chi+med
making medical devices safer

EPSRC Programme Grant EP/G059063/1

Public Paper no. 298

Control and Confounds in Cross-Cultural Crowdsourced Experiments

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

Gould, S. J. J., Cox, A. L., & Brumby, D. P. (2015). Control and confounds in cross-cultural crowdsourced experiments. Paper presented at the Workshop on "How WEIRD is HCI? Extending HCI Principles to other Countries and Cultures" at CHI 2015, Seoul, South Korea, April 2015.

PP release date: 6 February 2015

file: WP298.pdf

UCL UCLIC Swansea University Prifysgol Abertawe FIT LAB Queen Mary University of London CITY UNIVERSITY LONDON EPSRC Engineering and Physical Sciences Research Council

Control and Confounds in Cross-cultural Crowdsourced Experiments

Sandy J. J. Gould
UCL Interaction Centre
University College London
Gower Street, London WC1E 6BT
s.gould@cs.ucl.ac.uk

Anna L. Cox
UCL Interaction Centre
University College London
Gower Street, London WC1E 6BT
anna.cox@ucl.ac.uk

Duncan P. Brumby
UCL Interaction Centre
University College London
Gower Street, London WC1E 6BT
brumby@cs.ucl.ac.uk

Abstract
For as long as they have existed participant pools made up of undergraduate students at Western universities have elicited debate about their validity. Online crowdsourcing offers researchers a way of overcoming some of the practical issues associated with conducting research with broader more representative samples. But online experiments are not a panacea for the problems with participant pools. For many investigations, international differences in technology, motivation and access threaten to confound measures. It is therefore important to understand the effects of these differences when attempting cross-cultural replications. Some paradigms will be more susceptible to these issues of control than others. We explore briefly this space here.

Introduction
The attention paid to the homogeneity of participant pools has been a source of irritation for both the critics and defenders of classic laboratory-style psychological experimentation. Critics do not see enough change; defenders do not always agree that change is necessary. The lack of sample diversity in many studies has fed into the debate surrounding the replicability of psychological experiments.

The topic of replication has become a mainstay of psychology's discourse in recent years [8]. Although there have been dissenting voices [e.g., 12], there is a consensus that ensuring the reliability of findings from human studies is an important issue. The vogue in psychology for discussing the 'replication crisis' has also influenced the HCI research agenda [e.g., 13]. Researchers often attempt to match the conditions of experiments when conducting replications. But there is also need for a different kind of replication, one where findings are tested across a broad variety of people from around the world. Our understanding of people's behaviour should not be built entirely on research conducted with the small and biased populations that traditional psychology subject pools tend to draw upon.

Crowdsourcing cross-cultural replication

The idea that cross-cultural differences might affect the results of studies is pervasive and often accurate [see 4 for examples]. Conducting cross-cultural replications of studies can give researchers insight into the extent to which their topic of investigation is culturally bound.

Running these kinds of replications is not always straightforward, however. One of the difficulties in conducting these studies is that they often require significant logistical effort to carry out. Given these difficulties, many researchers have found the reach and simplicity of online studies enticing.

Online studies have aided cross-cultural comparisons in the past [e.g., 2,10]. However, despite the success of these efforts we think that caution is required when performing cross-cultural investigations and replications in online settings. The online environment means that there is not only variation in the participants that take part but also in the conditions and context in which people participate: in crowdworking settings people might be using laptops, desktops, tablets or phones. They might be working at home, at a café, in a library or moonlighting from the comfort of their office. This has the potential to introduce confounding factors into experiments. If due care is not given to monitoring these factors then they have the potential to undermine the validity of the conclusions that are drawn.

Online studies and experimental control

We need to be careful when employing online experiments as a shortcut to cross-cultural understanding, particularly when variations caused by the context of participation are not the target of research. Some studies have looked at broad cultural differences in the approaches to particular problems. For instance, Reinecke et al [10] used crowdsourcing techniques to investigate meeting scheduling in different demographics. In studies like this, contextual differences are of interest to the researchers. Indeed, they might be the target of the research.

But investigations of phenomena that might be confounded by the mode of participation are much more problematic. For instance, studies of low-level interactions with user interfaces [e.g., 6] are likely to be sensitive to differences in pointing devices, screen sizes and device speed. These sources of variation do not necessarily align with the demographic variation in crowdworkers [see 11] that makes them a target for cross-cultural comparisons in the first place. If variation over these factors is consistent across demographic groups there is no problem. However, if there are systematic biases then there certainly is a problem. For instance, the relative wealth of different demographic

groups may have a predictable effect on access to devices and connectivity.

Data quality has been a longstanding area of concern in online research [1,9], but these issues are magnified when the goal of a study is to understanding differences in performance based on some kind of demographic split. Of course, differences in working context cannot practically be changed in online environments. So we should instead commit to measuring as much as we can about people's working environments so that when we come to analyses of our data we are better able to detect the effects of confounding factors in our data.

Conclusion

Crowdsourcing replications online can reduce some of the logistical difficulties involved in cross-cultural studies. Online experiments are a reliable source of data for many research paradigms. But this approach also has drawbacks. The context of participation might vary in unexpected ways that can pose a threat to the validity of conclusions. In this short position paper we have explored the potential issues and described some techniques for limiting unintended sources of variation in performance.

Experience of cross-cultural research

We have little experience of cross-cultural research beyond what researchers fall into when crowdsourcing studies. We are also interested in carrying out online replications and many of the issues encountered in cross-cultural work are common to all replications.

Generalizability of studies

We are particularly interested in the generalizability of studies of multitasking behaviour and interruptions. This has been a major area of work in HCI. Many of the studies done have been situated in large Western multinationals [3,5,7]. It would be interesting to see how well these generalize (or indeed if they even should) more diverse settings.

References

[1] Dandurand, F., Shultz, T.R., and Onishi, K.H. Comparing online and lab methods in a problem-solving experiment. *Behavior Research Methods 40*, 2 (2008), 428–434.

[2] Eriksson, K. and Simpson, B. Emotional reactions to losing explain gender differences in entering a risky lottery. *Judgment and Decision Making 5*, 3 (2010).

[3] González, V.M. and Mark, G. "Constant, constant, multi-tasking craziness": managing multiple working spheres. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM (2004), 113–120.

[4] Henrich, J., Heine, S.J., and Norenzayan, A. The weirdest people in the world? *Behavioral and Brain Sciences 33*, 2-3 (2010), 61–83.

[5] Iqbal, S.T. and Horvitz, E. Disruption and recovery of computing tasks: field study, analysis, and directions. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM (2007), 677–686.

[6] Komarov, S., Reinecke, K., and Gajos, K.Z. Crowdsourcing Performance Evaluations of User Interfaces. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM (2013), 207–216.

[7] Mark, G., Voida, S., and Cardello, A. "A pace not dictated by electrons": an empirical study of work

without email. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM (2012), 555–564.

[8] Pashler, H. and Wagenmakers, E.-J. Editors' Introduction to the Special Section on Replicability in Psychological Science A Crisis of Confidence? *Perspectives on Psychological Science 7*, 6 (2012), 528–530.

[9] Peer, E., Vosgerau, J., and Acquisti, A. Reputation as a sufficient condition for data quality on Amazon Mechanical Turk. *Behavior Research Methods 46*, 4 (2014), 1023–1031.

[10] Reinecke, K., Nguyen, M.K., Bernstein, A., Näf, M., and Gajos, K.Z. Doodle Around the World: Online Scheduling Behavior Reflects Cultural Differences in Time Perception and Group Decision-making. *Proceedings of the 2013 Conference on Computer Supported Cooperative Work*, ACM (2013), 45–54.

[11] Ross, J., Irani, L., Silberman, M.S., Zaldivar, A., and Tomlinson, B. Who Are the Crowdworkers?: Shifting Demographics in Mechanical Turk. *CHI '10 Extended Abstracts on Human Factors in Computing Systems*, ACM (2010), 2863–2872.

[12] Stroebe, W. and Strack, F. The Alleged Crisis and the Illusion of Exact Replication. *Perspectives on Psychological Science 9*, 1 (2014), 59–71.

[13] Wilson, M.L., Mackay, W., Chi, E., Bernstein, M., Russell, D., and Thimbleby, H. RepliCHI - CHI should be replicating and validating results more: discuss. *CHI '11 Extended Abstracts on Human Factors in Computing Systems*, ACM (2011), 463–466.