

# Naturalistic Distributed Experimentation as a Source of New Insight

**Sandy J. J. Gould, Anna L. Cox, Duncan P. Brumby**

UCL Interaction Centre, University College London, Gower Street, London, UK  
s.gould@cs.ucl.ac.uk, anna.cox@ucl.ac.uk, brumby@cs.ucl.ac.uk

## Abstract

Human performance experiments are often conducted online with the help of paid crowdworkers and citizen scientists. This approach produces reliable data, but there are concerns that the inevitable loss of control that accompanies online experimentation might confound results. Researchers have therefore spent time considering how to regain control and mitigate the effects of confounds. In this abstract we argue that confounding factors in online work can be put to novel use, giving us insight into research questions we might otherwise be unable to answer.

## Insight through the unexpected

Paid-for crowdsourcing platforms like Amazon's Mechanical Turk have been used alongside purpose-built platforms like Lab in the Wild to obtain experimental data in human performance studies (Kittur, Chi, & Suh, 2008; Komarov, Reinecke, & Gajos, 2013). The environments in which these studies are run are unpredictable and do not share the characteristics or behavioural norms of laboratories. While this is a potential threat to validity, we argue that the 'wildness' of online studies affords opportunities to gain new insights into behaviour that would not be possible in the lab.

Running research studies online requires researchers to sacrifice a degree of control. Efforts have been made to compensate for this by adjusting tasks to, for instance, reduce satisficing behaviour (Kapelner & Chandler, 2010). In paid-for crowdsourcing settings, money can also be used to effect greater control over behaviour (Mason & Watts, 2010). The idea is to mitigate the loss of control in order to retain some of the qualities of lab work.

We think that for some research questions this approach does not play to the strengths of crowd-based research. Rather than attempt to minimise confounding factors in an environment where researchers have little control, we think that the behaviours people exhibit during online participation have the potential to be put to good use in naturalistic experiments. Potential confounds can be measured and used as covariates, but, more interestingly,

they can also be installed as an integral part of an experiment's research objectives.

Our own work provides an illustration of this approach. We have previously found that when completing online experiments, people often interrupt themselves to perform other tasks (Gould, Cox, & Brumby, 2013). This should be concerning. Even short interruptions disrupt performance and increase the prevalence of errors (Altmann, Trafton, & Hambrick, 2013; Altmann & Trafton, 2002). For human performance tasks using, for example, reaction times this kind of behaviour can affect results in unpredictable ways.

One can measure this and perhaps compensate for it. Or one can view it as behaviour that can be put to work to attack traditional problems in novel ways that deliver new insights. If online participants frequently interrupt themselves to perform other tasks, this gives us the chance to study the effect of real and meaningful interruptions on performance. This has the potential to give us more insight into natural multitasking behaviour than we can achieve with laboratory tasks that are unlikely to be meaningful or important to them. In this way we can begin to understand people's natural multitasking habits through the lens of traditional laboratory measures.

The challenge is to develop measures in an environment where people make use of a variety of devices and platforms that researchers have little control over. Browser-based tasks are often limited in the extent to which they can inspect system-level activity. Installed apps present compatibility issues (although app stores can help in this regard: see (Böhmer, Lander, Gehring, Brumby, & Krüger, 2014). Without developing browser add-ons or standalone software, some measures might need to be inferred from activity metrics like idle time (Mao, Kamar, & Horvitz, 2013) or the development of interaction methods that would be unnecessary in the lab.

We think that developing this approach has the potential to increase the reliability of online experiments, tell us something about online participants and deliver novel insights that could not be uncovered with traditional situated or laboratory investigations. It also has the

## Acknowledgements

Gould, Cox and Brumby are supported by the CHI+MED project, United Kingdom EPSRC grant EP/G059063/1.

## References

- Böhmer, M., Lander, C., Gehring, S., Brumby, D. P., & Krüger, A. (2014). Interrupted by a Phone Call: Exploring Designs for Lowering the Impact of Call Notifications for Smartphone Users. In *Proceedings of the 32Nd Annual ACM Conference on Human Factors in Computing Systems* (pp. 3045–3054). New York, NY, USA: ACM. doi:10.1145/2556288.2557066
- Gould, S. J. J., Cox, A. L., & Brumby, D. P. (2013). Frequency and Duration of Self-initiated Task-switching in an Online Investigation of Interrupted Performance. In *Human Computation and Crowdsourcing: Works in Progress and Demonstration Abstracts AAAI Technical Report CR-13-01* (pp. 22–23). Retrieved from <http://www.aaai.org/ocs/index.php/HCOMP/HCOMP13/paper/view/7485>
- Kapelner, A., & Chandler, D. (2010). Preventing Satisficing in online surveys. In *CrowdConf*.
- Kittur, A., Chi, E. H., & Suh, B. (2008). Crowdsourcing user studies with Mechanical Turk. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 453–456). New York, NY, USA: ACM. doi:10.1145/1357054.1357127
- Komarov, S., Reinecke, K., & Gajos, K. Z. (2013). Crowdsourcing Performance Evaluations of User Interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 207–216). New York, NY, USA: ACM. doi:10.1145/2470654.2470684
- Mao, A., Kamar, E., & Horvitz, E. (2013). Why Stop Now? Predicting Worker Engagement in Online Crowdsourcing. In *First AAAI Conference on Human Computation and Crowdsourcing* (pp. 103–111). AAAI. Retrieved from <http://www.aaai.org/ocs/index.php/HCOMP/HCOMP13/paper/view/7498>
- Mason, W., & Watts, D. J. (2010). Financial incentives and the “performance of crowds.” *SIGKDD Explor. Newsl.*, 11(2), 100–108. doi:10.1145/1809400.1809422