

# 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, targeting participants on Amazon Mechanical Turk (AMT).

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

## 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.

### General 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, or any organization or individual 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.

### Advantages 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. The advantages are obvious:

1. Data from a large number of participants can be collected quickly and at low costs. A few hours are typically sufficient for a standard cognition or perception experiment.
2. Since the data collection is anonymous, using AMT also minimizes experimenter effects and problems with contaminated subject pools at research departments.
3. For the same reason, experimental results become more replicable. If researchers decide to share their experiment code, it becomes easy to run the same study again with minimal effort.

Potential disadvantages of the method concern the quality of the data, and the role that the comparatively low reimbursement might play in lowering incentives to engage in a task. To address these questions, several authors have used AMT to replicate classic findings in their field. Paolacci and colleagues (2010), for example, replicated some well-known cognitive biases from the Judgment and Decision-Making literature using AMT data. Germine and colleagues (2012) found no systematic differences in the results of some widely-used perceptual paradigms using laboratory and online data. At NYU's Cognition and Computation lab we have successfully replicated the main findings of multiple classic studies in the concept learning literature [WAS LOOKING TO ALSO INCLUDE CRUMP ET AL PAPER FROM JOHNS PRESENTATION HERE, STILL IN PRESS?]. We also manipulated the monetary incentives of one of these tasks and found it had little effect on the performance in the task, although it did affect the drop-out rate. In addition to these experimental replications, researchers have addressed the objective reliability of AMT data. David Rand (2012), for example, conducted an extensive study into the reliability of AMT workers' demographic data and verified that self-reported demographic information is highly reliable.

In summary, we will illustrate that AMT data has stood up to a range of replicability tests so far. Given that some heterogeneity is always to be expected from replicated experiments

and keeping in mind that undergraduate subject pools introduce their own sampling biases, these should be encouraging for scientists who are using or planning to use AMT. Another advantage of using web-based experiments is that they can just as easily be repeated in the lab, if such cross-validation is desired.

### Running AMT experiments using psiTurk

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

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 NYUs Cognition and Computation lab have written a Python-based platform that allows users to create HITs for experiments with minimal effort. It provides the backend machinery that will enable researchers to plug in their own web-based experiments and create HITs on AMT directly. [MAYBE MORE DETAIL HERE? NOT SURE..]

The project is available on github for anyone to download and use to implement their own web experiments. In the workshop we will introduce the platform and show how attendants can run their own experiments on AMT. We will do so using example JavaScript experiments that will be turned into a HIT. These experiments will then be available for attendants to easily adapt to their own experimental needs.

### Audience

We expect that this workshop will appeal to cognitive science researchers who are conducting behavioral experiments in a wide number of areas. The workshop will be particularly informative for scientists who are unfamiliar with AMT, but are keen to transfer their experiments to the web (using JavaScript or Flash for example).

The workshop might also be useful for those who are already using AMT but would benefit from a platform like psiTurk to help them run customized web experiments without the need to engage in the backend machinery. While perhaps less useful to neuroscientists who require more elaborate experimental techniques, it might still be worthwhile for some to consider AMT as a platform for piloting the behavioral parts of experiments before taking them to the lab.

### Presenters

All presenters of the workshop have used AMT and psiTurk extensively to collect data and presented results based on this data at multiple conferences and published in XXX [YOU PROBABLY KNOW BEST WHAT ALL YOUR AMT CREDENTIALS ARE? HOPEFULLY 'EXTENSIVELY' WILL BE TRUE ABOUT MYSELF BY THEN... ]. John McDonnell was a main contributor to the psiTurk platform and also conducted previous workshops introducing AMT to other researchers. [ANY OTHER COOL STUFF YOU GUYS DID? PARDON MY IGNORANCE HERE...]

### References

- 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 *Proceedings of the 33rd annual conference of the cognitive science society* (pp. 66–67).
- 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.