**Performing Massively Open Online Social Experiments with Volunteer Science**

David Lazer, Waleed Meleis, Brooke Foucault Wells, Christoph Riedl, Jason Radford, Brian Keegan, Katya Ognyanova, Stefan Wojcik, Jefferson Hoye, Ceyhun Karbeyaz,

Technological innovation has expanded the scope of possibility for the kinds of research that can be performed. Social innovation in the form of online laboratories like Project Implicit, online panels like GfK’s TESS program, worker pools like Amazon’s Mechanical Turk, and citizen science projects like GalaxyZoo have emerged to take advantage of advances in technological capacity to push the boundaries of science. In the realm of online social experimentation, we created Volunteer Science to combine the open, citizen science model of Galaxy Zoo and Fold.it with social science experimentation occurring via Mechanical Turk and Project Implicit. In this presentation, we describe the capabilities Volunteer Science contributes to the social sciences landscape. Then we discuss the fourteen studies we conducted to validate our volunteer-based model of online social research. We conclude with the challenges remaining for Volunteer Science and the steps ahead of us.

Existing models of online social experimentation demonstrate the capacity for larger, less invasive, and lower cost research. Studies using workers on Mechanical Turk show that large numbers of subjects can be recruited at a fraction of the cost of existing student samples (Mason and Suri 2011, Mason and Watts 2012). And, as Crump, McDonnel, and Gurekis (2013, p. 16) acknowledged, running a replication study over the course of a long weekend is revolutionary. However, existing approaches to online experimentation remain relatively parochial and narrowly tailored to specific research paradigms. The weakness of this decentralized, ad hoc approach to online experimentation is a continued dependence on specific subject pools, experiments that remain difficult to replicate or transfer to other researchers, and a continuous re-inventing the wheel as every social scientist must solve the same problems such as recruiting and managing large numbers of participants. We created Volunteer Science to make the technical solutions we were creating in our own research available to the broader community.

Volunteer Science attempts to combine the strengths of existing approaches in an open, accessible, and ethically robust way. Currently, anyone with a web browser can visit Volunteer Science and participate in any of our more than thirty experiments or surveys. Only after people participate do they have to decide whether to donate their data to researchers and thereby become research subjects. Conversely, social scientists are given access to an online team management system enabling collaboration from anywhere in the world. Scientists are also given access to the more than twenty experimental templates developed on Volunteer Science including survey-based experiments, the prisoner’s dilemma, and reaction time experiments. In addition, researchers have access to our Application Programming Interface (API) which provides simplified access to technically complex functions like real-time subject communication, subject matching, consent management, and post-game rewards.

Volunteer Science integrates directly with existing subject pools like Mechanical Turk and physical laboratories; however, we have also developed a volunteer-based model of participation whereby anyone can participate. This citizen science-based approach is meant to provide a substantial reduction in subject costs, the ability to capture a demographically diverse audience, and a large pool of participants on which we collect a large variety of data. To test the platform and our ability to perform valid social research with these volunteers, we performed fourteen research studies.

Our validation studies were drawn from canonical research across the social sciences. In behavioral economics, we attempted to replicate four biases and successfully replicated all but one: replicated two anchoring effects and one framing effect but failed to replicate the risk-reward correlation bias. From cognitive psychology, we attempted and replicated both the Stroop and flanker effects (MacLeod 1991 and Eriksen 1995 respectively) which assess small effects in the reaction time between seeing different text-based stimuli. From personality psychology, we attempted and found supporting but not validating evidence for the five-factor personality inventory. That is, we found satisfactory results for the reliability of the five factors using Chronbach’s alpha (each >.75); however, we failed to confirm the five factors using Confirmatory Factor Analysis. From game theory, we validation our volunteers’ sensitivity to incentives using the prisoner’s dilemma and a commons dilemma. For both, we found that users changed their behavior in response to differing payoffs in the expected directions; however, subjects appear to be less sensitive to changes in high risk payoffs. We validated the underlying difficulty of two canonical problem solving tasks: the distribution of clues in Hidden Profile and, for the Travelling Salesperson, the number of “cities” inside the convex hull. Finally, we used to Cognitive Reflection tests and two statistical knowledge questions from the behavioral economics literature to compare our volunteers with other samples. In all three, our volunteers demonstrated a higher quantitative skill than found in other samples, typically of college students, but also including members of the public.

Overall, our validation study demonstrated the general validity of the platform while revealing some of its strengths and weakness. Another challenge is, even though more than 18,000 individuals participated in at least one of these studies, very few participated simultaneously. Thus, we are still a ways away from being able to perform and validate experiments which require multiple participants with volunteers. Finally, we are beginning to open our doors to larger circles of experimental groups. Yet, we have not completely worked out the procedural and financial system needed to regulate experiments and fund the platform. As we move forward, governance and sustainability will become a central area of work.

**References**

Crump, Matthew J. C., John V. McDonnell, and Todd M. Gureckis. 2013. “Evaluating Amazon’s Mechanical Turk as a Tool for Experimental Behavioral Research.” Edited by Sam Gilbert. *PLoS ONE* 8 (3): e57410.

Eriksen, Charles W. 1995. “The Flankers Task and Response Competition: A Useful Tool for Investigating a Variety of Cognitive Problems.” *Visual Cognition* 2 (2-3): 101–18.

MacLeod, Colin M. 1991. “Half a Century of Research on the Stroop Effect: An Integrative Review.” *Psychological Bulletin* 109 (2): 163–203.

Mason, Winter, and Siddharth Suri. 2011. “Conducting Behavioral Research on Amazon’s Mechanical Turk.” *Behavior Research Methods* 44 (1): 1–23.

Mason, Winter, and Duncan J. Watts. 2012. “Collaborative Learning in Networks.” *Proceedings of the National Academy of Sciences* 109 (3): 764–69.