RESEARCH ARTICLE

The Dark Side of Recruitment in Crowdsourcing: Ethics and Transparency in Micro-Task Marketplaces

Haoyu Xie [1,*], Eddy Maddalena[2], Rehab Qarout[3] & Alessandro Checco[4]

[1] *The University of Sheffield, Sheffield, UK (E-mail: hxie5@sheffield.ac.uk);*
[2] *University of Udine, Udine, Italy (E-mail: eddy.maddalena@uniud.it);*
[3] *University of Jeddah, Jeddah, Saudi Arabia (E-mail: rkqarout@uj.edu.sa);*
[4] *University of Rome La Sapienza, Roma, Italy (E-mail: alessandro.checco@uniroma1.it)*

Abstract. Micro-task crowdsourcing marketplaces like Figure Eight (F8) connect a large pool of workers to employers through a single online platform, by aggregating multiple crowdsourcing platforms (channels) under a unique system. This paper investigates the F8 channels' demographic distribution and reward schemes by analysing more than 53k crowdsourcing tasks over four years, collecting survey data and scraping marketplace metadata. We reveal an heterogeneous per-channel demographic distribution, and an opaque channel commission scheme, that varies over time and is not communicated to the employer when launching a task: workers often will receive a smaller payment than expected by the employer. In addition, the impact of channel commission schemes on the relationship between requesters and crowdworkers is explored. These observations uncover important issues on ethics, reliability and transparency of crowdsourced experiment when using this kind of marketplaces, especially for academic research.

Key Words: Micro-task crowdsourcing, Human computation

1. Introduction

Micro-task crowdsourcing platforms, e. g., Amazon Mechanical Turk (MTurk)[1] or Figure Eight (F8) (formerly known as CrowdFlower and currently as Appen[2]), enable researchers to reach large pools of participants, also called contributors or workers, for their experiments (Howe, 2006). The main advantage of using crowdsourcing in research is the access to low-cost digital labour from the platform's large pool of available crowdworkers (Heer and Bostock, 2010). Researchers, called requesters in the crowdsourcing context, design their experiments to be

---

[1] https://www.mturk.com
[2] https://appen.com

performed as a batch of atomic micro-tasks. These tasks are commonly called HITs (Human Intelligence Tasks), since they require human intelligence to be performed effectively. Such HITs include market survey or image annotation for training AI algorithmic models (Gadiraju et al., 2014). In essence, a HIT consists of a web page, typically a form, requiring workers to input specific information or perform actions. Based on responses to a survey of 1000 workers on CrowdFlower, Gadiraju et al. (2014) proposed a categorization scheme for the most popular HITs in the platform, having identified task categories including information finding, verification and validation, interpretation and analysis, content creation, surveys, and content access.

The micro-task crowdsourcing process takes place as follows: First, a researcher designs a HIT and deploys it into a crowdsourcing platform specifying the parameters for its execution, such as the number of workers required and the relative payment. Then, the platform allocates the batch of work to several workers according to specific policies. Finally, the platform collects and aggregates the results sending them back to the requester. In addition to the workers' payments, the requester also pays a service fee corresponding typically to 20-25% of the experiment cost to the platform.

Crowdsourcing lends itself to being used successfully in different areas, such as psychology (Gosling et al., 2015), social science (Auer et al., 2021), economics (Jacques and Kristensson, 2019), cognitive science (Stewart et al., 2017) and medical science (Petrović et al., 2020). More recently, crowdsourcing attracted the interest of several AI communities. Vougiouklis et al. (2020) conducted a series of crowdsourcing experiments to evaluate the quality of texts generated by machines. Finin et al. (2010) designed crowdsourcing tasks to collect named entity annotations for Twitter. Deng et al. (2016) improved an image recognition model with crowdsourcing via an online game. In Wan et al. (2019)'s study, traffic related data were collected via crowdsourcing to train vehicle route planning algorithms. Another notable example is Chimera, a large-scale classification model developed with the help of crowdsourcing to classify tens of millions of products at WalmartLabs (Sun et al., 2014).

1.1. Pricing of Crowdsourcing Tasks

While designing a task, requesters are called to make several choices. One crucial decision concerns the monetary reward to be given to the workers who complete the task successfully. Indeed, while an exiguous payment might not motivate the workers enough (Auer et al., 2021), a conspicuously large one might attract spammers or reduce the scale of limited-budget experiments. Different factors influence the reward policies of a task, including the amount of time required for its completion, its difficulty, or the worker skill set required (workers with special qualifications

have to be paid more). Finally, additional costs, e. g., platform fees, consisting typically of a percentage of task payment, need to be taken into account while deciding upon the reward strategies. In the literature, some efforts can be found to develop tools to assist price negotiation in crowdsourcing. For example, Horton and Zeckhauser (2010) proposed *hagglebot*, a bot to serve as an automated negotiating agent for requesters.

1.2. Ethical Issues in Crowdsourcing

One topic that has attracted the research community's interest concerns the ethical issues about workers' conditions reported over the years. Hara et al. (2018) showed that workers were paid a wage of 4-6 USD per hour on the MTurk platform whereas the US federal minimum wage corresponds to 7.25 USD per hour since 2009. Crowdworkers struggle with extremely low hourly income, and this issue has been used as the motivation for some studies (Whiting et al., 2019; Saito et al., 2019). Moreover, requesters might refuse to pay crowdworkers for a completed job either maliciously or because of badly-designed tasks and error-prone result assessment systems (Martin et al., 2016; Silberman et al., 2010). Lease et al. (2013) and Silberman et al. (2010) investigated crowdworkers' vulnerability to fraudulent tasks targeting their privacy and assets. Platforms often provide very limited information to crowdworkers about the identity of requesters and the quality of their tasks. While the amount of payment for a HIT is communicated to the crowdworker, an estimate of the expected time to complete it is missing. Most platforms also lack an estimate of the rejection rate of the requester who published that HIT batch, as well as general feedback from workers who have already completed HITs from the same batch. It is therefore arduous for workers to evaluate requesters' reputations and to assess the potential hourly wage before starting tasks (Fieseler et al., 2019; Martin et al., 2014). On the contrary, the information available for requesters is more detailed since they can typically access workers' performance history and qualifications (Kingsley et al., 2015).

1.3. Research Standards in Crowdsourcing

The academic community has responded to the aforementioned ethical issues by calling for crowdsourcing research standards to be implemented by universities, journals, and grantmakers. In particular, university ethics boards have been encouraged to design guidelines for the use of crowdsourcing in research, considering the specific platform- and labour-related issues of crowdworkers, like the lack of access to traditional employment protections (Williamson, 2016). Even though ethical guidelines vary by institution and country, the fairness of the payment should be an essential condition that needs to be satisfied (Silberman et al., 2018a). These concerns are significant when using paid micro-task crowdsourcing platforms, where

crowdworkers consider monetary rewards as the driving motivation for their participation (Martin et al., 2017), as opposed to e. g. in citizen science or volunteer-based crowdsourcing, where the driving motivation is intrinsic (Leimeister et al., 2009; Hossain, 2012). In particular, to guarantee transparency the amount of the participant reward should always be clearly specified in the design phase of a research project (Schmidt, 2013). As illustrated in this work, these issues are particularly pronounced in F8, where the compensation scheme seems to be unclear, changing over time, and linked to questionable activities such as gambling.

There have been several solutions in the existing literature that addressed the aforementioned ethical issues of crowdsourcing. Whiting et al. (2019) proposed *Fair Work*, a tool that computes payment to ensure a minimum wage for the workers. Saito et al. (2019) discussed how crowdworkers struggle with extremely low hourly income and proposed *TurkScanner*, a machine learning approach which predicts worker completion time to compute a fair hourly wage. Certainly, making crowdsourcing more ethical is not just about monitoring the fairness of rewards, but also about improving the reputation system. Gaikwad et al. (2016) attempted to encourage greater ethical standards among platform members by developing a reputation system that provides more honest private evaluations for both workers and requesters. In addition, reducing the risk of unethical behaviour is another approach. For instance, Fan et al. (2020) helped crowdworkers reach a more ethical hourly wage via a novel crowdsourcing reward mechanism, so their risks of being underpaid could be shared within a worker group.

1.4. Motivation - Opaqueness of Reward Schemes

The solutions discussed above represent an important first step in raising awareness about the importance of compensating workers fairly and in providing useful tools to achieve this goal. Unfortunately, when it comes to practice these solutions present some limitations. Indeed, when choosing a HIT reward scheme, requesters can only rely on the limited information provided by the platform about the actual reward process. Moreover, some platforms make use of recruitment channels, namely third-party services that work as an intermediary between platforms and workers (IG Metall, 2017). The cost of the recruitment channel service is then included in the worker's cost reported by the platform to the requester. Consequently, the requester cannot have full knowledge about which percentage of the reward will effectively reach the crowdworker. For the reader's convenience, we define in Table 1 the costs that requesters have to cover to run a crowdsourcing job.

By observing F8, we noticed that the commission amount varies over time and across channels. More importantly, such commissions are not explicitly provided when launching a task (as shown in Figure 1), and the exact amount is only revealed after a task is completed, as discussed in Section 6. Requesters might not be aware of the effects of the reduced rate on the workers' actual payments, which can fluctuate over time and potentially undermine the fairness of the compensation the workers

Table 1. Cost breakdown of a crowdsourcing task.

| Type of cost	Description
Task cost	The total cost paid by a requester to a crowdsourcing platform to run a HIT. The platform splits this whole amount into three parts: *platform fee*, *channel commission*, and worker reward.
Platform fee	The amount of money a platform earns from a requester for both the services of intermediation with the workers and task hosting. This cost is computed as a percentage (usually 20-25%) of the task cost. The platform fee is always known by the requester.
Channel commission	The cost of the recruitment channel service. The channel commission is not always disclosed to the requester, and it can vary considerably among channels and over time.
Worker reward	The actual reward the worker receives for completing the task, net of platform fee and channel commission. It might be monetary or in other forms (e. g., an online shop voucher), can vary over time, and can depend on output metrics (e. g., bonuses for output quality).
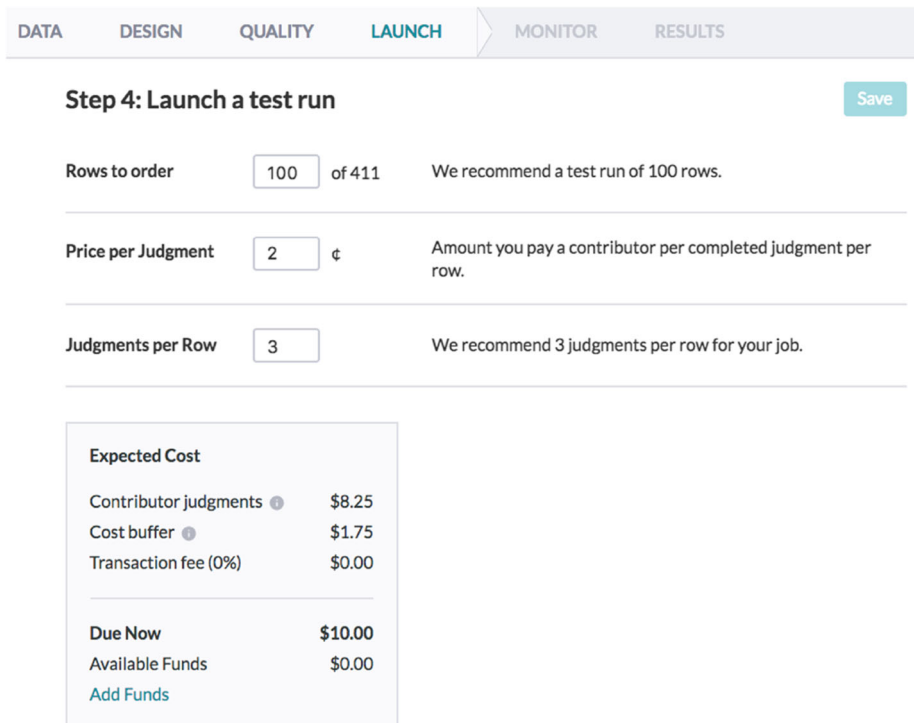
Figure 1. F8 task launch panel. The cost indicated as "contributor judgments" does not consider the channel commissions: the crowdworker might receive a lower reward.

receive. To the best of our knowledge, no study has analysed channel commissions in crowdsourcing platforms.

This paper investigates the distribution and variation over time of channel commissions in F8. Since the amount of such commissions is often not available to the requesters, we ran a survey task asking the workers in F8 how much they would be paid for the completion of the ongoing survey, and compared the payment amount for each channel. This data has been compared with the historical channel information in the marketplace to build a picture of the recruitment fee dynamics. We conducted a comparative analysis of the demographics, channel distribution, and reward scheme of 53065 tasks.

The rest of the paper is structured as follows: Section 2 presents a literature review of the ethical issues around payment in crowdsourcing tasks, as well as on motivation and incentive rewards effectiveness. Research questions are elaborated and explained in Section 3. In Section 4.1, a brief summary of the historical metadata is presented. Moreover, the reward schemes for each of the top five channels are explained in Section 4.2. The demographic information for the workers in the top five channels are discussed in Section 5, including their work experience and task acceptance criteria. The worker reward loss over time for each channel and task payment range is analysed in Section 6. In Section 7, the impact of unethical payment behaviour is explored from the perspective of the worker and the requester respectively in connection with our findings. Section 8 draws the conclusions of this work.

2. Literature Review

In this section, we illustrate related work on the ethical issues concerning rewards in paid micro-tasks. Furthermore, we also review the studies on the impacts of the reward scheme on workers' motivation and quality of outcome.

2.1. Ethical Issues with Crowdsourcing Rewards

The globalisation and cross-specialisation of crowdsourcing expose it to complex ethical judgements, as different countries and domains have their own unique regulations and ethical policies. Specifically, the US and India dominate the crowd labour force (Difallah et al., 2018; Kazai et al., 2013), and the workers from these two countries have different subjective perceptions of the fairness of payments from doing micro-tasks due to their local economic levels, and crowdsourcing practices are governed by different laws (Martin et al., 2016; Gellman, 2015). In addition, crowdsourcing projects from academic institutions are subject to higher ethical standards than those from commercial institutions because they are reviewed by ethics committees and governed by strict ethics guidelines (Shmueli et al., 2021; Gleibs, 2017; Martin et al., 2017).

Using crowdsourcing to collect data in social science studies brings together the benefits of having a broad demographic distribution and fast responses. However, many researchers have raised the issues of improper payments and low rewards, especially in studies where researchers do not have an extensive track-record in the domain of crowdsourcing task design, and particularly for those studies that regard crowdsourcing as a means to collect data and not as the main subject of study. For example, Callison-Burch (2009) claimed to have paid crowdworkers "a grand total of 9.75 USD to complete nearly 1,000 HITs". Andersen and Richard (2018) discussed the pay rate in social science experiments that had been carried out using crowdsourcing. Two experiments with different task lengths were conducted to measure the effect of payment rate on the quality of the workers' output. The findings of this study confirmed that the variance of pay rates did not have a significant effect on workers' output quality, but they did have a different kind of effect, such as the completion time. A similar study by Haug (2018) explored ethical issues in collecting data for survey research using crowdsourcing. Several studies considered using crowdsourcing as a fast, cheap, and effective tool for managing data for social science. However, others (Borromeo et al., 2017; Williamson, 2016; Fort et al., 2011) have raised concerns, claiming that low pay rates could challenge the ethics of the data collection process for such studies. Haug (2018) pointed out that in their experiments raising the payment increased the risk of having workers who were used to doing the same type of tasks, potentially affecting the bias of the collected data.

Paul and Lars (2018) developed a model to test the fairness of the payment during the task execution and after the task submission. Goel and Faltings (2019) discussed the fairness and the workers' trust in crowdsourcing platforms. They proposed a mechanism that used peers' answers to verify workers and reduced the number of gold questions needed in the task. Archambault et al. (2015, pp. 27-69) discussed the ethical issues around the use of crowdwork in academic research. The authors recommended following the guidelines provided by the Dynamo (Salehi et al., 2015) project and the Crowdworking Code of Conduct (Graham et al., 2020) as a guide for the researchers planning to use crowd tasks in their work.

In most of the studies around payment issues, researchers strive to pay attention to fair payments when using crowdsourcing tasks (Brawley et al., 2016; Ipeirotis, 2010a). Silberman et al. (2018b) noted the ethical responsibility of paying workers fair wages and discussed the importance of money as a motivating factor for most of the workers as it had been considered in previous studies (e.g., Ross et al., 2010; Ipeirotis, 2010b; Ho et al., 2013; Ye et al., 2017; Finnerty et al., 2013). Moreover, they pointed out that fair payment led to high-quality performance from the crowd. Researchers tried to develop models or implement criteria for calculating a fair payment depending on task type and expected completion time. However, even when the requesters pay a rate in accordance with the minimum wage, workers might still consider the payment unfair: we refer to Section 7.1 for a discussion of this phenomenon.

There is also an urgent need for transparent channel commissions of the platform. First, crowdworkers are considered independent contractors by the platforms. In other words, they are not granted the same protections as 'traditional' employees including minimum wage, employer-sponsored health care, or dismissal protection. As a result, their income is typically unstable and below the local minimum wage (Hara et al., 2018). Furthermore, while crowdworkers may be legally paid less than the local minimum wage, there is a growing consensus that paying too little for research-related crowdsourcing tasks is unethical, and that such tasks should pay at least the minimum wage (Shmueli et al., 2021; Qiu et al., 2019; Haug, 2018). In brief, as crowdworkers lack the security of a platform for their own basic income, researchers are expected to pay them with the full consideration that they are at far greater risk of low income than employees with guaranteed employment contracts.

In our study, we examined the existence of the above issues and focus on intermediary channels and the gap between the actual payment made by the requester and the payment received by the workers on the F8 platform.

2.2. Motivation vs. Reward in Crowdsourcing

Mason and Watts (2009) conducted one of the earliest studies that examined the effectiveness of financial incentives on crowdsourcing task outcomes. The authors discussed the impact of increasing the task rewards on the workers' expectations of the task, and the high rewards were found to make the tasks more attractive to the workers but did not increase the quality of the outcome. A similar study by Borromeo and Toyama (2016) compared the performance of an unpaid crowdsourcing task (self-hosted) with a paid one (via F8). The results of the task used were highly similar in the paid and unpaid conditions, but it took longer to finish the unpaid tasks. In contrast, Kost et al. (2018) defined incentive rewards as one of the four sources of experience meaningfulness for the workers. It was discovered from their experiments that the level to which the payment affects the workers is dictated by their real-world employment status and how much they rely on the crowdsourcing work.

In summary, it can be shown that the impact of the payment cannot be ignored even if it may have only a slight effect on the workers' performance. Ye et al. (2017) investigated the impact of the payment amount on the workers' performance in two types of crowdsourcing tasks. They introduced the concept of Perceived Fairness in Pay (PFP) and measured it in their experiments. This study aimed to clarify the relationship between fair payment and the quality of the results.

More studies investigated extensively the effect of fair payments and the loss of time in crowdsourcing tasks. Researchers discovered a significant gap between earnings, amount of time, and effort required to accomplish a task. They warned the academics and all requesters in general that discarding these details could threaten the attractiveness of crowdsourcing jobs. Hara et al. (2018) discussed workers' earnings on MTurk and considered the unpaid time, including the time spent finding

a task and working on tasks that are later rejected. The authors expressed their concerns about such wasted time, which ultimately affects the hourly wage.

Borromeo et al. (2017) discuss the implementation and evaluation of transparency and fairness principles on a crowdsourcing platform. On the one hand, the authors discussed the fairness in task assignment, completion time, and payment. On the other hand, they recommended having a special framework to encourage a more transparent process for requesters and platform developers. Ho et al. (2015) suggested a different payment scheme such as payment per unit and a bonus for achieving a specific target.

Furthermore, other researchers showed that workers could be motivated and work on a task with low or unfair payment or even work as volunteers if the task has deep meaning to them. Some researchers claimed that workers would respond to good humanitarian causes such as tasks for World Health Organisation (WHO) or disaster responses. For example, Spatharioti et al. (2017) pointed out that workers tend to do more work in, as the authors refer to it, a *"meaningful task"* such as a disaster response task.

Most studies used MTurk to analyse the correlation between the quality of the results and payments. Most of the work in MTurk is *"performance-based"* which means workers tend to submit a high-quality piece of work because they are afraid of rejection if their work does not meet the task criteria or the requesters' expectation (Ho et al., 2015). On the other hand, on the F8 platform low payment could affect the workers' performance differently. Since workers know that they are getting paid regardless of the requesters' job acceptance decision, low payment might not motivate them to expand efforts to submit high-performance results. Our focus in this study is the F8 platform and the variation of payment due to different commission rates taken by the channels. Based on an analysis of more than 53k HITs from previous crowdsourcing projects over four years, we identify the most common channels and explain how they operate.

3. Research Questions

In this paper, we focus on the transparency of the crowdsourcing marketplaces and the channels used to recruit crowdworkers. The business model of traditional crowdsourcing platforms relies on fees applied to the amount paid from requesters to workers for task completion. These fees are fixed by the platforms and are stated clearly to the requesters. In F8, recruitment channels are part of the value chain as well, and this has led to their unique business model. Such business models are not always transparent, and requesters struggle to comprehend how they operate to fairly compensate crowdworkers for completing their tasks. Our contribution aims to shed light on the policies of these channels, including recruitment rules, rates, and methods of payment. Our investigation focuses on F8 as a widely used crowdsourcing platform that offers its own in-house channel called `Elite`, and a number of external recruitment channels. The requesters can decide which

recruitment channels to include for completing a task during its configuration. In addition, all channels are included by default.

Our research questions are:

RQ1 What is the recruitment and reward model of such channels?
RQ2 What is the demographic composition of such recruitment channels?
RQ3 How do the recruitment commissions change over time and over the different channels?
RQ4 What is the impact of the recruitment channel choice on working conditions, e. g., on the hourly wage?

The first research question is addressed in Section 4.2, which provides a description of the reward scheme across the top five recruitment channels. The second research question, concerning the demographic composition of the channels, is critical in understanding the reliability of sampling when conducting experiments via crowdsourcing, and it is explored in Section 5. The third research question is addressed in Section 6 and is of fundamental importance to assess potential ethical issues in academic research. In addition, to answer the fourth research question, in this section we further explore the impact of different levels of reward transparency in recruitment channels on the labour conditions of workers in F8 in this section.

4. Data Collection and Research Methodology

The F8 platform was chosen as the primary focus of the study because it is one of the most popular and widely used crowdsourcing platforms/marketplaces currently utilised by academics. In order to identify the most popular F8 channels and collect reliable information from them, two issues need to be tackled: (i) the fluctuation of channel commissions and (ii) the potential inconsistency of F8 channel commission reports over time. To address (i), we built a metadata archive from a historical collection of F8 tasks. To address (ii), we cross-checked and validated F8 metadata with an ad-hoc survey that allows us to validate the consistency of the reported data. This survey also allowed us to collect additional information on the worker *per-channel* demographics and working profile.

4.1. Historical Metadata

We analysed a collection of 53065 tasks from 133 different jobs carried out across 38 months (from June 2015 to August 2018) from 6803 unique workers and 110 different countries. To create such an archive, we put together job results collected from multiple requesters. The vast majority of these results were collected from tasks that did not have any restrictions on worker expertise level or geographic location. It is important to keep in mind that this data collection uses an opportunistic method to compile all tasks that were available to us at the time as a form of meta-analysis, and as a result, sampling discontinuity over time may be present. While we

refer to other works for a systematic approach (e. g., Difallah et al., 2018) it is worth noting that the primary objective of this study is to identify issues and anomalies in the recruitment channels' payment schemes as reported to requesters by the platform, thus requiring less stringent statistical requirements on the underlying population. As described in detail in Section 5, the presence of such anomalies has been revealed through validating the channel commissions reported by F8 using an ad-hoc survey.

Some of the metadata, like the channel commission, was not available directly from the task output. For this reason, we built a web scraper to download additional metadata that was only presented in the F8 requester web interface. This allowed us to collect the F8 reported channel commission over time. By scraping the channel commission metadata, the actual rewards received by workers for completing tasks could be generated in Table 4. In addition, the scraped channel commission metadata helped us to study the fluctuation of channel commissions over time (Figure 13) and the variation in the size of rewards (Figure 14, 15). A summary of the information contained in the *Historical Metadata* dataset is shown in Table 2.

4.1.1. Most Popular Channels

Our main focus is the popularity of the channels among workers in relation to our Historical Metadata dataset. We computed the number of unique workers per channel, shown in Figure 2. To facilitate further analysis and ensure reliable statistics, in the remainder of this work we focused on the five most used channels, `NeoBux`, `Elite`, `Clixsense`, `InstaGC`, and `Swagbucks`, since all together these channels count 47998 units (93.6% of the total), and 6274 unique workers (97.2%).

The recruitment and reward model of the top-five channels are explained in Section 4.2. In addition, ethical issues of the Paid to Click (PTC) reward structure of some channels are discussed in Section 7.2.

Table 2. List of variables in the *Historical Metadata* dataset.

| Name	Type	Description
contributor_id	string	The ID of the worker who completed this task.
amount	float	The total amount of reward in dollars paid by the requester.
channel	category	The name of the channel which this worker joins to obtain this task.
channel_amount	float	The amount of reward finally assigned to the worker after the commission taken by the channel (in channel currency).
bonus_amount	float	The amount of bonus paid by the task creator to the worker.
job_id	string	The ID of this task.
timestamp	date	Date and time of submission of task results.
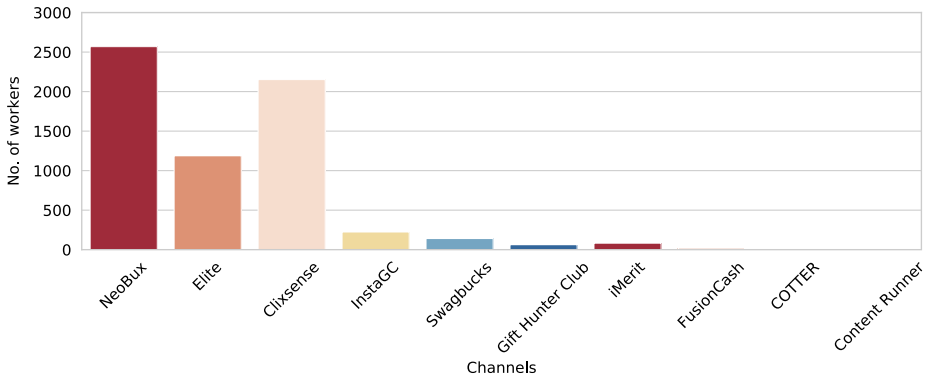
Figure 2. Number of unique workers per channel.

4.2. Reward Model

In this section, summarised in Table 3, the reward model of the top third-party F8 channels is illustrated. We obtained this information by registering as workers in the channels and observing their (often opaque) reward model.

Some of these channels are Paid To Click (PTC) services, in which users are paid by clicking on ads banner, reading and interacting with advertising emails, or by watching video ads. In addition to traditional PTC services, other ways to earn money in these channels include playing games, cash-back systems, e-shop sign-up offers, and micro-tasks/online surveys. This study will only focus on the latter because it is the only section that is exposed to F8.

Table 3. Third-party channels' reward scheme summary.

Channel	Reward method	Min withdrawal	Exchange Rate	Reward availability
Clixsense	PayPal, Paytoo, Payza, Neteller, Check	8 USD for PayPal, Paytoo and Payza, 10 USD for Check. 6 USD for Premium members	USD	7 days
InstaGC	eCheck, Direct deposit, Gift card, PayPal, Cryptocurrency	1 USD	100 points = 1 USD	1–4 days
NeoBux	PayPal, Payza	2 USD to 10 USD (up with no. withdrawals)	USD	unknown
Swagbucks	PayPal, Visa and Merchant gift cards	1 USD	100 SBs = 1 USD	1 month

4.2.1. Elite Channel

The `Elite`[3] channel is the official channel of F8, making it the most straightforward way to access a task in F8 directly from the platform web page. Unlike the other four channels considered in this study, `Elite` does not require any channel commission: workers are rewarded with the full amount paid by the requesters. `Elite` channel has a qualification system that works as follows: first, workers have to complete successfully at least 100 test questions without rewards to get qualified for working on the paid tasks; then, they are assigned a qualification level based on their accuracy. The qualification levels are: *level 1* for workers with at least 70% accuracy, *level 2* for workers with at least 80% accuracy, and *level 3* for workers with at least 85% accuracy. This effectively creates a barrier for workers struggling to complete such tasks to access this channel (e. g., non-native English speakers).

4.2.2. NeoBux Channel

As shown in Figure 2, `NeoBux`[4] is the most used channel in terms of both units and workers. `NeoBux` is a PTC platform established in 2008 which offers free registration for *Standard* membership. `NeoBux` pays members for carrying out simple tasks like clicking on ads. The number of clicks a member can do daily is limited and workers are required to be active daily to avoid suspension or cancellation of their membership. Workers can earn more by upgrading to *Golden* membership for 90 USD per year, for which they will get up to 2000 clicks per month at 0.01 USD, and rent referrals/subcontracting crowd work (where workers can spend credit to hire other workers' clicks). Workers can withdraw it to their PayPal and Payza accounts with a 2 USD minimum withdrawal limit for the first time. After that, they will be allowed to withdraw again when they reach a fixed minimum amount of 10 USD. The crowdsourcing tasks come as mini jobs for the workers to earn extra money, but it is typically not the main source of income for `NeoBux` users.

4.2.3. ClixSense Channel

Established in 2007, `Clixsense`[5] is one of the most popular online PTC platforms. On this platform, a weekly contest is held, in which the top ten workers (those who complete the most tasks) compete for a total prize pool of 100 USD, with 50 USD going to the best worker. The tasks include completing surveys, testing new products and downloading new apps, F8 tasks, watching videos, etc. `Clixsense` offers a *Standard* and a *Premium* membership option. The difference between these two options is the percentage of money received from doing the daily checklist and the amount earned from their referrals. Members are assigned referral links, and a worker will receive a 20% commission on the earnings of their

---

[3] https://elite.figure-eight.com

[4] https://www.neobux.com

[5] https://www.clixsense.com

referrals at `Clixsense`. Payments are issued every Monday if the worker has earned more than 8 USD for *Standard* members and 6 USD for *Premium* members. As a motivation, this channel offers a 5 USD bonus if the worker earns 50 USD. The minimum reward for a task is 1 cent and if the worker completes a task worth less than 1 cent, they will not get paid in `Clixsense` unless they complete another task for the same job.

4.2.4. InstaGC Channel

Established in 2011, `InstaGC`[6] channel, which is similar to `Clixsense` and `NeoBux` in terms of services and referral system, allows free registration. The only benefit of using it over the previous two is that the payout threshold is only 1 USD for 100 collection points. The payment is in the form of a gift card or a cash payment made through bitcoins or other electronic money transactions with a fee associated with the cash exchange process.

4.2.5. Swagbucks

This channel is related to `Swagbucks`[7] by Protege, a reward and loyalty operator that offers cashback and vouchers. Users can earn the so-called swagbucks (SBs), a virtual currency that can be used for online shopping directly or exchanged for US dollars. Users can earn SBs by using the `Swagbucks` search engine, playing games, watching videos, shopping online, answering surveys, and completing F8 tasks (which is the means of income this study focuses on). The primary ways to redeem SBs are PayPal, Visa gift cards, and Merchant gift cards. For each 100 SBs, a worker can redeem 1 USD at the end of the month.

From the point of view of currency stability, as defined by Kumar (2009), SBs are classified as a gambling asset (Novotnỳ, 2018), meaning that the virtual currency possessed high idiosyncratic volatility, high idiosyncratic skewness and low price.

`Swagbucks` associates with additional activities that potentially raise ethical issues when used for academic research: `Swagbucks` reward users for subscribing and using gambling services. These types of rewards constitute the majority of the offers appearing in the discover section and in the inbox when using the `Swagbucks` platform.

5. Survey

As the main experiment of this work, we restricted our focus to the five most popular channels as per the Historical Metadata dataset, as described in Section 4.1. We designed a survey, integrated it into a crowdsourcing task, and collected the

---

[6] https://www.instagc.com
[7] https://www.swagbucks.com

answers from 60 workers for each of the five most popular channels. While this number might not be sufficient to reliably draw conclusions on the demographics, it has proven to ensure significance for the analysis of the channel commissions (as discussed in Figure 11 and related statistical tests), and a qualitative analysis that achieved saturation on the open-ended questions (Hennink and Kaiser, 2022; Fofana et al., 2020; Rowlands et al., 2016). We ran the survey as a F8 task and paid each respondent 0.5 USD: We chose this amount on the basis on the expected completion time obtained with a pilot experiment (3 minutes), with the goal of providing an equivalent hourly wage of 10 USD, above UK minimum wage. While we made sure to prevent single accounts to complete the survey multiple times, we cannot guarantee that two accounts in different channels belonged to different crowdworkers. This could be verified with forms of user tracking using permanent cookies (Klein and Pinkas, 2019). However, that is beyond the scope of this work and would require specific ethical safeguards. While the actual time spent on the survey could not be measured accurately (especially for channels where crowdworkers engage in different activities in parallel), the median difference between worker acceptance and submission of the survey was under 3 minutes for the channels InstaGC and Swagbucks, and under 9 minutes for the other channels.

In the first part of the survey, we focused on the demographics of channels' users, asking participants about their age, gender, education, experience as crowdworkers, employment, type of device used to perform crowdsourcing tasks, monthly income, and criteria of HIT acceptance. The second part of the survey investigated the rewarding factor, with questions about how and how much the workers get paid by each channel. These responses, together with the scraped historical data of F8 reported channel commissions described in Section 4.1, allowed us to reconstruct and validate the actual worker reward over time for each channel. Here is an illustration of the demographics information collected in our survey, please see Section 6 for a detailed analysis of the channel commissions.

5.1. Age

In the first question, we asked the participants to specify their ages by choosing from a list of six age ranges. Figure 3 shows that participants' ages slightly differ across channels; e. g., 30% of NeoBux users were younger than 25, whereas no participants in that age range were recorded for Swagbucks.

5.2. Gender

We then investigated the participants' gender. The responses we collected show different trends in terms of gender distribution over channels. As shown in Figure 4, InstaGC and Swagbucks workers are mostly females (more than 60% in both the channel), whereas users of Clixsense, Elite, and NeoBux are mostly
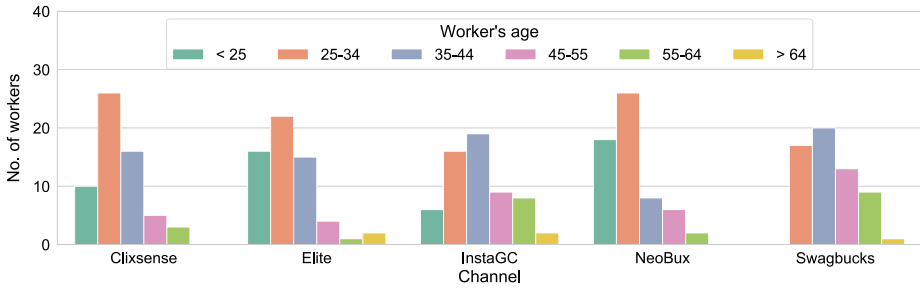
Figure 3. Survey responses on worker age.

male (more than 70% of the users in each of these channels). This difference might be related to the target customer segment associated with `InstaGC` and `Swagbucks` voucher redemption schemes. It is important to consider these differences to avoid unintended sampling biases.

5.3. Education

The third question of our survey concerned workers' education levels. We identified nine education levels, ranging from *no education* to *doctorate*. Figure 5 shows that for all channels the most common educational level is *bachelor's degree*. The vast majority of the users is included between *High school diploma* and *Professional degree*, and the only noteworthy difference between channels is the ratio between bachelor's and master's degree (e. g., `Clixsense` vs. `Elite`). It is worth noting that these results might contradict the narrative of a gig economy meant to provide extra income for workers at the early stage of their career, as also discussed in previous studies where this contested framing from platforms is referred as "beer money" (Bates et al., 2021; Berg, 2015; Tassinari and Maccarrone, 2020).
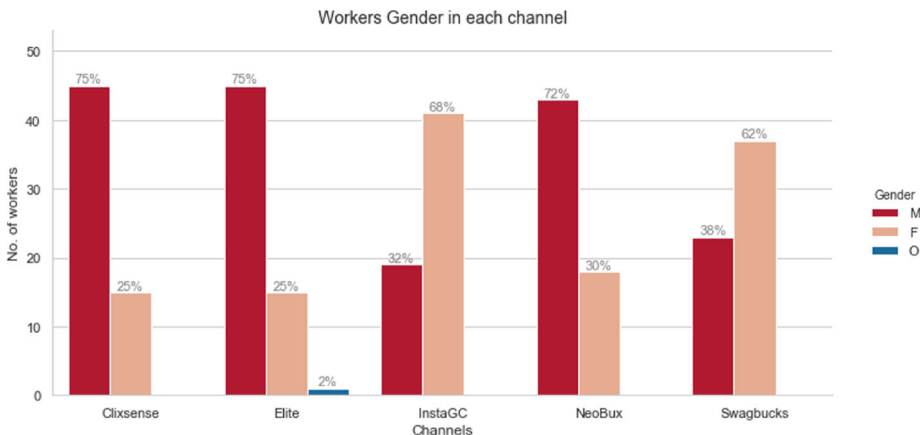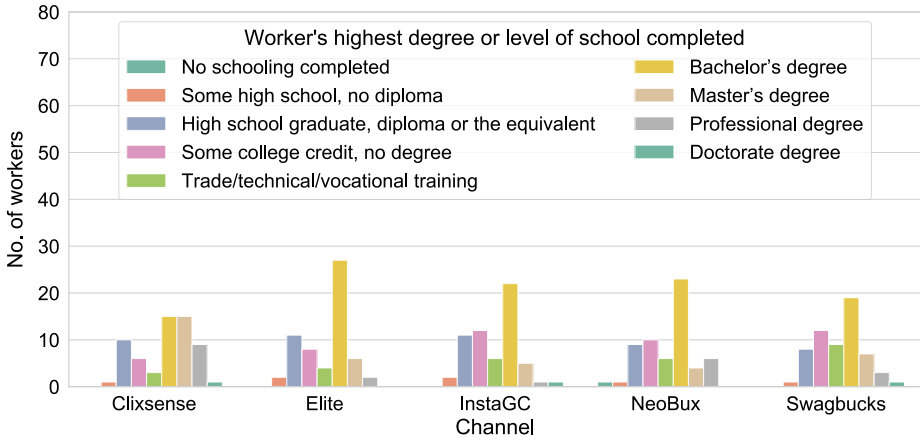

Figure 4. Survey responses on worker gender.

Figure 5. Survey responses on worker education.

5.4. Experience as Crowdworker

We asked participants to indicate the number of years of experience as a crowd-worker. Figure 6 shows that the vast majority of `InstaGC` and `Swagbucks` workers are experts, reporting more than two years of experience. The experience of the workers of the channels `Clixsense`, `Elite`, and `NeoBux` is more evenly distributed.

Employment

We then asked respondents to specify their employment status. Figure 7 shows the distribution of the results. A general trend, consistent across channels, seems to indicate that the majority of participants reported being "Employed for wages" or "Self-
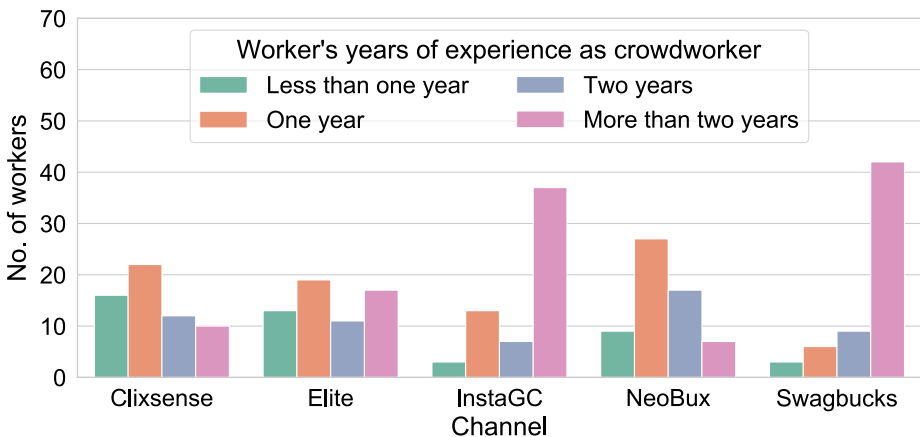
Figure 6. Survey responses on worker experience as crowdworkers.

employed". A noteworthy difference is the case of `InstaGC` and `Swagbucks`, where more than 58% of the respondent reported being "Employed for wages".

5.5. Desktop vs. Mobile Users

Previous works have shown a trend in micro-task crowdsourcing in offering tasks optimised for desktop rather than mobile devices (Mea et al., 2015). Consequently, users find accessing their crowdsourcing platforms to perform their HITs through desktop devices more convenient. To investigate if this behaviour varies across the recruitment channels, we asked participants what kind of device they use to perform crowdsourcing micro-tasks. Results confirmed that only 2% to 5% of the workers perform micro-tasks from mobile devices, without noteworthy differences across channels.

5.6. Income

We asked participants to report their monthly earnings from crowdsourcing micro-tasks. Figure 8 shows the reported monthly income for each channel.

There was a statistically significant (with threshold $p = 0.05$) difference between groups as determined by one-way ANOVA ($F(4, 295) = 14.885$, $p < 0.001$). A Tukey post hoc test revealed that `Swagbucks` reported monthly earnings are statistically different from all other channels. Additional significant differences were reported only between `InstaGC` with `Clixsense` and `Elite`.

5.7. Task Acceptance Criteria

Then, we focused on workers' task acceptance criteria. Participants could choose more than one answer. As shown in Figure 9, for all channels, the reward amount has
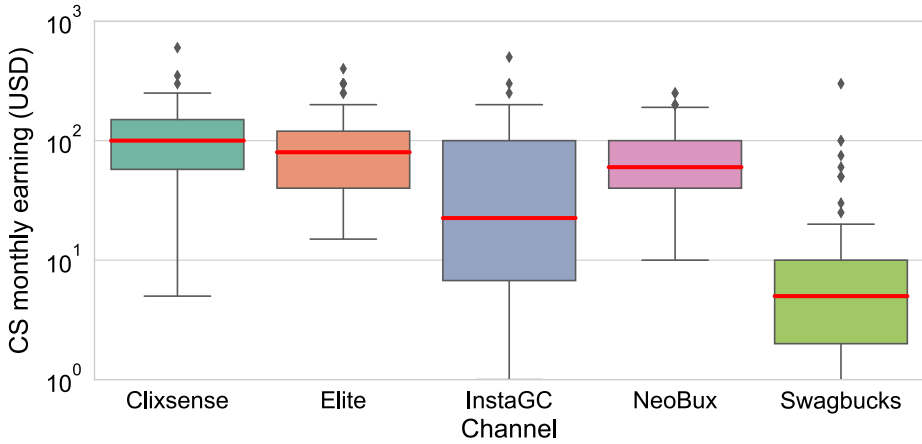

Figure 7. Survey responses on worker employment.

Figure 8. Reported monthly earning (log scale).

the greatest influence on the decision to accept a task. Task difficulty, completion time, and the interest aroused by the task are also a factor taken into consideration. It is also noteworthy that for `NeoBux` and `Swagbucks` some participants reported that the task was provided to them by the channel, without the possibility of choosing another one. The criteria used by the channels to decide which tasks to assign are not publicly disclosed. This approach is in stark contrast with `Elite` and MTurk, where a search engine is provided to the crowdworker to freely select the HITs to complete.

6. Channel Commission Analysis

In this section, we present the analysis of the actual worker reward over time per channel, obtained by first comparing the worker self-reported values with the
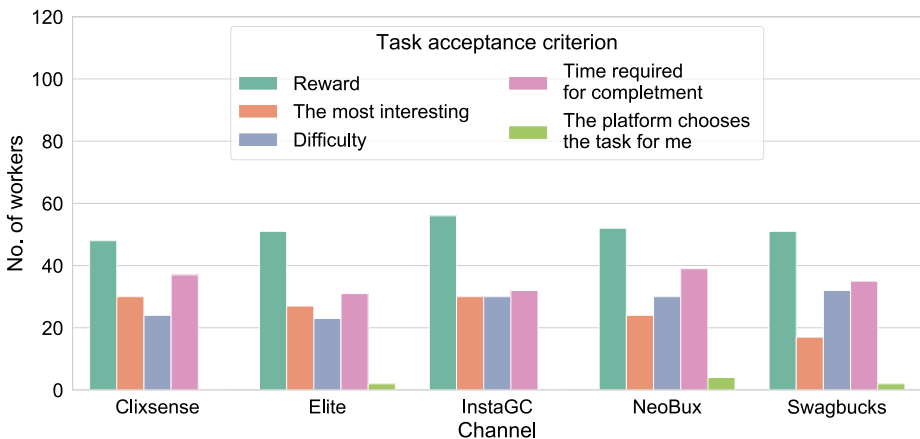

Figure 9. Survey responses concerning the criteria used to accept beginning to work on a task.

amount paid by the requesters, and then using these values to validate the scraped historical data of the F8 reported channel commissions.

6.1. Survey Completion Reward

We paid 0.5 USD (plus the 20% for the F8 *platform fee*) for the completion of the survey. Every channel, with the exception of `Elite`, applied an additional *channel commission*.

The F8's dashboard, depicted in Figure 10, shows the final amount to be paid to the worker for the completion of a task. The payment includes both the *platform fee* and the *channel commission*. It is worth noting that the latter cannot be accessed by the requester in advance, since they get published in the dashboard only after the completion of a task. Moreover, they change over time. Interestingly, the values reported present some inconsistencies: for example, at the time of the survey the platform reported for `Clixsense` a worker compensation two order of magnitude smaller than the ones of the other channels.
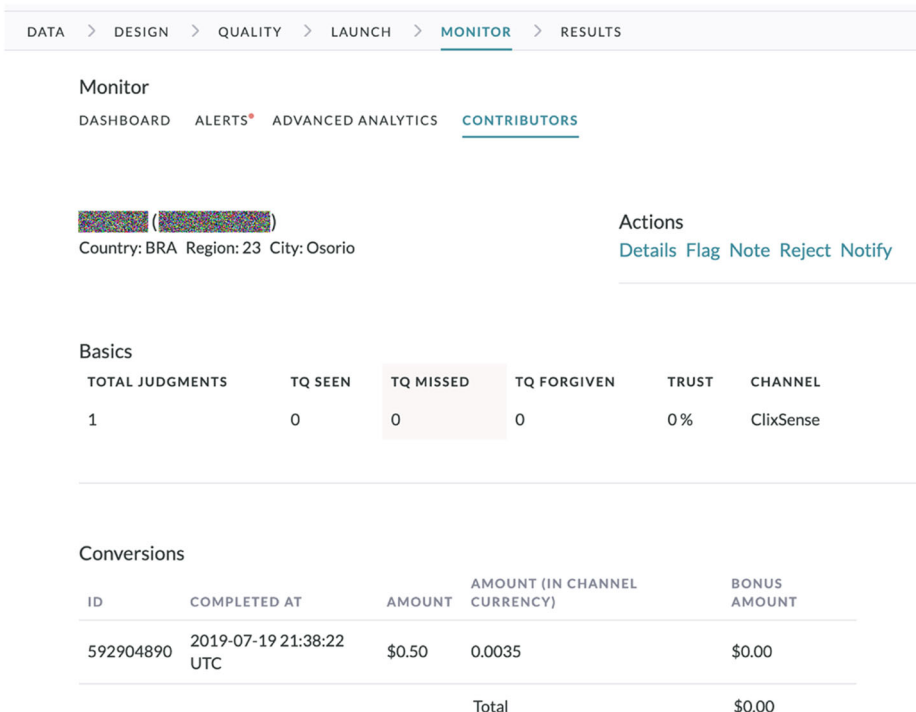

Figure 10. The F8 monitor shows details about a the contributions of workers in a survey's assignments. The "Amount" column refers to the payment allocated by the platform for the resolution of the assignment, whereas the "Amount (in channel currency)" column concerns the amount due to workers. We blurred worker's identifier and IP address for privacy reasons.
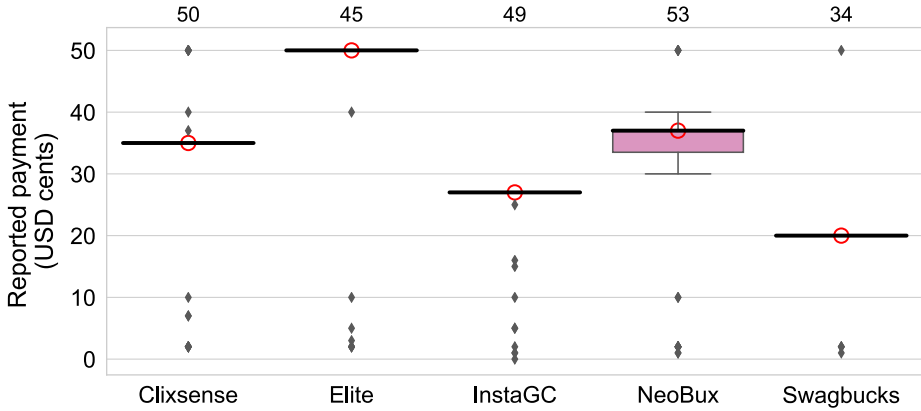
Figure 11. Payment per channel reported by the survey's participants who reported the channel currency as USD (231 out of 300 participants). On top of each box, the number of valid responses collected is shown. The red circles indicate the amounts reported by F8 (net of channel commission). Circles overlap exactly the medians (and thus are well within the IQR), suggesting that the channel commissions reported by the participants reflect those declared by F8. It is also worth noting that the dispersion is very low, as the IQR is equal to zero for all channels except `NeoBux`, where it is lower than 5c.

To validate, and possibly correct, these misalignments, in the survey we asked the following question: "How much money will you earn carrying out this task?". The responses are shown in Figure 11. The black bold line represents the median of the distribution of the workers' answers. Despite some outliers, the worker-reported rewards match those reported on the F8 dashboard, after a multiplicative correction of the case of `Clixsense` where the amount reported in the dashboard was 0.0035 instead of the worker-reported value of 0.35 USD. Moreover, the presence of noise in `NeoBux` interquartile range indicates a slightly higher disagreement in workers' responses, potentially caused by the more complicated reward procedure of these channels, as described in Section 4.2.

The effective rewards received by the workers per channel are summarised in Table 4. The channel `Elite`, which does not apply any commission, is the only one where the workers get the full amount we paid consisting of 0.5 USD. Because of the *channel commission* applied, the *worker payment* by the other four channels is

Table 4. Worker reward (USD) net of the *channel commission*, also reported as the percentage of requester payment, as obtained from the F8 dashboard for the task corresponding to the survey.

Clixsense	Elite	InstaGC	NeoBux	Swagbucks
.35 (70%)	.5 (100%)	.27(54%)	.37(74%)	.2(40%)

lower: 0.37 USD by `NeoBux`, 0.35 USD by `Clixsense`, 0.27 USD by `InstaGC`, and 0.2 USD by `Swagbucks`.

6.2. Withdrawal Delays

We focused on the type of reward workers get while using different channels. Around 60% of workers in `InstaGC` and `Swagbucks` reported receiving the reward as an instant electronic payment (e. g., Paypal), while over 35% of workers of the rest of the channels claimed of receiving money with a bank transfer, thus with additional delay.

As shown in Figure 12, the majority of workers in `Clixsense` and `Elite` claimed that it took a few days for their withdraws to complete. In comparison, it took only a few minutes for most workers in `InstaGC` to withdraw rewards from their accounts. This big difference on their withdrawal delays could be due to different reward methods that workers in each channel prefer (Table 3). The differences in reward delays is compounded by the use of multiple withdrawal methods and currencies (including cryptocurrencies) for some channels.

6.3. Reward Loss Over Time

As shown in Figure 13, channel commissions are subject to fluctuations, in some cases of considerable amounts e. g., for `NeoBux` and `InstaGC`. Nevertheless, the overall trend is confirmed: `NeoBux` compensates workers with 75-80% of the original payment, `Clixsense` and `InstaGC` the 60%, and `Swagbucks` is the channel that takes the highest commission since it pays workers with only 40% of the money initially allocated.
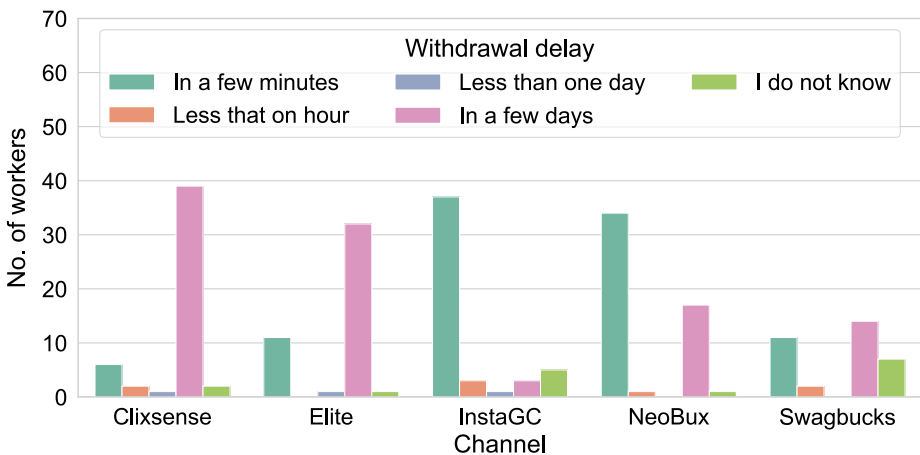

Figure 12. Workers reported different delays across channels in the withdrawal of their reward after completing a HIT.
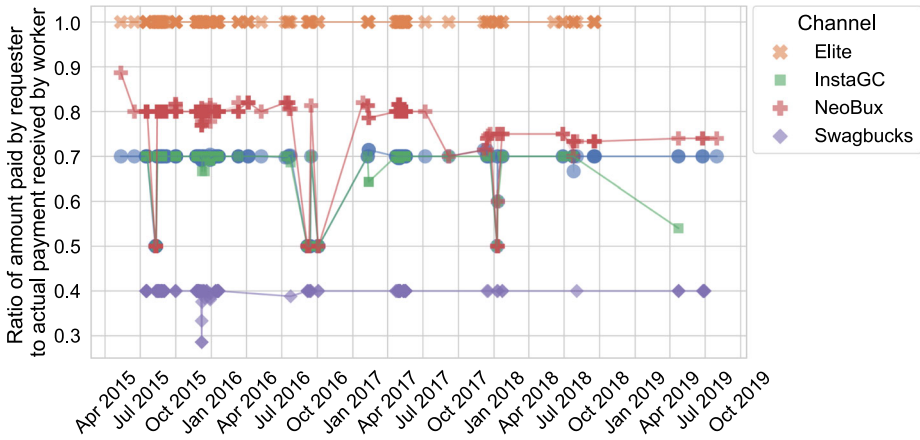
Figure 13. Ratio of requester's payment actually received by the worker observed from March 2015 to September 2019.

6.4. Worker Loss Per Channel by Reward Size

The channel commissions are also influenced by the amount of the reward. As shown in Figure 14, channels tend to retain a larger part for smaller rewards.

A use case scenario can simulate the workers' loss based on which channel they choose to use and the average reward size. Using the values from the *Historical Metadata* dataset, we estimated the amount of money lost per year, grouped by channel and requester payment bracket. The results, shown in Figure 15, indicate a sizeable cumulative loss, ranging from about 20% to 60%.
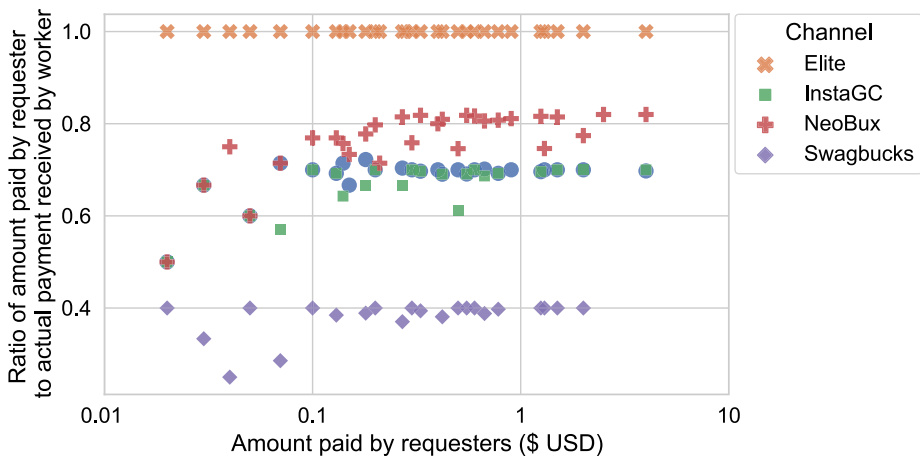

Figure 14. Reward ratio received by the worker vs amount paid by the requester

6.5. Workers Feedback

Workers expressed an overall positive sentiment towards the goal of the survey, with remarks about the need of demonstrating the "injustice" and unfairness of the payment process, especially because workers often feel that the relationship between employers and workers is quite "unbalanced". Despite such issues, some workers expressed their gratitude towards the opportunity of microtask crowdsourcing, which has been a "lifeline" in moments of financial uncertainty.

Moreover, several workers pointed out that, when working on crowdsourcing tasks, instructions sometimes are ambiguous and not clear, which leads them to abandon the task or submit wrong answers. Workers also pointed out that the number of tasks on the platform has decreased over time and the platform has been suffering from many technical issues.

7. Discussion

We discuss here our findings from both requester and crowdworker perspectives, consider their implications in the micro-task crowdsourcing ecosystem, and point out potential solutions and connections with related academic work.

7.1. Impacts of Unethical Crowdsourcing

7.1.1. Underpayment, Unfair Workload and Lack of Transparency for Workers
Crowdworkers are exploited by hidden rules on the platforms, such as being encouraged to complete a large number of zero-reward HITs in order to obtain higher reputations (including the number of tasks completed and task approval rate), which
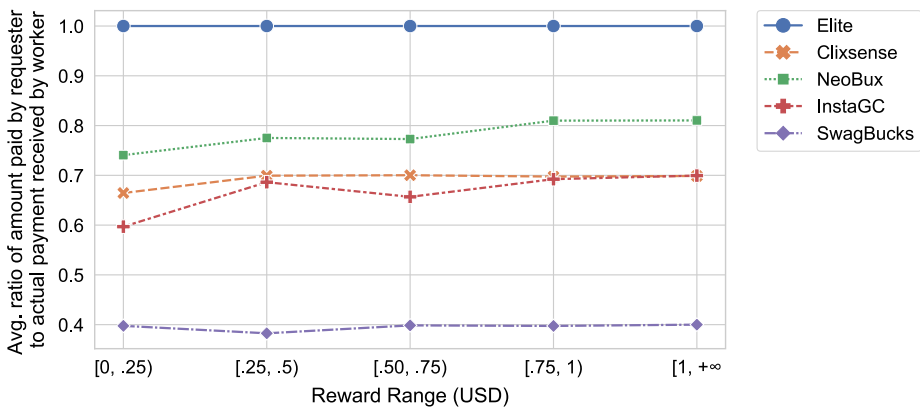

Figure 15. Mean ratio of amount paid by the requester to actual payment received by the worker computed on *Historical Metadata* dataset from 2015 to 2019.

makes them more likely to receive high-reward tasks in the future (Gupta et al., 2014). Our finding also reveals another opaque aspect of crowdsourcing market-places, which is the widespread presence of channel commissions. Worse still, crowdworkers are generally unaware of the existence of channel commissions, and thus have been subject to hidden exploitation for a long time. Thinking in terms of the opposite of this study's intention, it may be reasonable for channels to charge commissions as an incentive for recommending tasks to channel members. How-ever, it is worth thinking more deeply about how much commission is ethical and its impact on the overall market.

As one of the reasons for underpayment, badly designed tasks, technical errors, and interface design errors (McInnis et al., 2016) made by the job requesters may confuse workers and result in extra time, efforts, and even failure of submission or increased risk of rejection (Gadiraju et al., 2017). Even though they are paid, crowdworkers may need to spend additional time and effort searching for tasks, learning how to do those they are not familiar with, and waiting for the response to their questions from the requesters. Our findings confirm the existence of payment delays within the channels such as Clixsense. This indicates that payment delays come not only from job requesters who decide to accept or reject the results but also from the channels' own payment policies. A worker should be paid for the work done when the requester approves their submission. This delay contributes in deprecate the quality of the crowdwork experience.

The lack of understanding of channel commissions by crowdworkers identified in this study has led us to pay attention to the lack of transparency of informa-tion from the workers' perspective on the platform. Crowdsourcing platforms have always tried to avoid the traditional ways of human interaction in the work environ-ment, such as anonymising members and limiting the ways of interaction among them (Martin et al., 2017). As a result, the extremely low quality of communication directly undermines interpersonal trust, gradually causing a stripping away of the ethical guidelines of traditional work from crowdsourcing platforms. This could in turn contribute to a culture of extremely high channel commissions such as InstaGC and Swagbucks.

7.1.2. Reputation and Data Quality Risks for Requesters

Unethical crowdsourcing can create challenges for the requester or the institu-tions they belong to. As news of workers' exploitation spread through communi-cation and rating systems like Turkopticon and TurkerView, the reputation of the requester is harmed, making it more difficult for them to recruit workers in the future (ChrisTurk, 2022; Hanrahan et al., 2021; Gaikwad et al., 2016; Salehi et al., 2015; Irani and Six Silberman, 2013). Moreover, rating systems to keep crowdsourcing platforms accountable are gradually being developed (e. g., the Fairwork project by Fredman et al., 2020), potentially making virtuous platforms more attractive.

An increasing number of studies is looking at whether the use of data collected via crowdsourcing differs between business, public use and academia, and whether

this difference creates ethical challenges for both the data provider and the demand side (Gleibs, 2017; Vayena et al., 2015). However, the market's over-reliance on a single crowdsourcing platform leads to an inevitable rise in platform fees such as MTurk, thus challenging the fairness of payments to the crowdworkers in commercial and academic crowdsourcing projects (Haug, 2018; Gleibs, 2017). Therefore, equal attention needs to be paid to the fairness of the compensation given to workers or participants in commercial and academic crowdsourcing projects, and the difference in the treatment of this ethical challenge between the commercial and academic job requesters. Academic institutions keep a higher degree of ethical standards for worker compensation than commercial institutions (Shmueli et al., 2021; Gleibs, 2017). Moreover, academics have been actively looking for ways to improve payment fairness for crowdworkers (Fredman et al., 2020; Qiu et al., 2019; Whiting et al., 2019). As a result, academic job requesters often pay more than commercial job requesters (Rea et al., 2020).

A consensus agreement on the correlation between pay level of crowdworkers and data quality has still not been reached (Auer et al., 2021; Litman et al., 2015; Buhrmester et al., 2011). This is probably because the data quality is influenced by multiple factors, not just compensation. In other words, although the workers are not paid a fair reward, they might still maintain a high level of performance due to the penalties set by the platform (Auer et al., 2021). However, maintaining ethical rewards can help improve the data quality of those who treat the rewards as a primary source of income (Litman et al., 2015). In addition, the impact of pay level on worker satisfaction and turnover is clear, which in turn can affect the willingness of workers to continue working for the requesters or even refuse to continue working for the requesters altogether due to low rewards (Kees et al., 2017). And after being rejected by quality workers, job requesters may end up having no choice but to hire workers with lower-quality responses to complete the task, which in turn may reduce the quality of the work.

Based on the findings of this study, we encourage commercial job requesters to be fully aware of the risk of underpayment to workers arising from the floating channel commissions and to maintain a sufficient ethical standard of payment. This will also help to ensure the quality of the data and attract sufficient workers for continuous participation in the project (Auer et al., 2021; Litman et al., 2015).

7.1.3. Potential Solutions to Unethical Channel Commissions

One possible solution to help reduce unethical channel commissions could be to encourage workers and requesters to share the amounts they receive and pay for the same task through a browser plugin or script, similarly to other semi-automated "sousveillance" tools proposed in the literature (e. g., Checco et al., 2018). Designing such a tool would need to take into consideration potential issues in the data collection and reporting process, especially because channel commission change over time. This tool could calculate and share with the users the percentage of commission charged from a specific channel, which in turn could be monitored over

time. It could be used to create a leaderboard for channel commissions based on monitoring data, thus encouraging both workers and requesters to choose the channels with more reasonable commissions and reducing the extent to which they are exploited. The potential of this solution is not only to help facilitate ethical channel commissions, but also to be a useful attempt to promote cooperation between workers, and even between workers and requesters, in sharing information and thus improving unreasonable policies on the platform.

7.2. Ethical Considerations of PTC Services

Another point worth mentioning is the questionable nature of the PTC hierarchical payment scheme: users can make use of rent referral, that is effectively similar to subcontracting, where they can bet on the productivity of other users by renting their clicks for a set period of time. However, users need to pay a membership fee to gain access to advanced rent referral options, as well as pay their own money in the hopes of a potential return, bringing the platform dangerously close to a Ponzi Scheme (George, 2018). The functionalities of these platforms are close to those of Traffic Monsoon, where the main source of revenue was coming from the users that joined the platform as PTC workers. This practice caused a legal action from the USA Securities and Exchange Commission (SEC) (Penman, 2019).

While we focus on the crowdsourcing revenue source of these channels, we cannot dismiss the potentially unethical nature of these companies, both towards the advertisement systems (by gaming the advertisement statistics with ungenuine clicks allowing them to market themselves as a successful advertisement service) and towards the workers, that inevitably lose money while working under a complex pyramid system (George, 2018).

8. Conclusions

When budgeting for crowdsourcing tasks, requesters need to consider the overall cost of the commissioned task, which is often represented by the crowdsourcing platform as the sum of the platform fee and the contributors' reward.

However, crowdsourcing platforms like Figure Eight, can make use of outsourcing companies (external channels for crowdworkers recruitment). Some of these channels will withhold part of the reward as channel commissions and pose restrictions on accessing such rewards. While requesters can select which channels to include in their crowdsourcing tasks, no information about the channels' policies is provided. Even more importantly, the crucial information about the amount of channel commission is only revealed by the platform after the job is completed, and its value fluctuates over time. These practices make it extremely difficult to provide guarantees on the amount and modalities of compensation provided to the workers. These guarantees are required in a variety of situations, including data collection that requires compliance with ethical guidelines.

In this paper, we combined four years of historical data and an ad-hoc survey to identify the currently most popular channels, collected information about their policies and demographics, and investigated the gap between the reward paid by the requester and the part actually received by the workers. The survey allowed us to highlight the differences among workers recruited by different channels in terms of gender, years of experience as crowdworkers, and monthly earnings. Such differences should be taken into account when planning a research project, as they could influence the sampling process and may cause unintended biases.

The results of our investigation into channel commissions indicate an imbalance in the treatment of workers due to differences in channel policies. We showed that out of the top five channels only one, `Elite`, does not charge additional platform fees because it is the one owned by the platform itself. Workers who were surveyed indicated that they were receiving unequal payments and that they were unaware of the discrepancy between the intended amount by the requester and the amount they actually received due to channel commissions. We observed that some channels provide a variety of services to the workers, and doing a crowdsourcing task is only one of the extra jobs that they can do to get extra rewards or points. Furthermore, it has been discovered that some of the most common channels are Paid To Click platforms, which have been connected to potentially unethical behaviours towards the workers such as the use of complex pyramid systems and the reward of gambling activities. Regarding the worker earnings, our analysis shows a sizeable cumulative loss due to channel commissions, ranging from about 20% to 60% depending on the channels they belong to. This, in turn, leads us to discuss the potential impacts of unethical payments arising from opaque channel commission schemes.

We can generalise the lessons learned from our study and group them by the three main paid micro-task crowdsourcing actors: (i) Workers, who are the weak link in the chain, should be made aware of the different policies of the recruitment channels. They could increase their participation in dedicated online discussions, forums, and initiatives aimed at identifying unethical channels, boycotting them, and reporting them to the crowdsourcing platforms and to the requesters who might be unaware of the issue; (ii) platforms should develop public and transparent policies to guarantee that the recruitment channels operate fairly and ethically, excluding those that do not adhere to the stated policies; (iii) the requesters should become aware of the problems associated with the various recruitment channels, and prefer official ones (e. g., `Elite` in case of Figure Eight) when it is not possible to verify that other channels operate fairly and ethically.

In the future, we will investigate how economic changes in particular countries have affected workers' willingness to work on crowdsourcing platforms over the last four years. Moreover, some of the workers involved in our study will be interviewed for more details on the assumptions generated from this study. In addition, workers from other channels and crowdsourcing platforms will be surveyed as an extension of this research.

Declarations

Conflicts of interest

The authors declare that they have no known financial or personal conflicts that would have appeared to have an impact on the research presented in this study.

Open Access

References

Andersen, David J; and Richard R Lau (2018). Pay Rates and Subject Performance in Social Science Experiments Using Crowdsourced Online Samples. *Journal of Experimental Political Science*, vol. 5 , no. 3 , Winter 2018, pp. 217 – 229

Archambault, Daniel; Helen Purchase; and Tobias Hoßfeld (2017). *Evaluation in the Crowd. Crowdsourcing and Human-Centered Experiments: Dagstuhl Seminar 15481, Dagstuhl Castle, Germany, November 22–27, 2015, Revised Contributions*, vol. 10264. Springer

Auer, Elena M.; Tara S. Behrend; Andrew B. Collmus; Richard N. Landers; and Ahleah F. Miles (2021). Pay for performance, satisfaction and retention in longitudinal crowdsourced research. *PLOS ONE*, vol. 16

Bates, Jo; Alessandro Checco; and Elli Gerakopoulou (2021). *The Ambivalences of Data Power: New perspectives in critical data studies*, Palgrave, chap. Worker perspectives on designs for a crowdwork co-operative

Berg, Janine (2015). Income security in the on-demand economy: Findings and policy lessons from a survey of crowdworkers. *Comp. Lab. L. & Pol'y J.*, vol. 37, p. 543.

Borromeo, Ria Mae; Thomas Laurent; Motomichi Toyama; and Sihem Amer-Yahia (2017). Fairness and Transparency in Crowdsourcing. In *Proceedings of the 20th International*

Conference on Extending Database Technology, Venice, Italy, 21 Mar-24 Mar 2017. Konstanz, Germany: OpenProceedings, pp. 466–469

Borromeo, Ria Mae; and Motomichi Toyama (2016). *An investigation of unpaid crowd-sourcing*, vol. 6. Springer.

Brawley, Alice M; and Cynthia L S Pury (2016). Work experiences on MTurk: Job satis-faction, turnover, and information sharing. *Computers in Human Behavior*, vol. 54, pp. 531–546

Buhrmester, Michael; Tracy Kwang; and Samuel D. Gosling (2011). Amazon's Mechanical Turk: A New Source of Inexpensive, Yet High-Quality, Data? *Perspectives on Psycho-logical Science*, vol. 6, pp. 3–5

Callison-Burch, Chris (2009). Fast, cheap, and creative: Evaluating translation quality using Amazon's Mechanical Turk. In *Proceedings of the 2009 conference on empirical methods in natural language processing, Singapore, 6-7 August 2009*. Stroudsburg, PA, United States: Association for Computational Linguistics, pp. 286–295

Checco, Alessandro; Jo Bates; and Gianluca Demartini (2018). All That Glitters is Gold – An Attack Scheme on Gold Questions in Crowdsourcing. In *Sixth AAAI Conference on Human Computation and Crowdsourcing, Zürich, Switzerland, 6 July - 8 July 2018.* New York: ACM Press

ChrisTurk (2022). TurkerViewJS. https://turkerview.com/mturk-scripts/ 1-turkerviewjs.

Della Mea, Vincenzo; Eddy Maddalena; and Stefano Mizzaro (2015). Mobile crowdsourc-ing: four experiments on platforms and tasks. *Distributed and Parallel Databases*, vol. 33, no. 1, pp. 123–141

Deng, Jia; Jonathan Krause; Michael Stark; and Li Fei-Fei (2016). Leveraging the Wisdom of the Crowd for Fine-Grained Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, pp. 666–676

Difallah, Djellel; Elena Filatova; and Panos Ipeirotis (2018). Demographics and Dynamics of Mechanical Turk Workers. In *Proceedings of the Eleventh ACM International Confer-ence on Web Search and Data Mining (WSDM '18), Marina Del Rey, CA, USA, February 5 - February 9 2018*. New York: ACM Press, vol. 9, pp. 135–143

Fan, Shaoyang; Ujwal Gadiraju; Alessandro Checco; and Gianluca Demartini (2020). CrowdCO-OP: Sharing Risks and Rewards in Crowdsourcing. *Proceedings of the ACM on Human-Computer Interaction*, vol. 4, no. CSCW2, pp. 1–24

Fieseler, Christian; Eliane Bucher; and Christian Pieter Hoffmann (2019). Unfairness by Design? The Perceived Fairness of Digital Labor on Crowdworking Platforms. *Journal of Business Ethics*, vol. 156, pp. 987–1005

Finin, Tim; Will Murnane; Anand Karandikar; Nicholas Keller; Justin Martineau; and Mark Dredze (2010). Annotating Named Entities in Twitter Data with Crowdsourcing. In *Proceedings of the NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon's Mechanical Turk, Los Angeles, California, US, 6 June 2010*. USA: Association for Computational Linguistics, CSLDAMT '10, p. 80–88

Finnerty, Ailbhe; Pavel Kucherbaev; Stefano Tranquillini; and Gregorio Convertino (2013). Keep It Simple: Reward and Task Design in Crowdsourcing. In *Proceedings of the*

Biannual Conference of the Italian Chapter of SIGCHI, Trento, Italy, September 16 - September 20 2013. New York: ACM Press, CHItaly '13, pp. 14:1–14:4

Fofana, Fatoumata; Pat Bazeley; and Antoine Regnault (2020). Applying a mixed methods design to test saturation for qualitative data in health outcomes research. *PloS one*, vol. 15, no. 6

Fort, Karën; Gilles Adda; and K Bretonnel Cohen (2011). Amazon Mechanical Turk: Gold mine or coal mine? *Computational Linguistics*, vol. 37, no. 2, pp. 413–420

Fredman, Sandra; Darcy du Toit; Mark Graham; Kelle Howson; Richard Heeks; Jean-Paul van Belle; Paul Mungai; and Abigail Osiki (2020). Thinking Out of the Box: Fair Work for Platform Workers. *King's Law Journal*, vol. 31, no. 2

Gadiraju, Ujwal; Ricardo Kawase; and Stefan Dietze (2014). A taxonomy of microtasks on the web. In *Proceedings of the 25th ACM conference on Hypertext and social media - HT '14, Santiago, Chile, 1 September - 4 September 2014*. New York: ACM Press, pp. 218–223

Gadiraju, Ujwal; Jie Yang; and Alessandro Bozzon (2017). Clarity is a Worthwhile Quality: On the Role of Task Clarity in Microtask Crowdsourcing. In *Proceedings of the 28th ACM Conference on Hypertext and Social Media, Prague, Czech Republic, 4 July - 7 July 2017*. New York: ACM, HT '17. https://doi.org/10.1145/3078714.3078715

Gaikwad, Snehalkumar (Neil) S.; Mark Whiting; Karolina Ziulkoski; Alipta Ballav; Aaron Gilbee; Senadhipathige S. Niranga; Vibhor Sehgal; Jasmine Lin; Leonardy Kristianto; Angela Richmond-Fuller; Jeff Regino; Durim Morina; Nalin Chhibber; Dinesh Majeti; Sachin Sharma; Kamila Mananova; Dinesh Dhakal; William Dai; Victoria Purynova; Samarth Sandeep; Varshine Chandrakanthan; Tejas Sarma; Adam Ginzberg; Sekandar Matin; Ahmed Nasser; Rohit Nistala; Alexander Stolzoff; Kristy Milland; Vinayak Mathur; Rajan Vaish; Michael S. Bernstein; Catherine Mullings; Shirish Goyal; Dilrukshi Gamage; Christopher Diemert; Mathias Burton; and Sharon Zhou (2016). Boomerang: Rebounding the Consequences of Reputation Feedback on Crowdsourcing Platforms. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology - UIST '16, Tokyo, Japan, 16 October - 19 October 2016*. New York: ACM Press, pp. 625–637

Gellman Robert (2015) Crowdsourcing, citizen science, and the law: legal issues affecting federal agencies. Commons Lab, Woodrow Wilson International Center for Scholars

George H (2018). Neobux Review - My Experience with This PTC Ad Website. https://www.earningwithgeorge.com/neobux-review-on-this-clicking-page/.

Gleibs Ilka H (2017) Are all "research fields" equal? Rethinking practice for the use of data from crowdsourcing market places. Behavior Research Methods 49(4):1333–1342

Goel, Naman; and Boi Faltings (2019). Deep bayesian trust: A dominant and fair incentive mechanism for crowd. In *Proceedings of the AAAI Conference on Artificial Intelligence, Honolulu, Hawaii, USA, 27 January - 1 February 2019*. Palo Alto, California, USA: AAAI Press, vol. 33, pp. 1996–2003.

Gosling, Samuel D; and Winter Mason (2015). Internet research in psychology. *Annual review of psychology*, vol. 66, pp. 877–902

Graham, Mark; Jamie Woodcock; Richard Heeks; Paul Mungai; Jean-Paul Van Belle; Darcy du Toit; Sandra Fredman; Abigail Osiki; Anri van der Spuy; and Six M Silberman (2020). The Fairwork Foundation: Strategies for improving platform work in a global context. *Geoforum*, vol. 112, pp. 100–103

Gupta, Neha; David Martin; Benjamin V. Hanrahan; and Jacki O'Neill (2014). Turk-Life in India. In *Proceedings of the 18th International Conference on Supporting Group Work, Sanibel Island, Florida, USA, 9 November - 12 November 2014*. New York: ACM, pp. 1–11

Hanrahan, Benjamin V.; Anita Chen; JiaHua Ma; Ning F. Ma; Anna Squicciarini; and Saiph Savage (2021). The Expertise Involved in Deciding which HITs are Worth Doing on Amazon Mechanical Turk. *Proceedings of the ACM on Human-Computer Interaction*, vol. 5, pp. 128:1–128:23

Hara, Kotaro; Abigail Adams; Kristy Milland; Saiph Savage; Chris Callison-Burch; and Jeffrey P Bigham (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 - CHI '18, Montreal, QC, Canada, 21 April - 26 April 2018*. New York: ACM Press, pp. 1–14

Haug Matthew C (2018) Fast, Cheap, and Unethical? The Interplay of Morality and Methodology in Crowdsourced Survey Research. Review of Philosophy and Psychology 9(2):363–379

Heer, Jeffrey; and Michael Bostock (2010). Crowdsourcing Graphical Perception: Using Mechanical Turk to Assess Visualization Design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Atlanta, Georgia, USA, 10 April - 15 April 2010*. New York, NY, USA: ACM, CHI '10, pp. 203–212.

Hennink, Monique; and Bonnie N. Kaiser (2022). Sample sizes for saturation in qualitative research: A systematic review of empirical tests. *Social Science & Medicine*, vol. 292.

Ho, Chien-Ju; Shahin Jabbari; and Jennifer Wortman Vaughan (2013). Adaptive Task Assignment for Crowdsourced Classification. In *Proceedings of the 30th International Conference on Machine Learning (ICML-13), Atlanta, Georgia, USA, 17 June - 19 June 2013*. vol. 28, pp. 534–542

Ho, Chien-Ju; Aleksandrs Slivkins; Siddharth Suri; and Jennifer Wortman Vaughan (2015). Incentivizing High Quality Crowdwork. In *The International World Wide Web Conference Committee (IW3C2), Florence, Italy, 18 May - 22 May 2015*. Republic and Canton of Geneva, Switzerland: International World Wide Web Conferences Steering Committee

Horton, John J; and Richard J Zeckhauser (2010). Algorithmic wage negotiations: Applications to paid crowdsourcing. *Proceedings of CrowdConf*, vol. 4, pp. 2–5

Hossain, Mokter (2012). Users' motivation to participate in online crowdsourcing platforms. In *2012 International Conference on Innovation Management and Technology Research, Malacca, Malaysia, 21 May - 22 May 2012*. IEEE, pp. 310–315.

Howe Jeff (2006) The rise of crowdsourcing. Wired magazine 14(6):1–4

IG Metall (2017). CrowdFlower - Fair Crowd Work. http://faircrowd.work/platform/crowdflower/

Ipeirotis, P G (2010a). Analyzing the Amazon Mechanical Turk marketplace. *XRDS: Crossroads*, vol. 17, no. 2, pp. 16–21

Ipeirotis, Panagiotis G (2010b). Demographics of Mechanical Turk. *NYU Working Paper No. CEDER-10-01*

Irani, Lilly C.; and M. Six Silberman (2013). Turkopticon: interrupting worker invisibility in amazon mechanical turk. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '13, Paris, France, 27 April - 2 May 2013*. New York: ACM Press, p. 611.

Jacques, Jason T.; and Per Ola Kristensson (2019). Crowdworker Economics in the Gig Economy. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19, Glasgow, Scotland, UK, 4 May - 9 May 2019*. New York: ACM Press, pp. 1–10.

Kazai, Gabriella; Jaap Kamps; and Natasa Milic-Frayling (2013). An analysis of human factors and label accuracy in crowdsourcing relevance judgments. *Information Retrieval*, vol. 16

Kees, Jeremy; Christopher Berry; Scot Burton; and Kim Sheehan (2017). An Analysis of Data Quality: Professional Panels, Student Subject Pools, and Amazon's Mechanical Turk. *Journal of Advertising*, vol. 46, pp. 141–155

Kingsley, Sara Constance; Mary L Gray; and Siddharth Suri (2015). Accounting for market frictions and power asymmetries in online labor markets. *Policy & Internet*, vol. 7, no. 4, pp. 383–400

Klein, Amit; and Benny Pinkas (2019). DNS Cache-Based User Tracking. In *Proceedings of the NDSS Symposium 2019, San Diego, California, USA, 24 February - 27 February 2019*.

Kost, Dominique; Christian Fieseler; and Sut I Wong (2018). Finding meaning in a hopeless place? The construction of meaningfulness in digital microwork. *Computers in Human Behavior*, vol. 82, pp. 101–110

Kumar, Alok (2009). Who gambles in the stock market? *The Journal of Finance*, vol. 64, no. 4, pp. 1889–1933.

Lease, Matthew; Jessica Hullman; Jeffrey Bigham; Michael Bernstein; Juho Kim; Walter Lasecki; Saeideh Bakhshi; Tanushree Mitra; and Robert Miller (2013). Mechanical turk is not anonymous. *Available at SSRN 2228728*

Leimeister, Jan Marco; Michael Huber; Ulrich Bretschneider; and Helmut Krcmar (2009). Leveraging crowdsourcing: activation-supporting components for IT-based ideas competition. *Journal of management information systems*, vol. 26, no. 1, pp. 197–224

Litman, Leib; Jonathan Robinson; and Cheskie Rosenzweig (2015). The relationship between motivation, monetary compensation, and data quality among US-and India-based workers on Mechanical Turk. *Behavior research methods*, vol. 47, no. 2, pp. 519–528

Martin, David; Sheelagh Carpendale; Neha Gupta; Tobias Hoßfeld; Babak Naderi; Judith Redi; Ernestasia Siahaan; and Ina Wechsung (2017). Understanding the Crowd: Ethical and Practical Matters in the Academic Use of Crowdsourcing. In Daniel Archambault; Helen Purchase; and Tobias Hoßfeld (eds.), *Evaluation in the Crowd. Crowdsourcing and Human-Centered Experiments*, Cham: Springer International Publishing, vol. 10264

Martin, David; Benjamin V. Hanrahan; Jacki O'Neill; and Neha Gupta (2014). Being A Turker. In *Proceedings of the 17th ACM conference on Computer supported cooperative*

work & social computing, Baltimore, Maryland, USA, 15 February - 19 February 2014. New York, NY, USA: Association for Computing Machinery, CSCW '14, pp. 224–235

Martin, David; Jacki O'Neill; Neha Gupta; and Benjamin V Hanrahan (2016). Turking in a Global Labour Market. *Computer Supported Cooperative Work (CSCW)*, vol. 25, no. 1, pp. 39–77

Mason, Winter; and Duncan J Watts (2009). Financial Incentives and the "Performance of Crowds". In *Proceedings of the ACM SIGKDD workshop on human computation, Paris, France, 28 June 2009*. New York: ACM Press, pp. 77–85.

McInnis, Brian; Dan Cosley; Chaebong Nam; and Gilly Leshed (2016). Taking a HIT: Designing around Rejection, Mistrust, Risk, and Workers' Experiences in Amazon Mechanical Turk. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16), San Jose, California, USA, 7 May - 12 May 2016*. New York: ACM Press, pp. 2271–2282

Novotnỳ, Filip (2018). Are Cryptocurrencies Gambling Asset? (Unpublished Bachelor's dissertation). *Univerzita Karlova, Fakulta sociálních věd*.

Paul, Aplar; and Osterbrink Lars (2018). Antecedents of Perceived Fairness in Pay for Microtask Crowdwork. In *Twenty-Sixth European Conference on Information Systems (ECIS2018), Portsmouth, United Kingdom, 23 Jun - 28 Jun 2018*. Atlanta, Georgia, USA: Association for Information Systems.

Penman, Andrew (2019). How to turn £8,000 into £4 million: run a pyramid scam like Traffic Monsoon.

Petrović, Nataša; Gabriel Moyà-Alcover; Javier Varona; and Antoni Jaume-i Capó (2020). Crowdsourcing human-based computation for medical image analysis: A systematic literature review. *Health Informatics Journal*, vol. 26, pp. 2446–2469

Qiu, Chenxi; Anna Squicciarini; and Benjamin Hanrahan (2019). Incentivizing Distributive Fairness for Crowdsourcing Workers. In *Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems, Montreal, QC, Canada, 13 May - 17 May 2019*. Richland, SC: International Foundation for Autonomous Agents and Multiagent Systems, pp. 404–412

Rea, Stephen C.; Hanzelle Kleeman; Qin Zhu; Benjamin Gilbert; and Chuan Yue (2020). Crowdsourcing as a Tool for Research: Methodological, Fair, and Political Considerations. *Bulletin of Science, Technology & Society*, vol. 40, pp. 40–53

Ross, Joel; Lilly Irani; M. Six Silberman; Andrew Zaldivar; and Bill Tomlinson (2010). Who are the crowdworkers?: shifting demographics in mechanical turk. In *Proceedings of the 28th of the international conference extended abstracts on Human factors in computing systems (CHI EA '10), Atlanta, Georgia, USA, 10 April - 15 April 2010*. New York: ACM Press, pp. 2863–2872

Rowlands, Terry; Neal Waddell; and Bernard McKenna (2016). Are we there yet? A technique to determine theoretical saturation. *Journal of Computer Information Systems*, vol. 56, no. 1

Saito, Susumu; Chun-Wei Chiang; Saiph Savage; Teppei Nakano; Tetsunori Kobayashi; and Jeffrey P. Bigham (2019). TurkScanner: Predicting the Hourly Wage of Microtasks. In *Proceedings of the World Wide Web Conference (WWW '19), San Francisco, CA, USA, 13 May - 17 May 2019*. New York: Association for Computing Machinery

Salehi, Niloufar; Lilly C Irani; Michael S Bernstein; Ali Alkhatib; Eva Ogbe; and Kristy Milland (2015). We are Dynamo: Overcoming stalling and friction in collective action for crowd workers. In *Proceedings of the 33rd annual ACM conference on human factors in computing systems, Seoul, Republic of Korea, April 18 - April 23 2015*. New York: ACM, pp. 1621–1630

Schmidt, Florian Alexander (2013). The good, the bad and the ugly: Why crowdsourcing needs ethics. In *2013 International Conference on Cloud and Green Computing, Karlsruhe, Germany, 30 September - 02 October 2013*. New York: IEEE, pp. 531–535.

Shmueli, Boaz; Jan Fell; Soumya Ray; and Lun-Wei Ku (2021). Beyond fair pay: Ethical implications of NLP crowdsourcing. arXiv preprint 2104.10097

Silberman, M Six; Lilly Irani; and Joel Ross (2010). Ethics and tactics of professional crowdwork. *XRDS: Crossroads, The ACM Magazine for Students*, vol. 17, no. 2, pp. 39–43

Silberman, M Six; Bill Tomlinson; Rochelle LaPlante; Joel Ross; Lilly Irani; and Andrew Zaldivar (2018a). Responsible research with crowds: pay crowdworkers at least minimum wage. *Communications of the ACM*, vol. 61, no. 3, pp. 39–41

Silberman, M.S.; B. Tomlinson; R. LaPlante; J. Ross; L. Irani; and A. Zaldivar (2018b). Responsible research with crowds: Pay crowdworkers at least minimum wage. *Communications of the AMC*, vol. 61, no. 3, pp. 39–41

Spatharioti, Sofia Eleni; Sofia Eleni Spatharioti; Rebecca Govoni; Jennifer S Carrera; Sara Wylie; and Seth Cooper (2017). A Required Work Payment Scheme for Crowdsourced Disaster Response: Worker Performance and Motivations. In *Proceedings of the 14th International Conference on Information Systems for Crisis Response And Management (ISCRAM '17), Albi, Occitanie Pyrénées-Méditerranée, France, 21 May - 24 May 2017*. pp. 475–488

Stewart, Neil; Jesse Chandler; and Gabriele Paolacci (2017). Crowdsourcing Samples in Cognitive Science. *Trends in Cognitive Sciences*, vol. 21, pp. 736–748

Sun, Chong; Narasimhan Rampalli; Frank Yang; and AnHai Doan (2014). Chimera: large-scale classification using machine learning, rules, and crowdsourcing. *Proceedings of the VLDB Endowment*, vol. 7, pp. 1529–1540

Tassinari, Arianna; and Vincenzo Maccarrone (2020). Riders on the storm: Workplace solidarity among gig economy couriers in Italy and the UK. *Work, Employment and Society*, vol. 34, no. 1, pp. 35–54.

Vayena, Effy; Marcel Salathé; Lawrence C Madoff; and John S Brownstein (2015). Ethical challenges of big data in public health

Vougiouklis, Pavlos; Eddy Maddalena; Jonathon Hare; and Elena Simperl (2020). Point at the Triple: Generation of Text Summaries from Knowledge Base Triples. *Journal of Artificial Intelligence Research*, vol. 69, pp. 1–31

Wan, Xiangpeng; Hakim Ghazzai; and Yehia Massoud (2019). Mobile Crowdsourcing for Intelligent Transportation Systems: Real-Time Navigation in Urban Areas. *IEEE Access*, vol. 7, pp. 136995–137009

Whiting, Mark E.; Grant Hugh; and Michael S. Bernstein (2019). Fair Work: Crowd Work Minimum Wage with One Line of Code. In *Proceedings of the Seventh AAAI Conference*

on Human Computation and Crowdsourcing, Washington, USA, 28 October - 30 October 2019. Vancouver, British Columbia, Canada: PKP Publishing Services Network

Williamson, Vanessa (2016). On the Ethics of Crowdsourced Research. *PS - Political Science and Politics*, vol. 49, no. 1, pp. 77–81.

Ye, Teng; Sangseok You; and Lionel P Robert (2017). When Does More Money Work? Examining the Role of Perceived Fairness in Pay on the Performance Quality of Crowdworkers. In *Proceedings of the Eleventh International AAAI Conference on Web and Social Media, Montreal, Quebec, Canada, 15 May - 18 2017*. Palo Alto, California, USA: AAAI Press, Icwsm, pp. 327–336

Publisher's Note