M. 2000. The tyranny of transparency, *British Educational Research Journal*, 26(3), pp. 309–21.

Toxtli, C., and Savage, S. 2020. Enabling expert critique at scale with chatbots and micro-guidance, 13th International Conference on Advances in Computer-Human Interactions, ACHI 2020. International Academy, Research, and Industry Association IARIA.

Toxtli, C., Hara, K., Callison-Burch, C., and Kristy, M. 2019. Understanding the crowd markets that workers and requesters imagine. In *Proceedings of the 7th ACM Collective Intelligence (CI) Conference*. ACM.

Toxtli, C., Richmond-Fuller, A., and Savage, S. 2020. Reputation agent: Prompting fair reviews in gig markets. In *Proceedings of the 29th International Conference on World Wide Web, WWW '20*, International World Wide Web Conferences Steering Committee.

Toxtli, C., Suri S., & Savage, S. (2021). Quantifying the Invisible Labor of Crowd Work. *Proceedings of the ACM on Human-Computer Interaction*, *5* (CSCW1), 1-23.

Whiting, M. E., Hugh, G., and Bernstein, M. S. 2019. Fair work: Crowd work minimum wage with one line of code. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, Vol. 7, (pp. 197–206). PKP.

Williams, A. C., Mark, G., Milland, K., Lank, E., and Law, E. 2019. The perpetual work life of crowdworkers: How tooling practices increase fragmentation in crowdwork. In *Proceedings of the ACM on Human-Computer Interaction 3(CSCW)*, (pp. 1–28). ACM.

Yeoman, R. 2014. Conceptualising meaningful work as a fundamental human need, *Journal of Business Ethics*, 125(2), pp. 235–51.

Young, J., Swamy, P., and Danks, D. 2018. *Beyond AI: Responses to hate speech and disinformation*. Unpublished paper presented at Carnegie Mellon University. Retrieved from http://jessica-young. com/research/Beyond-AI-Responses-to-Hate-Speech-and-Disinformation. pdf.