Understanding Indian Crowdworkers

Neha Gupta, Andy Crabtree, Tom Rodden
School of Computer Science
University of Nottingham
{First name.Last name}@nottingham.ac.uk

David Martin, Jacki O'Neill
Xerox Research Centre Europe
6 chemin de Maupertuis, Grenoble France
{david.martin, jacki.oneill}@xrce.xerox.com

ABSTRACT
Amazon Mechanical Turk, the microtask-based platform is a huge opportunity for all its stakeholders - a means of livelihood for workers, quick and inexpensive way of getting work done for enterprises. In this paper, we explore who these workers are, what "work" means and the efforts workers take to do this work through an ethnographic study of Indian crowdworkers. To provide further analytic understanding of our findings we draw a comparison with our other studies of US based Turkers.

Author Keywords
Crowdsourcing, crowdworkers, Amazon Mechanical Turk

INTRODUCTION
The growth of internet and mobile-based communication provides workforces around the world with opportunities unheard of a decade ago, for e.g. the invention of crowdsourcing platforms. Jeffe Howe of Wired Magazine defined crowdsourcing as *"the act of a company or institution taking a function once performed by employees and outsourcing it to an undefined (and generally large) network of people in the form of an open call"*. One major type of crowdsourcing is microtask crowdsourcing – where many workers are used to complete large volumes of typically small and relatively low skilled tasks. From an organizational perspective, micro-task crowdsourcing has potential to function as business process crowdsourcing (BPC): many medium and large organisations are likely to have a variety of work such as data entry processing and translation that might be suitable for this type of crowdsourcing, currently done in-house or outsourced. One of the key challenges for BPC is that the work which might be crowdsourced is typically ongoing and high volume, often with tight turnaround times and quality metrics. For BPC to be viable, a scalable and quality workforce is required [1]. Whilst much of the crowdsourcing rhetoric takes this as given, we believe that it should be seriously considered and researched.

Whilst there are already a number of microtask platforms, of which Amazon Mechanical Turk (AMT) is perhaps the best known, it is not a simple matter to take business processes and delegate them either fully or partially out to the crowd [1]. In this paper, we present the preliminary findings of a study of Indian crowdworkers working primarily on AMT. The aim of this study was to get an in-depth insight into their working life – how they organized their activities (both work and non-work), what tools and technologies they used, the high and low points of crowdsourcing. We believe that taking such a view will enable us to design technologies and workflows of benefit to both organizations and crowdworkers. We also draw a comparison with US-based workers from other studies we are undertaking. This paper contributes to the body of research, which aims to elucidate and solve the various socio-technical, and ethical challenges that need to be addressed to make BPC viable [1].

RELATED WORK
Here, we describe relevant existing research focusing on crowdworkers, to which this paper contributes. The majority focus on 'Turkers' (crowdworkers working through AMT), partly because this is one of the most widely used platforms and partly because it is easy to access. Various surveys have been conducted into Turkers and their motivations. Ross et al.'s survey on AMT gives us the distribution of the crowdworkers as 56% U.S-based, 36% India-based and small percentage from countries like Canada, the U.K., and the Philippines [2]. These surveys also give us a preview of the demographics: 21-25 year olds, predominantly male, working to earn money or kill time, from the comfort of one's home [3].

Whilst these surveys give us a quantitative overview of who the crowdworkers are, they do not provide the in-depth nuanced view, which can enable us to design for them. For this, we need to turn to qualitative studies; like Irani and Silberman's studies, where they got Turkers to produce haikus and compile a workers' bill of rights [4, 5]. The responses given by Turkers enabled us to understand Turkers' issues such as unfair rejection of work, delay in payment, inapt wages and lack of communication from AMT and Requesters (those who provide the work). Research has also shown how crowdworkers evaluate tasks and Requesters according to criteria such as fairness of pay, responsiveness and so on [6, 7]. We have also been studying Turkers by analyzing publicly displayed posts on forums such as Turker Nation, Turk Alert, and Turkopticon, to understand the Turker perspective and reasoning about

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
CSCW '14, February 15–19, 2014, Baltimore, Maryland, USA.
Copyright 2014 ACM xxx-x-xxxx-xxxx-x/xx/xx...$15.00.

their work, Turker-Requester relationships and other practical and ethical matters [8].

Research has also evaluated design of platforms, for e.g. a study of low-income workers in India by Khanna et al. that revealed barriers preventing low-income worker groups from working in a marketplace like AMT. Using surveys and observations, they found the general barriers to be: complex task instructions and user interface, navigational and sequencing issues and the difference in culture. Based on these findings, they designed an interface with improved instructions, video tutorials and language localization; and tested it with users, baring a significant increase in their quality of work [9].

These studies cover a wide area, but we still only have a limited view of "who is behind the crowd" and what it means to be a crowdworker. In our study, we focus on AMT's second largest workforce – Indian Turkers. We set out to learn about their working habits, and the temporal, spatial and operational aspects of this work. What motivates them? What ethical issues do they face? What are their capabilities and what did they have to learn to become crowdworkers? What technologies do they use? How do they deal with technology-related challenges? What investments do they make? How can we better design the platform or related services to aid crowdwork? To answer some of these questions, and to understand what 'work' means, we studied crowdworkers in India over 5 months in the summer of 2013, using a number of methods. In this paper, we give a brief overview of the study, methods used to collect data, the challenges and potential of crowdwork.

SETTING AND METHOD

The subjects of this study were Indian crowdworkers, living and working from India. The majority of these workers use AMT regularly to find work. AMT was our main channel of recruitment and our prime focus to understand 'crowdwork'. However, we do not say, that Mechanical Turk is synonymous with 'crowdwork'.

We used a number of methods to collect data – starting with a short online survey of 25 questions. We used the survey to collect basic demographic and background information and as a way of making contact with possible participants. In addition to the surveys, we conducted open-ended qualitative interviews through Skype, phone and conferencing facilities, as well as face-to-face. During the interviews, we asked the participants to walk us through the various activities they undertake during crowdwork. Finally, we conducted observations with participants in their respective workplaces (typically their family homes or hostels). Since crowdworkers are distributed and anonymous, a research challenge was how to access them. Since this research was undertaken whilst on an internship to Xerox's Indian research centre (XRCI), we started with a list of crowdworkers who had previously worked for Xerox and who had chosen to contact the company personally and who were thus no longer anonymous. Typically, they had contacted Xerox for a clarification, to make a complaint, or to ask for more work. We contacted 69 out of 100 workers, the other 31 not being based in India were outside the scope of our study. Whilst we used their email addresses for the initial contact, all responses to the survey were anonymous. Participants were invited to complete the survey or take part in an interview or observation. They were paid via a HIT on AMT. A survey HIT paid $2.50, an interview HIT paid $7.00 and an observation HIT paid $20.00. Any extra effort or time spent with us was compensated for, by a bonus.

Once our HITs were live, we also started receiving emails from other workers who wanted to participate in the study. Thus, we conducted 68 surveys, 27 virtual interviews, 2 in-person interviews, and 12 observations in 3 locations. We have now begun to transcribe and analyze this data and have some interesting initial insights to share.

PRELIMINARY FINDINGS

Whilst this is a work in progress, we want to share some preliminary insights about 'workers from India', which we wish to develop further in order to share with the workshop participants. A brief look at our survey, gives us as a basic snapshot of demographics: the majority of our survey takers were male, in the age-group 24 – 31 years, predominantly from the suburbs of south India. More than half of them had started crowdwork only 1-3 years ago. Some selected initial insights from our data are as below.

"Stages" in Crowdworking (based on experience)

Like in any workplace, there is a hierarchy in the levels of experience and understanding among workers. We offer a classification of them as- novices, intermediate, or expert level workers. This is based partially on self-identification but also on our understanding of how different crowd-workers can be basically fitted into these groupings.

Novices: When a worker joins a crowdsourced platform and starts exploring how things work and tries different ways of finding jobs and making money. This stage is between 0-9 months of joining a crowdsourced platform.

Intermediate worker: When a worker has learnt the basics, acquired speed and a taste for the work, has made favorites, in terms of jobs, co-workers and Requesters and knows where to seek advice. This stage is between 1-2 years of joining a crowdsourced platform. These workers usually combine crowdwork with other conventional forms of work to generate income.

Expert worker: When a worker is "settled" with crowdwork, generates a steady flow of income, has set up a customer base (e.g. gets work from Requesters directly through emails). Some act as a broker/mentor for other (mostly novice) crowdworkers. They interact with and assist them regularly through social networking sites or personal communication such as phone, Skype or email.

Workers become experts after at least 2-3 years of work; crowdwork is generally their main source of income.

Workers' motivations, expectations and outputs were connected to the number of hours they were able to put into crowdwork every day or week. Workers spent as little as 2 hours to 40 hours (equivalent to a regular full-time job) every week. We also found that these work hours were not strict or uniform. People worked beyond these limits based on availability of tasks, free time and interest. For e.g. some users worked close to 56 hours every week as they did not want to break their flow of work.

Adapting to U.S. time

Most of the workers we interviewed, noticed that HITs were being created on AMT, in batches (bulk), during late hours of the night, presumably morning or 'work hours' in the U.S. We found that our crowdworkers were not shy of hard work; they spent countless hours late at night working on these HITs, if they were available in good quantities. At a stretch people liked to work on 100s, if not 1000s of such HITs.

Vignette 1:
A recent college graduate from Hyderabad, who spends most of his time working on AMT HITs says "I work from 7 – 7.30 pm in the evening every day, break for dinner with the family at around 9pm. I re-start work at 10 pm after some rest, and continue working until 1- 2 am on most nights. On some nights I get lucky and new batch HITs are uploaded. To make the most of it, I work till 3 or 4 am, till my eyes feel tired. Batch HITs are consumed very quickly.. imagine people are working on them all round the world. They are available tonight, gone tomorrow."

There are two sides to this adaptation. One is where the workers are happy to adapt their lifestyles to allow them to work at these late hours, following "late to bed, late to rise" principle and count on the support of their families to follow this regime. This support includes preparing meals for the crowdworkers and relieving them of their daily chores and household duties. Heavy (40+ hours per week) users of AMT admitted that they would not be able to put in their regular hours if not for the support of their families. On the other hand, it is not always desirable to work odd hours and compromises need to be made in other aspects of life as we see in the example below.

Vignette 2:
A part-time college student from Hyderabad, living in a large extended family system says "There is no work in the mornings. I have tried to find work, but there is nothing available… Right now, I run errands and help family with their work in the mornings and also, attend classes. I work at night until midnight or 1 am, but I would prefer not to. I would like to work in the mornings, have some spare time and proper rest at night, but the jobs only appear at night. In the morning, there is nothing for us."

Most workers are engaged in multiple activities everyday - complete with elements of familial chores, education or day job along with crowdwork. They are ready to work odd hours if their favorite jobs were available for decent money, if they felt the work was "worth their time and effort" but did not want to 'slog' for little returns, however sometimes this is unavoidable.

Vignette 3:
A final year engineering student from Madurai who wants to earn some cash before he lands a full-time job says "I work on AMT as it helps me earn some pocket money till I find a stable job. I work in the mornings and evenings, but not late nights due to my studies. The tasks in the morning are very low-paying. I have to spend hours working, to earn a few cents. In the evenings, you need to be lucky to find well-paying jobs. You don't have a choice but to take what you can find.

We see from these examples that jobs are not available round-the-clock, and that it is not always possible for workers to follow the schedules of the Requesters, which leaves them feeling dissatisfied.

Reputation and status

Many crowdworkers have listed AMT as their current employer on their Facebook profiles. The highlight of their work is that they are working for "Americans". They fathom jobs are being provided by an "American company" (in this case - AMT), where they work for "American clients" (Requesters) and are paid "American money".

Vignette 4:
A young data entry operator cum part-time student from Kerala, who is currently working in Bangalore says "In my previous job as an accountant, I met a few American people and found them to be very open. I like and respect them since then, they have a very high place in my mind... this company [AMT] gives me a chance to work for American clients and get paid in American money [dollar]), I am very proud of that. I love interacting with them via emails."

It gives workers a lot of satisfaction and pride in sharing the details of their "American" jobs with friends and family. Workers have to work on building a status and profile (of approval rates and number of HITs worked upon) – to catch reasonably-paying jobs and to sustain Requesters. They share this information on networking sites and forums happily, to get self-satisfaction and to inspire new workers.

DISCUSSION

There are a number of new findings already coming out of our study, which we believe, are of interest.

Pay and Status

There are differences and similarities between US and Indian crowdworkers. Whilst in the US, microtask crowdsourcing is often considered as rather low status work, largely because of its poor pay and relatively low

skill nature; whereas, in India, a certain amount of status is conferred on the work *because* it largely comes from American clients. US-based crowdworkers do take pride in their work, skills and expertise but bemoan how little of this is recognized outside their community. Conversely, crowd work has a better general status in India – it is not a 'dirty little secret' – shown by the fact that it is more generally discussed in families and society, and is not differentiated as a different/special type of work. There is a difference in cost of living to account for, when comparing wages between Indian and US Turkers even when performing the same work. The benefits, in terms of the prestige of working for a US Company and development of skills on your CV that work for Indian crowdworkers, are not so clear with US-based workers. However, it is interesting to note that even for Indian Turkers low wages are an issue; crowd working is still seen as somewhat of a stopgap and is less desirable than a 'regular' job.

Batch Scheduling

Time differences mean that Indian Turkers have to work outside of normal working hours. Whilst this can be beneficial for those who are working and studying during the day, it is not without its downsides, just as shift work in traditional jobs can be difficult to manage. This also raises some intriguing questions about batch scheduling and knowledge of who your workers are. US companies might even target the timing of their release of batches in order to better coincide with normal working hours in India. Of course, if your batches are being completed on time to a high standard this is not a problem. However, if there are difficulties, which may well be multiplied in a BPC situation it makes sense to think more about how batches are scheduled and released in relation to fluctuations in worker availability. Having knowledge of and some form of relationships with your workers can help with this.

Career Path/Life-Cycle

In our study of US-based crowdworkers, we understood that they self-identified roughly as being in one of two categories 'newbies' or 'experienced' workers. Newbies were characterized by the fact that they were learning the ropes, understanding the marketplace, who bad and good Requesters were, working on their skills and technology set-up and doing often lower paid bulk work as a means of getting a good HIT count and approval ratings. Experienced workers, as in this study, had gone through the basic learning curve and now had regular Requesters and types of HIT they worked on, through which they received acceptable wages. Comparatively we can see something similar occurring in India. An interesting thing that comes out of this is that, it would be good to know even more about the career path and life-cycles of crowdworkers. What sort of activities relate to the different stages? What are the drop-out rates? What is the average length of career in crowdsourcing? What are the reasons that people enter and exit? Businesses that understand these issues could help bring new workers on-board and support a career trajectory. They might also be better able to retain valuable workers.

CONCLUSION

Until now microtask platforms and workflows have been primarily designed with the Requester or organization in mind. For BPC this needs to change – design of tools, technologies and workflows needs to take into account the requirements and concerns of the crowdworkers. Since BPC needs to be sustainable over the long-term, and since an experienced workforce is likely to benefit both the worker and the organization even in very low-skilled work [1], organizations must think of their workers' concerns, issues and well-being, even if they don't face them directly and work from a remote location. Currently we are analyzing the data collected during our fieldwork. We hope we have been able to provide an indication of the type of material we are developing. We will soon be able to provide a more detailed response to the many questions we have raised and in doing so, an understanding of what it means to be an Indian crowdworker and provide comparisons between workers from India and US.

REFERENCES

1. O'Neill J., Martin D. Relationship-based Business Process Crowdsourcing? In Proc. of IFIP Conference on Human-Computer Interaction (2013) 429-446.

2. Ross, J., Irani, L., Silberman, M. S., Zaldivar, A., and Tomlinson, B. Who are the crowdworkers?: shifting demographics in mechanical turk. In Proc. of CHI EA '10, ACM (2010), 2863–2872.

3. Ipeirotis, P. Demographics of Mechanical Turk. CeDER Working Papers No. CeDER-10-01. (2010) http://hdl.handle.net/2451/29585

4. Silberman, M. S. What's fair? Rational action and its residuals in an electronic market. Unpublished manuscript, (2010)

5. Silberman, M., Ross, J., Irani, L. and Tomlinson, B. Sellers' problems in human computation markets. In Proc. of HCOMP '10. ACM Press (2010), 18-21.

6. Bederson, B. B. and Quinn, A. J. Web workers unite!addressing challenges of online laborers. In Proc. CHI EA '11. ACM (2011), 97-106.

7. Felstiner, A. Working the crowd: employment and labor law in the crowdsourcing industry. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1593853 (2011).

8. Martin, D., Hanrahan, B. J., O'Neill, J., Gupta, N., Being a Turker. To appear in CSCW 2014 (2014)

9. Khanna, S., Ratan, A., Davis, J., Thies, W.: Evaluating and Improving the Usability of Mechanical Turk for Low-Income Workers in India. In: ACM DEV 2010 (2010)