OPEN ACCESS Freely available online

PLOS ONE

Evaluating Amazon's Mechanical Turk as a Tool for Experimental Behavioral Research

Matthew J. C. Crump[1]*, John V. McDonnell[2], Todd M. Gureckis[2]

1 Department of Psychology, Brooklyn College of CUNY, Brooklyn, New York, United States of America, **2** Department of Psychology, New York University, New York, New York, United States of America

Abstract

Amazon Mechanical Turk (AMT) is an online crowdsourcing service where anonymous online workers complete web-based tasks for small sums of money. The service has attracted attention from experimental psychologists interested in gathering human subject data more efficiently. However, relative to traditional laboratory studies, many aspects of the testing environment are not under the experimenter's control. In this paper, we attempt to empirically evaluate the fidelity of the AMT system for use in cognitive behavioral experiments. These types of experiment differ from simple surveys in that they require multiple trials, sustained attention from participants, comprehension of complex instructions, and millisecond accuracy for response recording and stimulus presentation. We replicate a diverse body of tasks from experimental psychology including the Stroop, Switching, Flanker, Simon, Posner Cuing, attentional blink, subliminal priming, and category learning tasks using participants recruited using AMT. While most of replications were qualitatively successful and validated the approach of collecting data anonymously online using a web-browser, others revealed disparity between laboratory results and online results. A number of important lessons were encountered in the process of conducting these replications that should be of value to other researchers.

Citation: Crump MJC, McDonnell JV, Gureckis TM (2013) Evaluating Amazon's Mechanical Turk as a Tool for Experimental Behavioral Research. PLoS ONE 8(3): e57410. doi:10.1371/journal.pone.0057410

Editor: Sam Gilbert, University College London, United Kingdom

Received November 20, 2012; **Accepted** January 21, 2013; **Published** March 13, 2013

Funding: The authors have no support or funding to report.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: mcrump@brooklyn.cuny.edu

Introduction

One challenging aspect of experimental psychology research is the constant struggle for data. Typically, researchers depend on university undergraduates who participate in studies in exchange for experience, course credit, or money. Research progress depends on the ebb and flow of the semester. As a result, it can take weeks, months, or even years to conduct a large behavioral study. This issue is even more salient for researchers at smaller universities.

One appealing solution is to collect behavioral data over the Internet. In theory, online experimentation would allow researchers to access to a large and diverse pool of potential subjects worldwide, using automated replicable techniques free of unintended experimenter effects. However, the main obstacle to conducting Internet-based research is finding people who are willing to participate and compensating them.

Recently, a number of online crowdsourcing services have been developed which connect individuals willing to perform online tasks with other individuals willing to pay for work to be done. Perhaps the most popular system is Amazon's Mechanical Turk (AMT). AMT is useful for behavioral researchers because it handles recruitment and payment in a fairly automatic way. Most importantly, there are a large number of people who use AMT making it a great way to advertise and distribute studies (over 100,000 active users in 2007 [1]).

There are a number of recent summaries about using AMT for research [2]. In addition, the service has been validated as a tool for conducting survey research [3,4], one-shot decision-making research [5,6], collective behavior experiments [7,8], for norming stimuli, and conducting behavioral linguistics experiments [9,10].

However, less is known about the viability of conducting behavioral experiments typical of those used in cognitive science and cognitive psychology. Such studies are unique in that they typically involve multi-trial designs, sustained attention on the part of participants, millisecond timing for response recording and stimulus presentation, and relatively complex instructions. These features present two key challenges for online data collection. First, there are technical challenges in programming web-based experiment protocols and then ensuring the browser systems of the participant and experimenter support the same features. Second, experiments where memory and timing are important are likely more sensitive to incidental aspects of the testing environment that are difficult to control online (e.g., presence of distractions, problems with display, pausing for long periods in the middle of the task, and misreading or misunderstanding of instructions).

The aim of the present paper is to validate AMT as a tool for behavioral cognitive research, with a specific focus on complex multi-trial designs. We focus on AMT simply because it is the most popular system currently available and the one most researchers would likely consider. If the data obtained from AMT can replicate classic findings in the field with reasonable fidelity, it will validate the potential of the service for use in cognitive behavioral research. In this sense our study joins a number of recent articles exploring the relationship between data collected online and in the