Participation in Human Computation

Benjamin B. Bederson
Alex Quinn
Human-Computer Interaction Lab
Computer Science Department
University of Maryland
{bederson, aq}@cs.umd.edu

INTRODUCTION

As researchers building systems in Human Computation [1,3,4,5,6,7], we have consistently found ourselves facing issues not only related to the structure of human and computer participation in collaborative problem solving, but also to the nature of employing human workers. While not at all surprising, issues relating to participation, motivation, cheating, privacy and ethics are fundamental to this kind of work. So, for this workshop, we are submitting two position papers – one more technical one led by my graduate student Alex Quinn – and this on that focuses on participatory issues.

While these social risks may at first glance seem beyond the scope of HCI researchers (and particularly technologists), we feel that it is precisely our responsibility to address them from the beginning since it is designers, not ethicists or policy makers, who have the power to influence what is built and to mitigate risks before any harm is done. This concern arises partly because we sometimes hear human participants referred to in ways that make us think their human-ness is being de-valued (i.e., "remote person call"). We also think it is of strategic importance to the field because technologies tend to build on one another and so once a technology gets started, it can be difficult or impossible to reverse the social effects.

We have been pleased to see work that begins to address these issues on ethics (e.g., Silberman and Zittrain), economics (e.g., Mason & Watts), and cheating (by many researchers). But we don't feel that it goes far enough, and we would like to argue for an even larger focus on this.

PARTICIPATION AND ECONOMICS

From the earliest days of computation, the essential goals of just about all computer science work is to create solutions that optimize speed, cost and quality. Typical computing solutions are inexpensive, and technical work focuses on improving speed and quality. The field of management has as a main focus, the study of human-only solutions with a focus on increasing speed and decreasing cost while maintaining high quality. Both computer-only and human-only solutions offer solutions at some point in the speed-cost-quality trade-off continuum, so it is perhaps not surprising that interest in human computation has come along to offer different points in this trade-off space.

The field of human computation is mostly at the stage of having one-off solutions – that is, systems that are built to solve a particular problem aiming for a particular point in the trade-off space (i.e., offering better quality than a fully automated approach while not being too expensive). And while there has been some work on understanding the general economics (i.e., work by Mason & Watts that looked at the impact of changing payment rates to workers), this area remains underexplored.

We have joined forces with an economics professor (Ginger Jin) and a business professor (Siva Viswanathan) to start exploring these issues more methodically. We did a pilot study looking at different payment schemes (fixed rate, piece rate, performance bonus, and reputation) to push forward these issues. While getting some interesting results, our efforts also emphasized just how difficult this work is. One significant issue is the challenge of figuring out how to generalize such studies since there are so many variables that are difficult to control. How much does such work depend on the problem domain? Time of day? Country of workers? Platform used? Reputation of requestor? And while the community is typically impressed by how inexpensive web workers are, doing significant studies that control a number of variables can end up not being inexpensive at all.

CHEATING DETERRENCE

Controlling cost and maintaining quality when working with Mechanical Turk has required a delicate balance. The research community have made some strides, and we have developed our own expertise building on that – but we are still searching for the ideal strategy.

On one hand, we can place restrictions on who can do the tasks, and design the interface to make cheating inconvenient for the worker. For example, we sometimes design the form so that randomly chosen answers will be easily detectable. However, these measures are not perfect and they have the potential to slow down the rate of work.

Another option is to skip those controls and just analyze the results carefully in order to identify and reject payment to workers who have turned in a large proportion of bad work. The challenge here is in correctly identifying the invalid work and responding appropriately. Early on, our strategy was to seed the job with ground truth, and reject all work from workers whose overall score fell below a baseline

level. Of course, there is always the possibility that such a good worker might misunderstand the instructions. Therefore, we would send a warning after ten wrong answers and block the worker after twenty. At first, this seemed a good strategy, but we learned that blocking workers and/or rejecting a lot of tasks at once can lead to the worker being permanently banned by Amazon. When this happens, Amazon also refuses payout of the worker's payment account and has not provided any explanation to either the worker or the requester. Thus, in designing a scheme for mitigating dishonest work, we also have to consider ethical implications with respect to the workers.

ETHICS

Although the designers of Mechanical Turk chose to hide the identify of workers from the requesters who post tasks, anytime we have made a mistake in conducting our experiments, we quickly received non-anonymous email communication from real human workers who were affected. When we were designing the experiment, we could regard the workers as a computational resource. Only after there was a problem did the relationship become more personal.

Unfortunately, the consequences of such problems can be very real for workers, especially if they depend on the money. For example, recently we mistakenly rejected some work done by several workers. One of them was subsequently banned by Mechanical Turk. The worker, who is in India, said he lost the entire $130 that was in his Amazon payment account. Despite prodding by us and the worker, Amazon refused to reverse the action. The lesson is that although these systems place a degree of separation between us and the workers, in essence it is still an employment relationship. Although the university IRB does approve our experiments, they have little experience and thus provide very little meaningful oversight. In effect, we enjoy relative impunity from the usual labor laws and IRB regulations when running experiments on Mechanical Turk, and so we have an even greater responsibility to consider the interests of workers as we design experiments.

Of course, this goes well beyond short-term experiments. As we think about future models for human computation and crowdsourcing, we are trying to account for the ethical implications to the people who would be affected. As technologists and researchers, we are in a powerful position to shape the technology that goes forward so that the issues are not left to policymakers to rectify later.

An obvious step we can take now is to design systems to make the terms of employment clear to workers before work begins. Currently, Mechanical Turk displays the reward per task, but workers do not know how long the task is expected to take, the expected effective hourly wage, or the requester's policy for approving or rejecting work. We could design our tools so that disclosing such information is the default and the norm.

Another issue is anonymity. Currently, workers are anonymous to requesters. This may offer protection from bad requesters, but it creates little accountability for cheaters, thus forcing requesters to take measures to deter cheating. In addition, having workers anonymous makes it easier for requesters to ignore the interests of the workers. In the near term, we have tried to maintain contact with workers by email. However, this is not the default with Mechanical Turk. Perhaps future systems will be designed based on the assumption that workers and requesters know more about one another in the beginning. If we design our tools to make communication with workers easier and provide more opportunities for building working relationships of trust, this may become the norm, leading to more positive treatment of both workers and requesters, and generally shaping the way both think about the relationship.

Going forward, we expect the importance of this issue to become increasingly obvious. As new advances in automation technology obviate the jobs of some of today's blue collar workers, some of that labor force might gradually transform into an online group for supporting human computation applications. Based on the current trajectory of this research community, we fully expect to see continued improvement in the ability to harness the online workforce for increasingly sophisticated tasks that have value in a business. If this happens, demand will increase, as will the value of the labor, and so we may well see wages increase, so that it becomes a more viable source of employment for workers. Of course, that demand may be met with abundant supply, so wages could decrease as well. In any case, it is important that we design our technologies now to support the productive and ethical labor environment people will be calling for as the use and visibility of these systems expands.

CONCLUSION

Any system can be abused, but we think that a system's design can encourage less problematic use. To that end, we think it is important that our community debate, create and follow a set of guidelines for best practices. We think that if implemented, such guidelines would result in more ethical utilization of this labor source that could result in better economics with higher quality and more efficient work, benefiting all parties. We created a first draft of these guidelines based on our own experience in [2].

ATTENDANCE

Benjamin B. Bederson (who would like to attend the workshop) is an Associate Professor of Computer Science and the previous director of the Human-Computer Interaction Lab at the Institute for Advanced Computer Studies and iSchool at the University of Maryland. His research is on human computation, mobile device interfaces, interaction strategies, digital libraries, and children's education. He is also Cofounder and Chief Scientist of Zumobi, the first premium mobile app network.

REFERENCES

1. Bederson, B.B., Hu, C., & Resnik, P. (2010) Translation by Interactive Collaboration between Monolingual Users, Proceedings of Graphics Interface (GI 2010), 39-46.
2. Bederson, B.B., & Quinn, A. (2010) Web Workers Unite! Addressing Challenges of Online Laborers. University of Maryland, Human-Computer Interaction Lab Tech Report #HCIL-2010-27.
3. Hu, C., Bederson, B.B., & Resnik, P. 2011. MonoTrans2: A New Human Computation System to Support Monolingual Translation, In Proc. CHI '11., (in press).
4. Quinn, A., & Bederson, B.B. 2011. Human Computation: A Survey and Taxonomy of a Growing Field, Proceedings of CHI 2011, (in press).
5. Quinn, A., Hu, C., Arisaka, T., & Bederson, B.B. (2008) Readability of Scanned Books in Digital Libraries, Proceedings of ACM CHI (CHI 2008), ACM Press, 705-714.
6. Resnik, P., Buzek, O., Hu, C., Kronrod, Y., Quinn, A., & Bederson, B.B.* (2010) Improving Translation via Targeted Paraphrasing, Proceedings of Conference on Empiracal Methods in Natural Language Processing (EMNLP 2010), (in press).
7. Suh, B., Ling, H., Bederson, B. B., & Jacobs, J. W. (2003). Automatic Thumbnail Cropping and Its Effectiveness. UIST 2003, ACM Symposium on User Interface Software and Technology, CHI Letters, 5(2), 95-104.