Sellers' problems in human computation markets

M. Six Silberman, Joel Ross, Lilly Irani, and Bill Tomlinson
Department of Informatics
University of California, Irvine
msilberm | jwross | lirani | wmt@uci.edu

ABSTRACT

"Tools for human computers" is an underexplored design space in human computation research, which has focused on techniques for buyers of human computation rather than sellers. We characterize the sellers in one human computation market, Mechanical Turk, and describe some of the challenges they face. We list several projects developed to approach these problems, and conclude with a list of open questions relevant to sellers, buyers, and researchers.

Categories and Subject Descriptors

H.5.m [**Information Interfaces and Presentation (e.g., HCI)**]: Miscellaneous

General Terms

Design, Economics, Human Factors

Keywords

Human computation, Mechanical Turk, Markets

1. INTRODUCTION

"Tools for human computers" is an underexplored design space in human computation research.

With very few exceptions (notably [1], [2], [3]), human computation research has focused exclusively on problems facing users (or "requesters") of human computation. Put shortly, the user of human computation wishes to get good data from "human computers" quickly, without paying much (e.g. [4], [5]). With the growth of paid crowdsourcing systems, most famously Amazon's Mechanical Turk ("AMT") platform, and interest in market mechanisms in human computation (e.g., [6], [7], [8], [9], [10] [11]) and human-computer interaction (e.g., [12], [13], [14], [15], [16]), the number of people doing human computation tasks for money has also grown. Sellers[1] in the market for human computation, like buyers, have interesting and difficult problems to solve, some of which could benefit from the technical expertise of human computation researchers. In this note we first characterize sellers in the AMT labor pool and describe some of their problems. We present several projects, including one developed by the authors, that approach some of these problems. Finally, we list open design problems and empirical questions of interest to human computers to which human computation researchers might contribute.

KDD-HCOMP '10 Washington, DC USA

2. THE CROWD AND ITS PROBLEMS

The AMT labor pool hosts a growing international population earning less than USD 10,000/yr., some of whom are reliant on their Turking income to make basic ends meet. The uncertainty associated with HIT payment complicates human computers' work and reduces their effective wage. This uncertainty is due in part to the apparent prevalence of fraudulent requesters, to whom certain design decisions have made AMT particularly attractive.

Intuitively, we might expect that, just as buyers of human computation aim to minimize expense at a fixed quality (or maximize quality within a cost constraint), sellers of human computation wish to secure payment with a minimum time expenditure, even if this means "gaming the system" by providing responses they know are of low quality. Fraudulent sellers do appear to optimize in this way, but a reading of survey responses and forum discussions reveals a concern for what is "fair" and "reasonable" rather than a desire to maximize short-term personal earnings at requester expense.

Ross et al. [17], [18], extending work by Ipeirotis [19], [20], present longitudinal demographic data on the "increasingly international" AMT. They find that AMT hosts a growing population of young, male, Indian workers earning less than USD 10,000/yr. Additionally, almost a third of Indian Turkers surveyed reported that they always or sometimes relied on their Turking income to "make basic ends meet" (27% in May 2009 and 31% in Feb. 2010). Between May 2009 and Feb. 2010, the fraction of US Turkers surveyed reporting reliance held steady at 13±1%.

Many Turkers see themselves as laborers doing work to earn money. In survey data collected in Feb. 2009 (n = 878) [21], the most commonly reported motivation for doing HITs was making money. 91% of respondents mentioned

[1]We use "sellers," "workers," and, when referring to AMT, "Turkers" interchangeably.

a desire to make money, whether as a form of supplemental income or in order to purchase extras. Turking to pass the time, in contrast, was mentioned by only 42% of respondents. Feb. 2010 data ($n = 1000$) from Ipeirotis [22] confirms the importance of money compared to other motivations, with most respondents reporting they do not do HITs for fun or to kill time, and ∼25% of Indian respondents and ∼13% of US respondents reporting that AMT is their primary source of income.

What challenges, then, face these "professional" crowdworkers? We can begin to answer this question by examining the "mTurk Suggestions" board of Turker Nation (`turkernation.com`), a free web forum operated by and for Turkers. The first thread[2] is called

*** IMPORTANT READ: HITs You Should NOT Do! ***

As of late April 2010, the thread has 57 replies and 15,048 views since it was started in late October 2009. User `spamgirl`, a Turker Nation administrator with 55,360 posts on the forum between Dec. 2008 and Apr. 2010, started the thread with the following suggestions:

Do not do any HITs that involve: *filling in CAPTCHAs; secret shopping; test our web page[3]; test zip code; free trial; click my link; surveys or quizzes (unless the requester is listed with a smiley in the Hall of Fame/Shame); anything that involves sending a text message; or basically* ***anything that asks for any personal information at all*** *— even your zip code. If you feel in your gut it's not on the level, IT'S NOT.* ***Why?*** *Because they are scams and you will not get paid for your work. Worse, your personal information will be used to send you spam, or steal your identity, or bill you fraudulently.* [23]

She also advises Turkers to avoid doing search engine optimization (SEO) tasks, creating accounts on other sites (forums, Twitter, Facebook, etc.), posting a requester's content to online classified listings ("or even your own blog"), downloading software to test, or accepting payment for a review or rating. Newcomers to the forum might read the thread as alarmist, but the prevalence of scam HITs and the difficulty of finding legitimate work on AMT is an ongoing discussion among Turkers. User `dreamingfree` responded to the thread, "I don't understand why so many of the kinds of hits that you mention are still ubiquitous on Mturk. I have trouble finding any that aren't in these catagories! :-/" In early Nov. 2009, user `baubles` replied, "The black hats are out in full force today. Dozens of hits asking for email addresses after clicking on their links. Be careful out there." A few days later, user `kaine13` posted, simply, "I'm so damn sick of these..." User `a4x401` offered a newcomer's take:

Being a newbie and having relatively decent pc skills I have been checking all this stuff out and am somewhat upset [about] the things that I have discovered! It's no wonder that people don't trust the requesters, yes I did some of those hits that one should not do and found myself having to repair my pc and remove some pop-ups. After having done that I really got into checking out the program and realized that it's too easy to manipulate it due the fact that work can be rejected after it's finished but the work is still done. All [a requester] has to say is "not to our satisfaction"!!!!! The other way is to just leave the hits open; you still collect your work but don't have to pay! My favorite part is hits that are way too complicated to complete in the time frame allowed! Why is there no control on any of this stuff? I'm sill waiting for hits to be approved from Oct 31 [on Nov. 6], what's up with that? After all this I discovered this forum and have seen the complaints from all the others; why is there no control?[4]

`spamgirl`'s warnings pertain to HITs where (a) there is uncertainty about whether the requester *intends* to pay workers, (b) the requester intends to spam or defraud workers, and/or (c) the task is against AMT's terms of use. (The private incentive to avoid HITs that violate terms of use, according to `spamgirl`, is that if the offending requester's account is suspended, workers who did their tasks will not be paid.) `a4x401` corroborates the malware problem and highlights a few additional ones: (d) requesters can collect and use work without paying workers; (e) some tasks have prohibitively short time limits; and (f) requesters often take longer than seems reasonable to pay, especially given the small amount.[5]

Intrigued by earlier discussions of these phenomena, Irani in Oct. 2008 posted a HIT asking respondents to articulate a hypothetical "Turkers' Bill of Rights" [24]. 67 respondents agreed to have their submissions displayed publicly. These were coded with an eye toward recurring topics of worker concern. 35 responses indicated a concern with unfairly or arbitrarily rejected work and 26 with slow payment. Seven explicitly mentioned a "minimum wage" or "minimum payment" per HIT, and 14 a minimum wage/payment and/or more "fair" compensation generally. Eight indicated a frustration with lack of response to email communication on the part of both requesters and AMT staff, often in the context of complaints about rejections perceived unfair or tasks in violation of the platform's terms of service.

Respondents sought more information about requesters and the ultimate use of work supplied by Turkers. Several respondents, for example, indicated a frustration with tasks requiring third-party website membership or disclosure of an email address, and with tasks perceived as possibly implicated in the production of email or blog spam. Some were interested in a forum in which Turkers could air concerns publicly without censorship or condescension, and worker visibility and dignity more generally. Others were interested in a way to build long-term work relationships with prolific requesters, and worker-requester relations generally. Two respondents mentioned unionization.

Respondents expressed concerns about malicious or fraudulent requesters who systematically reject best-effort work in order to obtain it without payment; data privacy in survey tasks; technologically defective HITs or HITs with unclear or inadequate instructions;[6] arbitrariness of bonuses; and tasks in which workers were asked to rate work submitted by others, presumably as a basis for approving or denying that work. Three respondents said they liked the platform just as it was and that its design was appropriate for "inde-

[2]The thread is "sticky," appearing at the top of the thread list whether or not it has the most recent reply, indicating its importance according to moderators.

[3]These appear generally intended not to test functionality but to drive traffic, often to sites requiring email address registration.

[4]Edited for punctuation.

[5]Workers are automatically paid after 30 days if the requester does not reject the work. Some workers have pointed out that a month is a long time to wait for 5 cents.

[6]These affect workers' ratings negatively by forcing them to either return the task or leave it incomplete.

pendent contractors" who did the work by choice.

In Jan. 2010, Hartmann [25] asked Turkers about the biggest challenges facing crowdworkers. Respondents noted the difficulty of finding "credible work"; of communicating with requesters and Amazon staff; of finding "tasks that pay enough to be worth the effort"; the annoyance of "getting work rejected for no good reason"; and the frustration of technical errors by requesters or by Amazon leading to damaged worker reputations (leaving workers unable to accept work they could have otherwise done). These responses indicated the persistence of previously reported concerns.

In these responses and discussions we can identify at least eight recurring themes articulated by workers: (1) uncertainty about payment; (2) unaccountable and seemingly arbitrary rejections; (3) fraudulent tasks; (4) prohibitive time limits; (5) long pay delays; (6) uncommunicative requesters and administrators; (7) cost of requester and administrator errors borne by workers; and (8) low pay. There are of course others; for more discussion see the "Suggestions" boards on Turker Nation and mTurk Forum (`mturkforum.com`).

3. APPROACHES TO SELLER PROBLEMS

Software tools exist, some built by Turkers, that attempt to help Turkers manage these problems. Many are client-side scripts that add functionality to the AMT interface; some include a web database component. One platform aims to compete with AMT, in part by addressing some of these problems.

Workers and requesters on mTurk Forum have made a number of Turking tools, including a list of all requesters [26], [27], a script for recording your own worker history (which is not preserved by AMT, but is useful knowledge for tax purposes) [28], [29], and, indirectly addressing some of the problems mentioned above, a client-side script to hide HITs posted by particular requesters [30].

Irani and Silberman built and maintain Turkopticon (`turkopticon.differenceengines.com`), a Firefox add-on and web database application that adds a drop-down interface element next to each requester's name in the HIT listing displayed to workers. When the worker mouses over the name of a requester about which the database has data, averages are displayed for four requester attributes, 'communicativity,' 'generosity,' 'fairness,' and 'promptness,' motivated by problems 1-3, 5, 6, and 8 above.

CloudCrowd, launched in Sep. 2009, is positioned as a more "worker-friendly" alternative to AMT. In his post to mTurk Forum introducing the service [31], CEO Edelstein writes that CloudCrowd will offer "a more efficient [worker] interface," payment through PayPal (allowing workers to collect currencies other than USD and INR, the only choices for Turkers), and "credibility" ratings (in place of acceptance rates as in AMT) as the measure of worker quality.

4. OPEN PROBLEMS AND QUESTIONS

These projects are only a tentative step toward addressing the problems facing Turkers and developing a richer understanding of the structure and dynamics of human computation markets. Many questions remain, including:

How does database, interface, and interaction design influence individual outcomes and market equilibria? This has been explored in online auctions (especially eBay; e.g., [32], [33], [34], [35]) but not human computation.

What are the economics of fraudulent tasks (scamming and spamming)? Work in this thread might draw on research on the economics of botnets (e.g., [36], [37], [38]).

What decision logics are used by buyers and sellers in human computation markets? Buyers who read HCOMP papers and workers who "game the system" may maximize financial return, while other buyers and sellers may satisfice. What are the various problems being solved (and with what strategies) by different actors in human computation markets and how do these shape market outcomes?

What's fair in paid crowdsourcing? Economists Akerlof and Shiller [39] argue that "considerations of fairness are a major motivator in many economic decisions" that has been overlooked in neoclassical explanations that assume economic decision makers act rationally. They lament that "while...there is a considerable literature on what is fair or unfair, there is also a tradition that such considerations should take second place in the explanation of economic events" (pp. 20, 25). AMT requesters and administrators say tasks should be priced "fairly," but fairness is difficult to define and thus to operationalize in practice. Concepts like a reservation wage (as in [2]) are useful here, but do not settle the matter, which is complicated economically and culturally by the global reach of human computation platforms.

Gaps remain in our demographic understanding of AMT, despite at least two ongoing parallel investigations. For example, US-based researchers (including the authors of the present paper) have designed surveys with income questions where the lowest bin is "less than USD 10,000/yr." Most people in India make less than USD 10,000/yr., so these survey results are lossy. As Indians (and citizens of countries less affluent than the US generally) come to comprise a larger fraction of the Turking population these surveys will need to be revised and redeployed. Additionally, language proficiency is an occasionally overlooked confound in quality control strategies; lack of English fluency, while perhaps a good proxy for unreliability in English language tasks, does not imply intent to "game the system" (cf. [16]).

As new platforms and tools come online and mature, comparative studies in all of these areas will become possible, and longitudinal studies more feasible. Much of the prevalence of the eight recurring themes is owed to two conditions that characterize AMT: information asymmetry and the power differential between requesters and administrators on one hand and workers on the other. How new human computation systems are designed (mostly through software) to make information available and distribute decision-making powers to different actors will be of practical interest to buyers and sellers, and empirical and theoretical interest to scholars.

Students in a joint Harvard/Stanford law schools course on "difficult problems in cyberlaw" [40] developed a list of hypothetical projects that might address some of the problems discussed above, including "Turkopticon-style MTurk plugins that achieve various ends, such as adding social features to [AMT] and increasing accountability" [41]. Human computation researchers are well-suited to evaluate such proposals, contributing to understanding of human computation market design and user experiences of buyers and sellers.

5. CONCLUSION

Fraud, spam, malware, and frustrating user experiences are endemic to the internet broadly, not just human compu-

tation markets. We do not suggest that human computation researchers should take it upon themselves to solve these problems, or even that they can be "solved" definitively. We do suggest however that it may be both practically useful and theoretically productive to ask how human computation markets can be made less hospitable to these "black hats."

Most of us complain about our jobs, no matter how well-paid; that Turkers do also is hardly surprising. That their work takes place entirely within markets built in software presents a moderately novel situation, however, in which the difference between making rent and failing to do so can be made by a software (market) design decision. We believe the expertise of human computation researchers is well-suited to explore and address some of the problems raised by human computers, and that our understanding of human computation will be both broadened and deepened in the process.

6. ACKNOWLEDGMENTS

We thank the users of Turkopticon; the Turkers who completed our HITs; the members of Turker Nation and mTurk Forum, especially `numbers` and `spamgirl`; Beatriz da Costa; Kavita Philip; Lukas Biewald; Paul Dourish; and Gloria Mark. This work was funded by the Donald Bren School of Information & Computer Sciences at UC Irvine, the California Institute for Telecommunications and Information Technology, the Sloan Foundation, and the National Science Foundation (Grant No. 0644415).

7. REFERENCES

[1] Zittrain, J. Ubiquitous human computing. *Phil. Trans. R. Soc. A* **366**: 3813-3821, 2008.

[2] Horton, J. J. and L. Chilton. The labor economics of paid crowdsourcing. arXiv:1001.0627v1 [cs.HC], 2010.

[3] Horton, J. J. The condition of the Turking class. arXiv:1001.1172v1 [cs.CY], 2010.

[4] Snow, R., et al. Cheap and fast—but is it good? *Proc. EMNLP '08*.

[5] Feng, D., et al., Acquiring high-quality non-expert knowledge from on-demand workforce. *Proc. ACL-IJCNLP '09*.

[6] Hsieh, G. et al. Can markets help? *Proc. CSCW '08*: 535-544.

[7] Hsieh, G. and S. Counts. mimir. *Proc. CHI '09*: 769-778.

[8] Hsieh, G., et al. Why pay? *Proc. CHI '10*: 305-314.

[9] Burnham, T. and R. Sami. A reputation system for selling human computation. *Proc. HCOMP '09*.

[10] Jain, S. and D. C. Parkes. The role of game theory in human computation systems. *Proc. HCOMP '09*.

[11] Mason, W. and D. J. Watts. Financial incentives and the "performance of crowds." *Proc. HCOMP '09*.

[12] Kittur, A., et al. Crowdsourcing user studies with Mechanical Turk. *Proc. CHI '08*: 453-456.

[13] Heer, J. and M. Bostock. Crowdsourcing graphical perception. *Proc. CHI '10*: 203-212.

[14] Jeon, G. Y., et al. Re-examining price as a predictor of answer quality. *Proc. CHI '10*: 325-328.

[15] Seuken, S., et al. Hidden markets: UI design for a P2P backup application. *Proc. CHI '10*: 315-324.

[16] Downs, J. S., et al. Are your participants gaming the system? *Proc. CHI '10*: 2399-2402.

[17] Ross, J., et al. Who are the crowdworkers? *Proc. CHI '10*: 2863-2872.

[18] Ross, J., et al. Who are the crowdworkers? Slides, *alt.CHI*, 14 April 2010, Atlanta, GA. `www.ics.uci.edu/~jwross/pubs/RossEtAl-altCHI2010-slides.pdf`.

[19] Ipeirotis, P. Mechanical Turk: The demographics. `behind-the-enemy-lines.blogspot.com/2008/03/mechanical-turk-demographics.html`.

[20] Ipeirotis, P. Turker demographics vs. internet demographics. `behind-the-enemy-lines.blogspot.com/2009/03/turker-demographics-vs-internet.html`.

[21] Ross, J. and B. Tomlinson. Worker motivations in Amazon Mechanical Turk. Unpub. manuscript, 2009.

[22] Ipeirotis, P. The new demographics of Mechanical Turk. `behind-the-enemy-lines.blogspot.com/2010/03/new-demographics-of-mechanical-turk.html`.

[23] `spamgirl` et al. HITs you should NOT Do! `turkers.proboards.com/index.cgi?board=kgeneral&action=display&thread=5126`.

[24] Irani, L. Turk Work. `turkwork.differenceengines.com`.

[25] Hartmann, B. Crowdsourcers respond to crowdsourcing. `blogs.law.harvard.edu/difficultprobs/2010/01/07/crowdsourcers-respond-to-crowdsourcing`.

[26] `techlist`. MTurk Requesters. `techlist.com/mturk/requesters`.

[27] `techlist` et al. Current requesters. `mturkforum.com/showthread.php?t=1100`.

[28] `lightdark` et al. Recording your history, would you like to? `mturkforum.com/showthread.php?t=1109`.

[29] `lightdark` et al. Scripts for user data. `mturkforum.com/showthread.php?t=424`.

[30] `ergo` et al. Block (hide) requester script for Firefox+Greasemonkey. `mturkforum.com/showthread.php?t=1088`.

[31] `CloudCrowd`. Intro from CloudCrowd, an alternative to Mechanical Turk. `mturkforum.com/showthread.php?t=1169`.

[32] Ba, S., et al. The dynamics of the electronic market. *Inf. Sys. Frontiers* **2**(1): 31-40, 2000.

[33] Hahn, J. The dynamics of mass online marketplaces. *Proc. CHI '01*: 317-324.

[34] Resnick, P., et al. The value of reputation on eBay. *Experimental Econ.* **9**(2): 79-101, 2006.

[35] Steiglitz, K. *Snipers, Shills, and Sharks: eBay and Human Behavior*. Princeton, 2007.

[36] Franklin, J., et al. An inquiry into the nature and causes of the wealth of internet miscreants. *Proc. CCS '07*: 375-388.

[37] Li, G., et al. Botnet economics: uncertainty matters. In M. Johnson, ed., *Managing Information Risk and the Economics of Security*: 245-267. Springer, 2008.

[38] Namestnikov, Y. The economics of botnets. Kaspersky Lab Tech. Rep., 2009.

[39] Akerlof, G. A. and R. Shiller. *Animal Spirits*. Princeton, 2009.

[40] Zittrain, J. and E. Stark. Cyberlaw: Difficult Issues. `cyber.law.harvard.edu/cyberlaw_winter10/Main_Page`.

[41] `andrewj` and `spamgirl`. MTurk "App Store"? `turkers.proboards.com/index.cgi?board=mturksuggest&action=display&thread=5543`.