

# Who are the Turkers? Worker Demographics in Amazon Mechanical Turk

Joel Ross, Andrew Zaldivar, Lilly Irani, Bill Tomlinson

Department of Informatics

University of California, Irvine, USA

{jwross, azaldiva, lirani, wmt} @ uci.edu

## ABSTRACT

Amazon Mechanical Turk (MTurk) is a crowdsourcing system in which tasks are distributed to a population of thousands of anonymous workers for completion. This system is becoming increasingly popular with researchers and developers. In this paper, we survey MTurk workers about their demographic make-up and usage behavior. We find that this population is diverse across several notable demographic dimensions such as age, gender, and income, but is not precisely representative of the U.S. as a whole. Indeed, certain homogeneous aspects of the population, such as education level and nationality, may impose limits on the appropriateness of Turkers as a target community for some interventions or research areas. An awareness of the demographics and behaviors of MTurk workers is important for understanding the capabilities and potential side effects of using this system.

## Author Keywords

Mechanical Turk, demographics, user surveys, crowdsourcing

## ACM Classification Keywords

H.5.3 Group and Organization Interfaces: Computer-supported cooperative work; H.5.3 Group and Organization Interfaces: Web-based interaction

## INTRODUCTION

Amazon Mechanical Turk<sup>1</sup> (MTurk) is an online crowdsourcing [5] system that allows users to distribute work to a large number of workers. This work is broken down into simple, one-time tasks which workers are paid to complete. Such tasks are frequently those that are difficult for computers and yet simple for humans (e.g., image labeling), creating a kind of "artificial artificial intelligence" [2]. Requesters create Human Intelligence Tasks, or HITs,

specifying the amount paid for the job's completion—usually ranging from as little as \$0.01 for a quick task up to a few dollars for more involved jobs, such as transcribing audio clips. Workers (or "Turkers") who log into the MTurk website are able to pick and choose which tasks they perform (after previewing the HIT), thus creating a micro-task marketplace [8]. Launched in 2005, Amazon reports that the system has now more than 200,000 workers registered, and there are about 50,000-100,000 HITs to work on at any given time.

MTurk has become increasingly popular as a tool for research, being used for performing user studies [4,8], image labeling [11], natural language processing [10], and relevance evaluation [1]. However, little research has considered the almost entirely anonymous users who are actually performing this work—only a randomly generated ID number represents workers to requesters—and most such previous work has been informal (e.g., [6]). Distributed crowdsourcing systems often rely on this anonymity: for example, the ESP Game [13] crowdsources image tagging, using players' mutual anonymity to help ensure that they describe an image in terms of its contents, rather than based on their knowledge of each other. Even Wikipedia, the online encyclopedia that anyone can edit, benefits from distributing its creation to anonymous users as well as registered editors [9].

Nevertheless, it is important to know the context of anonymous users in these systems in order to understand how they may affect and be affected by the research performed. For example, a population made up entirely of similarly situated subjects may give more homogenous responses to a usability questionnaire than a more diverse crowd. Furthermore, obscuring worker identity may dissociate those users from requesters and potentially contribute to workers being exploited: because workers are decontextualized, requesters may be more likely to offer lower, unfair prices on HITs, or even refuse to pay for work performed. Thus understanding and even humanizing Turkers is vital to the overall health of the system and its users, and may even increase the effectiveness with which researchers can engage and interact with this population.

## METHOD

In order to begin to understand the kinds of people who

<sup>1</sup> MTurk can be accessed at <http://www.mturk.com>.

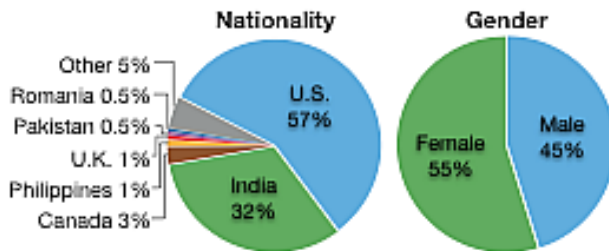


Figure 1. Nationality and gender of MTurk workers.

work in MTurk, we created an exploratory survey to gather simple demographic data about the worker population. We also asked users about their Turking habits, such as how often and what kinds of HITs they complete.

The survey was delivered to users in the form of a HIT on the MTurk website—workers could choose to take the survey and be compensated through the system, as with any other task. Workers accepted the HIT and were redirected to a web-based survey, at the end of which they were given a confirmation code to enter back into MTurk. Note that because the external link to the survey was available in the HIT preview, workers were able to view (and even complete) the survey before actually accepting the HIT. The survey was made available at 3pm on a Wednesday, and left open for a full seven days. Workers were compensated \$0.10 for their time (the median completion time was just over 2 minutes).

Previous research [1,8,11] has described the potential unreliability of MTurk workers. While this literature has suggested the use of either qualification pre-tests and/or explicitly verifiable questions, these were not appropriate for the current study—qualification tests would exclude a portion of the Turker population, and there was no way of verifying user demographics. Indeed, although MTurk HITs default to being answerable only by users with at least a 95% approval rating (meaning that 95% of the worker's submitted HITs have been approved by the requester of the work), we removed all restrictions for accepting the survey—allowing any worker to complete the HIT—in order to reach as broad a user population as possible.

Thus there is the possibility some respondents may have given purposefully false answers. Collecting responses through a HIT creates the possibility of demand characteristics (where subjects change behavior in response to being measured): workers may have shaped their responses based on what they believed we wanted to hear in order to be assured of being paid for their time. In addition, the method of delivering this survey means that respondents were self-selecting—thus our survey may be biased towards Turkers who enjoy taking surveys and are willing to provide information about themselves, rather than reflecting the worker population as a whole. Nevertheless, we believe this sample is still large enough to provide meaningful insight into the demographics of MTurk workers.

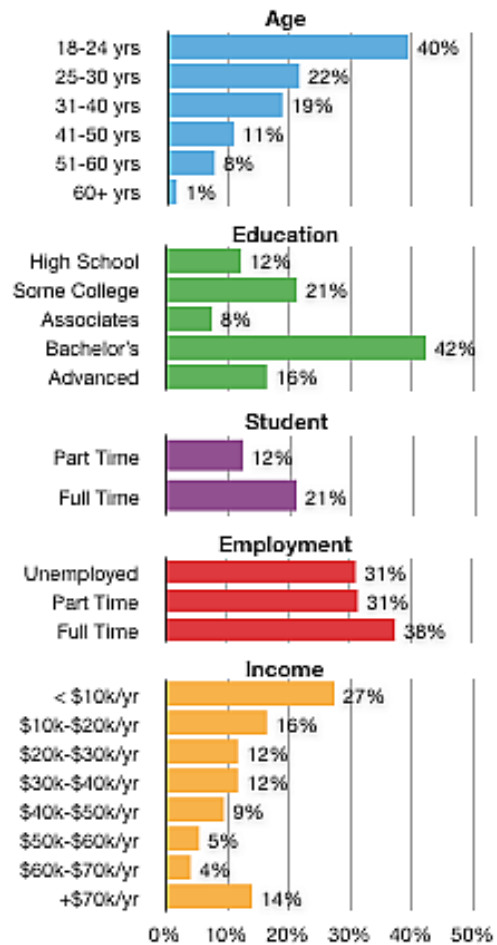
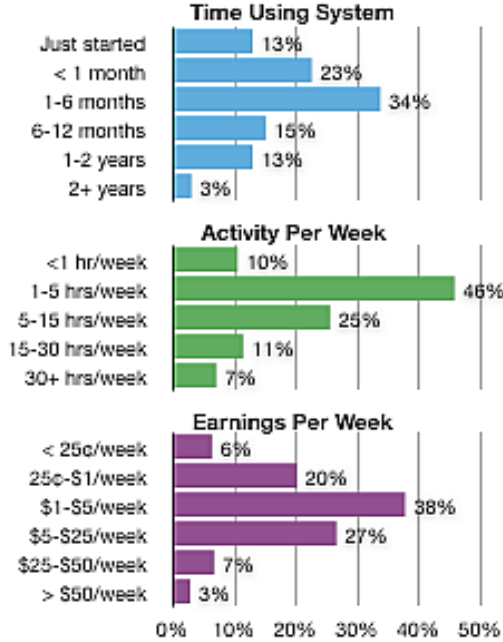


Figure 2. Demographics of MTurk workers.

## SURVEY RESULTS

The survey was available as a HIT for a full week. During this time, 573 people submitted valid completed surveys (three respondents reported to be under 18, and are not included in the results presented here).

According to our survey results, 57% of MTurk workers are from the United States, while 32% are from India—the remaining respondents are from countries ranging from Australia to Ukraine. Respondents reported an average age of 31 years old (min 18, max 71, median 27), and the majority of respondents (55%) are female. More than half (66%) of respondents have a college or advanced degree, and 33% are either full- or part-time students. While 38% of respondents are employed full-time, nearly a third (31%) are currently unemployed. The median annual reported income was between \$20,000 and \$30,000. These demographics reveal a significantly international and highly educated population, though one with lower levels of employment and income. Indeed, 18% of Turkers reported sometimes or always relying on MTurk to "make basic ends meet" (Figure 4). While only a minority of workers rely on the pay earned from completing HITs, they still make up a significant percentage of the Turker population.



**Figure 3. Reported system usage statistics**

In general, this survey suggests that MTurk users are relatively new to the system, with most (69%) Turkling for less than 6 months. The average respondent spends about 8 hours per week performing HITs and earns around \$10 during that time (see Figure 3). Turkers earn less than U.S. minimum wage, but are not generally completing HITs in such a way as to constitute a full-time job: Turkling seems to be a part-time activity for most users. This effect could be a result of the relatively low levels of pay—Turkers might Turk more if they made more money from it. Furthermore, this level of payment may still be substantial to some—the less than minimum-wage Turk earnings may mean the difference between paying a bill or not [7].

The type of HIT respondents reported completing most frequently is detailed in Table 1. Notably, a majority (52.9%) of users reported complete surveys more often than other types of HITs (anecdotally, surveys do not appear to be more common than other types of HITs). This supports the idea that our survey saw some amount of self-selecting—Turkers who often completed surveys were more likely to respond to our questionnaire. This may have biased our population sample—the demographics presented here may more represent workers who do survey-type tasks than audio transcription tasks, for example.

However, a set of chi-square tests comparing each factor from those who mostly complete surveys and those who mostly perform other tasks finds that this difference is not significant across many of the factors examined in this survey. Age, gender, education, and employment all have similar distributions no matter which task type respondents prefer. But there was a strongly significant difference in

Finding information (ex. contact information on a website)	12.2%
Comparing if items/locations/etc. are the same	6.6%
Image labeling/tagging (also locating information in an image)	6.5%
Image filtering/moderation	3.5%
Writing tasks (articles, summaries, reviews, editing/rewriting)	9.1%
Audio transcription	1.4%
Surveys, polls, questionnaires	52.9%
Data entry	4.5%
Other	3.3%

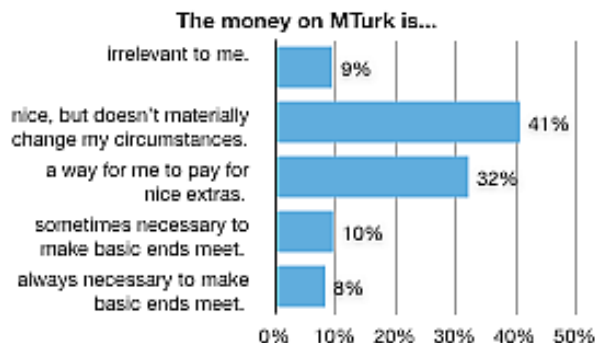
**Table 1. Task completed most often on MTurk.**

other categories: Turkers who answer surveys work less in MTurk, earn less in MTurk, have higher incomes, rely on MTurk less, and are more likely to be from the US. This difference is like a result of surveys often being less lucrative on MTurk—workers looking primarily to maximize their income are likely to find other HITs with a higher pay rate. Yet because our survey paid more than the reported average earnings (\$3/hr vs. \$1.25/hr), we likely attracted a wider sample population. Thus many of the demographics presented here are still seem to be representative of the system as a whole.

## DISCUSSION

This survey suggests that MTurk workers make up a diverse group, including a range of ages, education levels, and socio-economic strata, though primarily from highly industrialized societies. The requirements of an Internet connection and English language skills restrict the potential range of MTurk workers. The results somewhat matched those in [6], as we also find that "the profile of the typical Turker is not of a person that completes tasks for a living in a developing country." However, we find a much more international population than in [6], with a greater number of users from India. This may indicate that MTurk has gained substantially more international members in the 18 months since this previous survey was performed.

However, unlike [6], we find this diverse population to be only somewhat representative of the U.S. population as a whole. Compared to the entire population [12], MTurk workers from the U.S. are younger (median age 30 vs. 36.6), much more highly educated (63% vs. 25% with college degrees), and include a significantly greater number of female members (69% vs. 51%). Turkers also tend towards lower levels of annual income. However, these demographics somewhat resemble the "elite Internet users" of iStockPhoto (a similar crowdsourcing system for creating stock images) [3], though Turkers are less homogenous and tend to have lower levels of income. So while the MTurk population may perhaps be representative of the U.S. *internet-using* population, it cannot truly be seen to be a microcosm of the country as a whole.



**Figure 4. Reported reliance on money earned on MTurk.**

Turkers from India, on the other hand, are much more often male (69%), even younger (median age 25), and even more highly educated (74% have a college degree or higher). Although they have slightly lower unemployment levels than the sample as a whole (26% vs. 31%), they are almost twice as likely to report themselves as relying on the income from MTurk (29% vs. 18%). This suggests that MTurk may have two different groups of users simply based on nationality. On the other hand, some or all of this difference may be a function of demand characteristics or even cultural values—Indian Turkers may seek to portray themselves through surveys differently than U.S. Turkers.

Thus these results suggest that Turkers are not precisely representative of the U.S. population, and so should not necessarily be treated as such. Indeed, the homogeneity of Turkers' education levels and nationality may limit the appropriateness of MTurk as a target community for some research and interventions. For example, user surveys conducted through MTurk may not produce equivalent results as more traditional (and more expensive) methods of polling. Nevertheless, the ease and low cost of performing such studies means MTurk should not be excluded as a tool for research, and may in fact be more appropriate for many Internet-based HCI applications and interventions than a broader demographic.

Furthermore, we hope that this survey helps to give some insight into the humans who enable this "artificial artificial intelligence." A significant portion of these workers rely on the pennies they earn performing human computation, some treating the system as a full-time (if low-paying) job. Being aware of the circumstances of these anonymous workers is important if CHI researchers and other system developers are to act responsibly toward them as we conduct our research. MTurk may present itself as a form of AI, but behind it are real and potentially vulnerable people.

## CONCLUSION

MTurk workers are a diverse group of users, though not demographically representative of the broader U.S. population. Nevertheless, Turkers may provide a viable user base for a variety of CHI research efforts. By providing

demographic information and usage patterns of MTurk workers, we seek to enable more researchers to engage this community in various surveys, online prototypes, and other interventions, while at the same time discerning the actual people in the cloud of labor. Our future work will continue to explore the real people that enable this system, emphasizing the human element in human computation.

## ACKNOWLEDGMENTS

The authors thank Six Silberman, Gloria Mark, and the Social Code Group. This research has been supported by a Sloan Research Fellowship, the Donald Bren School of Information and Computer Sciences, and the California Institute for Telecommunications and Information Technology (Calit2).

## REFERENCES

- Alonso, O., Rose, D.E., and Stewart, B. Crowdsourcing for relevance evaluation. *SIGIR Forum* 42, 2 (2008), 9-15.
- Barr, J. and Cabrera, L.F. AI gets a brain. *Queue* 4, 4 (2006), 24-29.
- Brabham, D.C. Moving the crowd at iStockphoto: The composition of the crowd and motivations for participation in a crowdsourcing application. *First Monday* 13, 6 (2008).
- Evans, B.M. and Chi, E.H. Towards a model of understanding social search. *Proc. CSCW 2008*, ACM (2008), 485-494.
- Howe, J. The rise of crowdsourcing. *Wired Magazine* 14, 6 (2006).
- Ipeirotis, P. Mechanical Turk: The Demographics. *A Computer Scientist in a Business School*, 2008. <http://behind-the-enemy-lines.blogspot.com/2008/03/mechanical-turk-demographics.html>.
- Irani, L. and Silberman, S. Turkopticon: The Sourced Crowd is Made of People. Presentation given at Dolores Labs, 10 June 2009.
- Kittur, A., Chi, E.H., and Suh, B. Crowdsourcing user studies with Mechanical Turk. *Proc. CHI 2008*, ACM (2008), 453-456.
- Olleros, F. Learning to Trust the Crowd: Some Lessons from Wikipedia. *2008 International MCETECH Conference on e-Technologies*, (2008), 212-216.
- Snow, R., O'Connor, B., Jurafsky, D., and Ng, A.Y. Cheap and Fast—But is it Good? Evaluating Non-Expert Annotations for Natural Language Tasks. *Proc. EMNLP 2008, ACL (2008)*, 254-263.
- Sorokin, A. and Forsyth, D. Utility data annotation with Amazon Mechanical Turk. *Proc. CVPRW 2008, IEEE (2008)*, 1-8.

12. US Census Bureau. US Census Bureau 2009 Statistical Abstract of the United States. <http://www.census.gov/compendia/statab/2009edition.html>.
13. Von Ahn, L. and Dabbish, L. Labeling images with a computer game. *Proc. CHI 2008*, ACM (2004), 319-326.