LabintheWild: How do Design Uncompensated, Feedback-driven Online Experiments

Nigini Oliveira
Computer Science & Engineering, DUB group
University of Washington
nigini@cs.washington.edu

Eunice Jun
Computer Science & Engineering, DUB group
University of Washington
emjun@cs.washington.edu

Trevor Croxson
Computer Science & Engineering, DUB group
University of Washington
tcroxson@gmail.com

Krzysztof Z. Gajos
School of Engineering and Applied Sciences
Harvard University
kgajos@eecs.harvard.edu

Katharina Reinecke
Computer Science & Engineering, DUB group
University of Washington
reinecke@cs.washington.edu

Abstract
We present LabintheWild, an online experiment platform that provides participants with the opportunity to learn about themselves and compare themselves to others. In the past four years, LabintheWild has attracted approximately 3.5 million participants of diverse ages and education levels who came from more than 200 countries and regions. Leveraging our experience with LabintheWild, we show how researchers can design engaging and robust online experiments that provide personalized feedback. In particular, our interactive tutorial highlights challenges and best-practice guidelines for designing volunteer-based online experiments for diverse participant samples.

Author Keywords
Online experimentation; behavioral experiments; tutorials

ACM Classification Keywords
H.5.m [Information interfaces and presentation (e.g., HCI)]: Miscellaneous

Introduction
Researchers in HCI and various social science disciplines have been increasingly turning to online platforms, such as Amazon's Mechanical Turk, to conduct behavioral experiments with human subjects. Compared to traditional laboratory experiments, online studies offer faster and less trou-

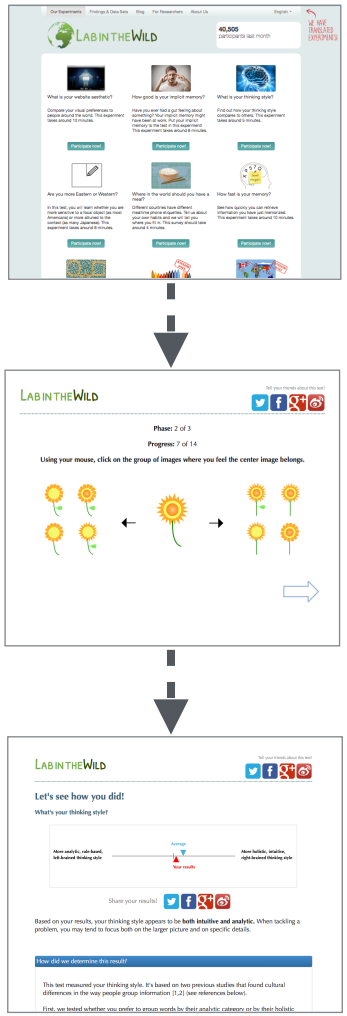

Figure 1: The LabintheWild front page, an example trial page, and a personalized results page.

blesome participant recruitment, as well as larger and more diverse samples [6]. However, such financially incentivized online studies have limits: Studies requiring effortful, but unverifiable, contributions from workers (e.g., subjective responses on Likert scales) remain difficult to conduct on Mechanical Turk because some workers, motivated by quick financial gain, provide plausible, but untrue, answers [8]. Further, Mechanical Turk's user base is insufficiently diverse to conduct studies spanning many countries (for example, the plurality of Turkers come from the United States and India [6]). Further still, the subject pool accessible via online labor platforms is additionally restricted to those who have signed up to receive payment in one of two currencies (U.S. dollars and Indian Rupees [6]). This sign-up barrier systematically excludes participants from certain countries, demographics, and personality traits (see, e.g., [3]).

With this demonstration of our online experiment platform LabintheWild [7], we show a complementary approach for conducting large-scale studies: studies with online volunteers in which participants receive no financial compensation but instead benefit from personalized feedback. On LabintheWild (as on other similar platforms, such as ProjectImplicit, TestMyBrain, and OutofService), experiments enlist participants using short incentive slogans, such as "How fast is your reaction time?", or "Test your social intelligence!" (see Figure 1). After participating in an experiment, volunteers can view a personalized results page that shows how they compare to others.

Such volunteer-based online experiments have several strengths compared to paid online experiments. First, the personalized feedback encourages participants to provide truthful answers [2, 7]. Second, existing volunteer-based platforms have attracted more diverse participant samples than in-lab experiments and those conducted on Mechanical Turk, with participants reporting wider age ranges, more diverse educational backgrounds, and a far more expansive geographic dispersion (see, e.g., [2, 7, 5]). Also, while estimates suggest that a research study (with unlimited financial resources) could access up to 7,300 unique participants via Mechanical Turk [9], several of our LabintheWild studies attracted more than 40,000 participants (and two exceeded 300,000). Despite these strengths, volunteer-based online experiments remain far less prevalent than financially compensated online experiments. A major reason is that designing engaging and robust volunteer-based experiments for diverse participants requires different knowledge and experience compared to financially compensated experiments.

The goal of this demonstration is to equip other researchers with the know-how to design volunteer-based online experiments that incentivize participation with personalized feedback. Building on more than four years of experience with LabintheWild, we developed an interactive experiment tutorial that highlights the most common challenges and best-practice guidelines. Researchers will be able to use the tutorial to answer questions, such as:

- What are the main differences in design between financially compensated and uncompensated experiments?
- What kind of experiments are possible?
- How do I effectively incentivize participation?
- How do I design a rewarding experience for participants?
- How can I design experiments for a diversity of participants (including demographics and devices used to access the studies)?
- How do I ensure data quality in volunteer-based online experiments?
- What is the process for launching my own experiment on LabintheWild?

The LabintheWild Platform

Our volunteer-based online experiment platform LabintheWild was launched in 2012 with the goal of reaching larger and more diverse participant samples than what was possible with either conventional lab studies or through online labor markets such as Amazon Mechanical Turk. We made several key design decisions to achieve this goal:

First, LabintheWild provides unrestricted online access. The experiments are conducted without experimenter supervision, which allows participation in large numbers around the clock independent of location and time zones. No sign up and account creation is needed for participation. The experiments are also open to anyone to participate without requiring people to sign up. This decision lowers the barrier to participation and protects participants' privacy.

Second, LabintheWild provides personalized feedback instead of financial compensation. Paying participants would limit both the size and the diversity of the participant pool because of limited financial resources and because only some participants have the ability to receive online payments or are interested in earning virtual money. Instead, we leverage the human urge to learn about themselves and to compare themselves to others [1]: After participating in an experiment, participants are shown a personalized results page (see Figure 1), which explains how they did and how their performance or preferences compare to others. The feedback usually corresponds to our own research questions or to some other aspect of the experiment that might be interesting to participants. For each experiment, we produce a short slogan that advertises the type of feedback participants will receive (e.g., "Are you more Eastern or Western?"). We use these slogans to advertise each experiment on LabintheWild's front page and to share the experiments via social media. The personalized feedback serves three main purposes. First, it encourages participants to take part in experiments because it enables self-reflection and social comparison. Second, it ensures data quality: Participants are intrinsically motivated to provide honest answers and exert themselves. As a result, experiments conducted on LabintheWild and other volunteer-based experiment platforms produce reliable data that matches the quality of in-lab studies [2, 4, 7] (including studies that require effortful subjective judgement, which pose a challenge on Mechanical Turk [8]). Third, the personalized feedback serves as a word-of-mouth recruitment tool. Participants share their results on social networking sites, blogs, or other web pages, which generates a self-perpetuating recruitment cycle [7]. For example, most LabintheWild participants come to the site from online social networks such as Facebook and Twitter. Since launching four years ago, LabintheWild has been visited more than 3.5 million times, with an average of over 1,000 participants per day completing an experiment.

Third, we design our experiments to take 5–15 minutes to ensure that participation does not become too tedious or exhausting. This also imposes constraints on the experiment design. Studies that require showing large numbers of stimuli are shortened by presenting participants with different randomized subsamples of the full set of stimuli, and we account for the resulting differences in sample frequencies in the analyses.

These design decisions have resulted in a continuous stream of traffic and more diverse participant samples than have been reported for laboratory experiments or those conducted on Mechanical Turk (see [7] for additional details).

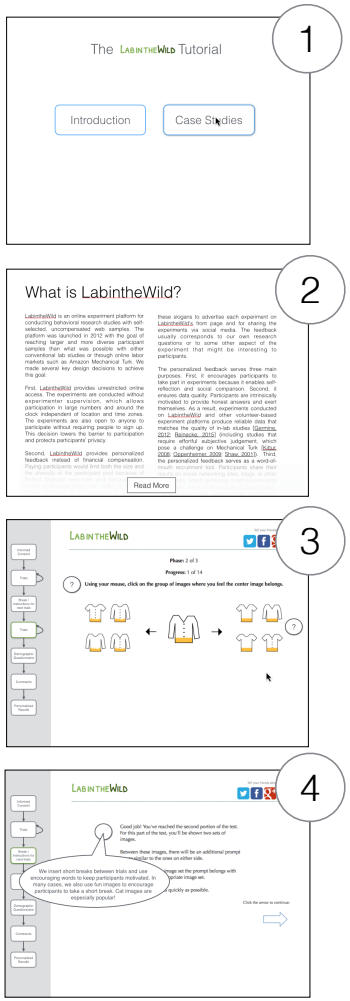

Figure 2: Our interactive online tutorial includes generalizable information about developing volunteer-based online experiments as well as three case studies of successful LabintheWild experiments.

Tutorial for Volunteer-Based Online Experiments

To enable other researchers to conduct studies with diverse participants, we developed an interactive tutorial for designing engaging and robust volunteer-based online experiments (see Figure 2). The tutorial builds on more than four years of experience with LabintheWild, several tens of thousands of comments from participants, as well as on the results of various meta-studies conducted on the platform and its experiments. The web-based tutorial is implemented as a slide show which first provides a high-level overview of the general concept of volunteer-based online experiments (Figure 2.2) before delving into three case study examples of successful LabintheWild experiments (Figure 2.3), including the research question that led to the implementation of the experiment and the experiment outcome. While researchers are able to try the experiments through the eyes of participants, they can click and expand speech bubbles that explain specific design choices (see Figure 2.4). The tutorial ends with a summary of key design choices and a step-by-step guide on the requirements and process for launching an experiment on LabintheWild. During the demonstration at the conference, researchers will be able to explore LabintheWild as well as our interactive tutorial on a large touch-screen display and on several iPads, as well as on their own devices.

Relevance and Contribution

Volunteer-based online experiments offer a promising, complementary research methodology to financially compensated online experiments that is currently underutilized in the research community. With this live demonstration of LabintheWild, its experiments, and our interactive tutorial, we aim to provide researchers with the knowledge and tools for designing and launching engaging and robust online experiments on LabintheWild or other volunteer-based experiment platforms.

REFERENCES

1. L Festinger. 1954. A theory of social comparison processes. *Human relations* 7(2) (1954), 117–140.
2. Laura Germine, Ken Nakayama, Bradley C Duchaine, Christopher F Chabris, Garga Chatterjee, and Jeremy B Wilmer. 2012. Is the Web as good as the lab? Comparable performance from Web and lab in cognitive/perceptual experiments. *Psychonomic bulletin & review* 19, 5 (2012), 847–857.
3. Joseph K Goodman, Cynthia E Cryder, and Amar Cheema. 2013. Data collection in a flat world: The strengths and weaknesses of Mechanical Turk samples. *Journal of Behavioral Decision Making* 26, 3 (2013), 213–224.
4. Samuel D Gosling, Simine Vazire, Sanjay Srivastava, and Oliver P John. 2004. Should we trust web-based studies? A comparative analysis of six preconceptions about Internet questionnaires. *American Psychologist* 59, 2 (2004), 93–104.
5. Joshua K Hartshorne and Laura T Germine. 2015. When does cognitive functioning peak? The asynchronous rise and fall of different cognitive abilities across the life span. *Psychological science* (2015), 0956797614567339.
6. Winter Mason and Siddharth Suri. 2012. Conducting behavioral research on Amazon's Mechanical Turk. *Behavior research methods* 44, 1 (2012), 1–23.
7. Katharina Reinecke and Krzysztof Z. Gajos. 2015. LabintheWild: Conducting Large-Scale Online Experiments With Uncompensated Samples. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing (CSCW '15)*. ACM, 1364–1378.
8. Aaron D. Shaw, John J. Horton, and Daniel L. Chen. In *Proceedings of the ACM 2011 conference on Computer supported cooperative work*. New York, NY, USA.
9. Neil Stewart, Christoph Ungemach, Adam JL Harris, Daniel M Bartels, Ben R Newell, Gabriele Paolacci, and Jesse Chandler. 2015. The average laboratory samples a population of 7,300 Amazon Mechanical Turk workers. *Judgment and Decision Making* 10, 5 (2015), 479.

Reception Demo Summary

Our LabintheWild platform and the interactive experiment design tutorial will provide CSCW attendees with the opportunity to try out experiments and to learn more about the design of volunteer-based online experiments. The LabintheWild platform was first introduced in a CSCW'15 paper, so we expect at least some attendees to be familiar with it. Since then, we have received numerous requests from researchers in the community to launch their own experiment on LabintheWild (several have actually done so). The main challenge for these researchers was usually to understand the difference between financially compensated and uncompensated online experiments, the latter requiring more thought and effort in designing a rewarding experience for participants. We have developed the interactive experiment design tutorial with the questions from these other researchers in mind and believe that a demonstration of the platform, its experiments, and the interactive tutorial will lead to a greater number of researchers being able to make use of this fairly new research methodology.

We will demonstrate the LabintheWild platform and the interactive tutorial simultaneously on two large touch screens. The interactive experiment design tutorial will allow attendees to clearly see design differences between financially based online experiments (such as those conducted on Mechanical Turk) and volunteer-based experiments that provide participants with personalized feedback. Attendees will be able to learn how to turn their own research questions into a rewarding experience for participants, and how to launch an experiment on LabintheWild in order to attract large and diverse participant samples. Apart from the two large touch screens, we will additionally have 2-3 iPads that researchers can use to engage with the platform and/or the tutorial. In addition, attendees can use their own devices to access both.

If accepted, we will demonstrate our LabintheWild platform and the interactive experiment design tutorial as a Reception Demo.

Team: Nigini Oliveira is a PhD candidate at the Universidade Federal de Campina Grande in Brazil and a research scientist at the University of Washington where he researches how the LabintheWild community can collaborate in the research process. Eunice Jun is a PhD student in the Computer Science department at the University of Washington, working towards designing LabintheWild experiments to be maximally rewarding for participants from various countries and cultures. Trevor Croxson is a research scientist at the University of Washington who develops LabintheWild experiments and supports the maintenance of the platform. Krzysztof Gajos is a Full Professor at Harvard University and Katharina Reinecke is an Assistant Professor in the Computer Science department at the University of Washington. Krzysztof and Katharina co-founded LabintheWild in 2012. All five authors developed the interactive experiment tutorial together and are closely involved in LabintheWild.

Demonstration Requirements: This demonstration requires a standard-sized table for the two touch screens, as well as an Internet connection. Devices and extension cords will be brought by the authors.