Using Mechanical Turk and PsiTurk for Dynamic Web Experiments

Anna Coenen (anna.coenen@nyu.edu)
Doug Markant (doug.markant@nyu.edu)
Jay B. Martin (jbmartin@nyu.edu)
John McDonnell (john.mcdonnell@nyu.edu)

Department of Psychology, New York University
New York, NY 10003

Keywords: Amazon Mechanical Turk; PsiTurk; Online Experiments; Crowdsourcing

Objectives

This half-day workshop will demonstrate how to build custom web-based experiments that rely on participants from Amazon Mechanical Turk (AMT). Attendees will learn how to deploy web-based experiments using PsiTurk, a Python-based platform that simplifies the process of setting up experiments and interacting with AMT.

Workshops discussing the AMT marketplace have been offered at previous Cognitive Science Society meetings (e.g., Mason & Suri, 2011). This workshop will complement those by stepping through a working demo that attendees can use to follow along and run on their personal computers. Importantly, the demo will illustrate how AMT can be used with dynamic, externally-hosted experiments, rather than the basic survey templates currently offered on AMT.

The workshop will have two parts. First, we will outline some of the general advantages and principles of using AMT for online behavioral experiments, including a basic introduction to the AMT website and the data collection process more generally. Second, we will show participants how to use the PsiTurk platform to run any web-based experiment on AMT. This portion of the workshop will emphasize "hands-on" training in AMT and PsiTurk that will teach attendees how to deploy their own web-based experiments.

Outline of the Workshop

Throughout the workshop we will use both slides and live demonstrations of how to use AMT and PsiTurk for running web experiments.

Introduction to Mechanical Turk

We will start by introducing the basic structure behind AMT and demonstrate how to run a simple project.

AMT is the largest online service in the US that offers a marketplace for tasks that need to be solved by human rather than machine intelligence. Human Intelligence Tasks (HITs) are submitted by *requesters*, such as corporations, researchers, organizations, or individuals in need for human participants. They can be completed by *workers* in exchange for a reimbursement that is set by the requester. Workers can also be awarded bonuses or have their payment rejected based on how they completed a HIT. We will walk attendants through a simple example of how to post a HIT, oversee the data collection, and reimburse participants on the AMT website.

Benefits and drawbacks of online experiments

Next, we will cover some of the advantages and pitfalls associated with using AMT for behavioral research.

For cognitive psychologists the appeal of using AMT lies in running computer experiments that would otherwise be completed in the lab, typically by undergraduate students. Online experiments have several advantages:

1. Data from a large number of participants can be collected quickly and at low costs. A few hours are typically sufficient for recruiting a full set of participants in a standard cognition or perception experiment.

2. Since the data collection is anonymous, using AMT minimizes experimenter effects and problems with contaminated subject pools at research departments.

3. For the same reason, experimental results become more replicable. Because subjects do not interact with an experimenter, there is no possibility for experimenter confound. If one researcher runs the code for another's experiment, it is, in principle, a pure replication: there is no source of systematic experimental deviation.

4. In general, web-based experiments are easier to share with other researchers since they are designed to run in standard web browsers and do not require any additional software. This facilitates the re-use of experimental code either for the purposes of direct replication or the design of new experiments.

Potential disadvantages of the method concern the quality of the data, including the possibility that comparatively low reimbursement might lower incentives to engage in a task. To address these questions, several authors have used AMT to replicate classic findings in their field. Paolacci, Chandler, and Ipeirotis (2010), for example, replicated a number of well-known cognitive biases using AMT data. Germine et al. (2012) found no systematic differences in the results of some widely-used perceptual paradigms using laboratory and online data. Rand (2012) also conducted an extensive study into the reliability of AMT workers' demographic data and verified that self-reported demographic information is highly reliable. At NYU's Cognition and Computation lab we have successfully replicated the main findings of multiple classic studies in the concept learning literature (reported in Crump, McDonnell, & Gureckis, in press, as Experiments 8–10), but found that it was critical to test participants for comprehension of the experimental instructions. We also manipulated the monetary incentives of one of these tasks and found it had little effect on the performance in the task, but did affect the dropout rate. In addition to these experimental replications, researchers have addressed the objective reliability of AMT data. Our workshop will delve into the findings of the literature so far on what sorts of experiments do and do not work on AMT.

Running AMT experiments using PsiTurk

Finally, we will demonstrate how researchers can run experiments from their own website using AMT and PsiTurk.

Mechanical Turk offers some basic templates for simple online studies that can be built directly on the website. However, it can also be used to run any web-based experiment programmed directly by the researcher via the *External Question* type. To facilitate this process, John McDonnell and Todd Gureckis from NYU's Cognition and Computation lab coauthored and continue to maintain a Python-based platform that allows users to create HITs for experiments with minimal effort. It provides a back-end framework, handling interaction with Amazon's servers to credit participants, and logging participants' data and identifying information in a database. This allows researchers to build a user-facing front-end providing their own experimental code without having to write software to handle these logistical issues. The platform is available at `http://github.com/NYUCCL/PsiTurk`.

Over the course of the workshop, we will introduce the platform and show how attendants can run their own experiments on AMT. We will do so using a demo experiment coded in JavaScript that will be turned into a HIT. The code for this demo will be available for attendants to easily adapt to their own experimental needs.

Audience

This workshop will appeal to cognitive science researchers who are conducting behavioral experiments in a wide number of areas. For those who are unfamiliar with AMT, the lecture portion of the workshop will explain the mechanics of AMT and review methods for designing and delivering experiments to participants. The interactive portion of the workshop will be particularly informative for scientists who wish to use AMT to run dynamic experiments that go beyond simple surveys, for example involving timing of stimulus presentation, collection of reaction times, or interactions with complex stimuli.

The workshop may also be of use to researchers who are unsure whether online research can accommodate their needs. For example, neuroscientists might be interested in using AMT as a platform for piloting experimental paradigms and online experiments in general for reducing dropout rates for follow-up tasks, but may be unsure whether their paradigms can be easily translated into online experiments. One important theme of the workshop will be the capabilities and limitations of online experiments in general.

Although AMT is currently only available to requesters in the United States, we believe that researchers from other parts of the world could still benefit from the workshop. They might be able to use AMT through collaborations with laboratories in the United States, for example. Also, PsiTurk offers a general framework for running web experiments which can also be helpful for users of other online services.

Preparation

We suggest that participants download the PsiTurk platform before attending the workshop and attempt to set it up before attending the workshop. If they do so, they will be able to follow along during the demonstration segment in which we launch an experiment on AMT.

Presenters

All presenters of the workshop have used AMT and PsiTurk extensively to collect data, and have expertise in writing web-based experiments in JavaScript. John McDonnell is the coauthor and maintainer of the open-source PsiTurk framework for behavioral experiments on AMT. He has also validated AMT as a platform for studying learning using Turkers as participants (Crump et al., in press). The other speakers have several projects in preparation based on AMT data collected using PsiTurk. All of the speakers will be available throughout the workshop to assist attendants in setting up PsiTurk and using the AMT platform.

References

Crump, M. J. C., McDonnell, J. V., & Gureckis, T. M. (in press). The promise of mechanical turk: How online labor markets can help theorists run behavioral experiments. *PLoS One*.

Germine, L., Nakayama, K., Duchaine, B., Chabris, C., Chatterjee, G., & Wilmer, J. (2012). Is the web as good as the lab? comparable performance from web and lab in cognitive/perceptual experiments. *Psychonomic bulletin & review*, 1–11.

Mason, W., & Suri, S. (2011). How to use mechanical turk for cognitive science research. In L. Carlson, C. Hölscher, & T. Shipley (Eds.), *Proceedings of the 33rd annual conference of the cognitive science society* (pp. 66–67). Austin, TX: Cognitive Science Society.

Paolacci, G., Chandler, J., & Ipeirotis, P. (2010). Running experiments on amazon mechanical turk. *Judgment and Decision Making*, *5*(5), 411–419.

Rand, D. (2012). The promise of Mechanical Turk: How online labor markets can help theorists run behavioral experiments. *Journal of Theoretical Biology*, *299*, 172–179.