

How to use Mechanical Turk for Cognitive Science Research

Winter A. Mason (winteram@yahoo-inc.com)

Yahoo! Research
111 W. 40th St.
New York, NY 10018 USA

Siddharth Suri (suri@yahoo-inc.com)

Yahoo! Research
111 W. 40th St.
New York, NY 10018 USA

Keywords: methods; mechanical turk; crowdsourcing

Tutorial Overview and Objectives

In this half-day tutorial we will describe a new tool that has emerged in the last 5 years for conducting online behavioral research: crowdsourcing platforms. The term *crowdsourcing* has its origin in an article by Howe (2006), who defined it as a job outsourced to an undefined group of people in the form of an open call. One of the main benefits of these platforms to behavioral researchers is that they provide access to a large set of people who are willing to do tasks—including participating in research studies—for relatively low pay. The crowdsourcing site with one of the largest subject pools is Amazon’s Mechanical Turk (AMT), so it is the focus of this tutorial.

Originally, Amazon built Mechanical Turk specifically for “human computation” tasks. The idea behind its design was to build a platform for humans to do tasks that are very difficult or impossible for computers, such as extracting data from images, audio transcription and filtering adult content. In its essence, however, what Amazon created was a labor market for micro-tasks (Huang, Zhang, Parkes, Gajos, & Chen, 2010). Today Amazon claims hundreds of thousands of workers, and roughly ten thousand employers, with AMT serving as the meeting place and market (Pontin, 2007; Ipeirotis, 2010). For this reason, it also serves as an ideal platform for recruiting and compensating participants in online experiments.

In this tutorial, we will begin by discussing some of the advantages of doing experiments on Mechanical Turk. Specifically, there are four main advantages to using Mechanical Turk as a platform for running online experiments:

1. While researchers at large universities typically have access to large numbers of undergraduates participating in experiments in exchange for academic credit, these subject pools may be much smaller or even non-existent in smaller colleges and universities, or may be unavailable to all researchers. The options for non-academic researchers are even fewer, with recruitment generally limited to ads posted online and flyers posted in public areas. Mechanical Turk offers a very large pool of online participants for these researchers.

2. Workers on Mechanical Turk tend to be from a very diverse background, spanning a wide range of age, ethnicity, socio-economic status (SES), language, and country of origin. Unfortunately, the population of workers on AMT is not representative of any one country or region, but it does open the doors to cross-cultural and international research (Eriksson & Simpson, 2010) at a very low cost and can broaden the validity of studies beyond the undergraduate population.
3. Studies on Mechanical Turk can be conducted at a very low cost, which clearly compare favorably to paid laboratory participants. For example, (Paolacci, Chandler, & Ipeirotis, 2010) replicated classic studies from the judgment and decision-making literature at a cost of approximately \$1.71/hour, and obtained results that neatly paralleled the same studies conducted with undergraduates in a laboratory setting.
4. All too often, research is delayed because of the time it takes to recruit participants and recover from errors in the methodology. For instance, many academic researchers experience the drought / flood cycle of undergraduate subject pools, with supply of participants exceeding demand at the beginning and end of a semester, and then dropping to almost nothing at all other times. The participant availability on Mechanical Turk is relatively stable, with fluctuations in supply largely due to variability in the number of jobs available in the market. Moreover, experiments can be built and put on Mechanical Turk easily and rapidly, which further reduces the time to iterate the cycle of theory development and experimental execution.

We will then discuss how the behavior of workers compares to laboratory subjects, citing work by researchers from computer science and psychology. Then, we will walk through the mechanics of putting a task on Mechanical Turk including recruiting subjects, executing the task, and reviewing the work that was submitted. We will also provide solutions to common problems that a researcher might face when executing their research on this platform such as techniques for conducting synchronous experiments, methods to ensure high quality work, how to keep data private, and how to maintain code security.

Materials

The majority of the tutorial will be in a presentation format, utilizing slides to explain the process of building and running an experiment on Mechanical Turk. Ideally, if there is access to the internet, we will actually build and run a classic cognitive science study on Mechanical Turk, with participants following along to create their own task if possible. This will allow us to not only demonstrate how to do research on Mechanical Turk, it will also demonstrate how quickly and efficiently research conducted on Mechanical Turk can be.

Relevance to Cognitive Science

Many kinds of cognitive science research rely on human subjects and have methods that can be conducted online, and we will use many examples from classic and recent cognitive science as motivating examples. For instance, the first study in (Goldstone, 1995), presenting numbers and letters to participants with differing hues and allowing participants to match the hue, could have been done with participants on Mechanical Turk with very little cost and effort and may be the motivating example we use in the tutorial. Research by one of this year's invited speakers (Chatterjee, Southwood, & Basilico, 1999) in which participants judged the congruence between action sentences and corresponding images could also be conducted easily and at low cost on Mechanical Turk.

For these kinds of research projects, and in fact, for any research that can be conducted online, Mechanical Turk offers low-cost research with a large, diverse, readily-available pool of participants with a built-in means for running the study and paying participants. This is an opportunity for researchers who do not have ready access to a subject pool, have limited resources for paying participants, are interested in doing cross-cultural research (roughly 50% of workers are not from the United States), or are interested in rapid data collection.

Potential Audience

Because Mechanical Turk is a tool for any researcher who can do studies online, we expect this tutorial to have very broad appeal. The tutorial will be especially relevant to those who are unfamiliar with Mechanical Turk or want to know details about good practices when doing research on the site and ways to ensure reliable data collection. This tutorial will not be particularly useful to individuals who are already using Mechanical Turk for sophisticated research, or who can only conduct their research in the field, in the laboratory (e.g., cognitive neuroscientists) or with special populations (e.g. developmental cognitive scientists).

Credentials

Over the past three years, we have run 6 independent studies of very different types on Mechanical Turk. We have also participated in, reviewed papers for, and/or helped organize multiple workshops on crowdsourcing, including: the Human Computation (HCOMP) Workshop in 2009, 2010 at Knowledge Discovery and Data mining (KDD) and 2011 at

the conference of the Association for the Advancement of Artificial Intelligence (AAAI), Workshop on Crowdsourcing for Search Evaluation (CSE) at SIGIR in 2010, CrowdConf 2011, the workshop on "Computational Social Science and the Wisdom of Crowds" in 2011 at Neural Information Processing Systems (NIPS), and Crowdsourcing for Search and Data Mining (CSDM) 2011 at Web Search and Data Mining (WSDM) conference.

We have two published papers reporting research conducted on Mechanical Turk (Mason & Watts, 2009; Suri & Watts, 2011), as well as two additional manuscripts in press or under review (Mason & Watts, 2011; Mason & Suri, 2011), including a methodological review on how to conduct behavioral research on Mechanical Turk, on which this tutorial is largely based.

Special Requirements

The only technical requirement for this tutorial is reliable internet access and a projector.

References

- Chatterjee, A., Southwood, M., & Basilico, D. (1999). Verbs, events and spatial representations. *Neuropsychologia*, 37(4), 395–402.
- Eriksson, K., & Simpson, B. (2010). Emotional reactions to losing explain gender differences in entering a risky lottery. *Judgment and Decision Making*, 5(3), 159–163.
- Goldstone, R. (1995). Effects of categorization on color perception. *Psychological Science*, 6(5), 298–304.
- Howe, J. (2006). The rise of crowdsourcing. *Wired Magazine*, 14(6), 1–4.
- Huang, E., Zhang, H., Parkes, D. C., Gajos, K. Z., & Chen, Y. (2010). Toward automatic task design: A progress report. In *Proceedings of the ACM SIGKDD workshop on human computation* (pp. 77–85).
- Ipeirotis, P. G. (2010, September). *Analyzing the Amazon Mechanical Turk marketplace* (Tech. Rep. No. CeDER-10-04).
- Mason, W. A., & Suri, S. (2011). *Conducting behavioral research on amazon's mechanical turk*. (<http://ssrn.com/abstract=1691163>)
- Mason, W. A., & Watts, D. J. (2009). Financial incentives and the performance of crowds. In *Proceedings of the ACM SIGKDD workshop on human computation* (pp. 77–85).
- Mason, W. A., & Watts, D. J. (2011). *Collaborative problem solving in networks*.
- Paolacci, G., Chandler, J., & Ipeirotis, P. G. (2010). Running experiments on Amazon Mechanical Turk. *Judgment and Decision Making*, 5, 411–419.
- Pontin, J. (2007, March). Artificial intelligence, with help from the humans. *The New York Times*.
- Suri, S., & Watts, D. J. (2011). Cooperation and contagion in networked public goods experiments. *PLoS ONE*. (To Appear)