

RepliCHI - CHI should be replicating and validating results more: discuss

Max Wilson, Wendy E. Mackay, Ed Chi, Michael Bernstein, Dan Russell, Harold Thimbleby

▶ **To cite this version:**

Max Wilson, Wendy E. Mackay, Ed Chi, Michael Bernstein, Dan Russell, et al.. RepliCHI - CHI should be replicating and validating results more: discuss. CHI EA '11: Proceedings of the 2011 annual conference extended abstracts on Human factors in computing systems, May 2011, Vancouver, Canada. pp.463–466, 10.1145/1979742.1979491 . hal-01000423

HAL Id: hal-01000423
https://hal.science/hal-01000423

Submitted on 4 Jun 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

RepliCHI – CHI Should be Replicating and Validating Results More: Discuss

Max L. Wilson
FIT Lab
Swansea University
Swansea, UK
m.l.wilson@swansea.ac.uk

Wendy Mackay
INRIA and Stanford University
LRI, Bâtiment 490
Université de Paris-Sud
91405 ORSAY FRANCE
mackay@lri.fr

Confirmed Panelists:

Ed Chi
PARC
Strength: CHI2012 Organizer

Dan Russell
Google
Strength: Industry + Anthropology

Michael Bernstein
MIT
Strength: Systems + Our Future

Harold Thimbleby
FIT Lab, Swansea University
Strength: Science Background

Copyright is held by the author/owner(s).
CHI 2011, May 7–12, 2011, Vancouver, BC, Canada.
ACM 978-1-4503-0268-5/11/05.

Abstract
The replication of research findings is a cornerstone of good science. Replication confirms results, strengthens research, and makes sure progress is based on solid foundations. CHI, however, rewards novelty and is focused on new results. As a community, therefore, we do not value, facilitate, or reward replication in research, and often take the significant results of a single user study on 20 users to be true. This panel will address the issues surrounding replication in our community, and discuss: a) how much of our broad diverse discipline is 'science', b) how, if at all, we currently see replication of research in our community, c) whether we should place more emphasis on replication in some form, and d) how that should look in our community. The aim of the panel is to make a proposal to future CHI organizers (2 are on the panel) for how we should facilitate replication in the future.

Keywords
HCI, Research, Science, Replication

ACM Classification Keywords
H5.2. User Interfaces: Evaluation/methodology.

General Terms
Experimentation, Reliability, Verification

Introduction

Replication of research and the reproduction of results is a cornerstone of science for many disciplines, but in HCI we focus on novelty and reject work that appears to have been done before. We do not reward, or even facilitate the publication of, research that looks back to confirm or challenge the research performed by others. It is important to discuss replication of research in our diverse HCI community for several reasons, including:

1. While focusing on novel work helps us move forwards, without looking backwards we may not be building on solid ground.
2. HCI Research is typically performed on a small sample of the world's population (often CS students!) within a small geographic location.
3. HCI includes a range of disciplines, from traditional science to sociological, ethnographic, design, and qualitative methods that can be hard to replicate.
4. Previous findings, published in the history of CHI, can be measured in their scale of contribution, by whether the results are replicable and applicable.

Background

Replication in research exists in a spectrum where some communities directly replicate work, others replicate and then extend research, and the HCI community rarely replicates work at all. In disciplines like physics, chemistry, medicine and mathematics, research is published so that the community can address the research and investigate the findings. Typically, research is only presumed to be true, and thus built upon, if it can be (at least in principle) replicated by others. While direct replication of research is perhaps not directly published, researchers often first replicate and then extend. If results, during replication of research, are contrary to prior research, then they can be published as amendments. Famously in mathematics, for example, Wiles' proof of Fermat's Last Theorem was proven by several other academics to have errors, which were fixed nearly 18 months later.

This example, however, presents a particular case. A colleague and friend from pure mathematics noted that: "[pure mathematics] is neither an art where different opinions are acknowledged and tolerated, nor a science where experiments can be replicated". HCI, however, can fall into both categories[1]. Art is largely seen as a novel-only forum, where newly recognized modern art *can* be influenced by former work, but itself breaks new ground. We often treat HCI research in a similar way, especially in the narrow focus of UIST. HCI research, however, is also scientific and experimental.

Experimental sciences, such as physics and chemistry, do encourage replication to facilitate progression. Observed phenomena that cannot be easily re-created are often studied for years, with the conditions of any subsequent successes being newly published. Similarly, there are often many scientists working on similar research around the world. In HCI, it is not uncommon to have multiple groups working on one topic, as we discuss later. It is also common in scientific communities for researchers to replicate research before building on it. Again, research that does not confirm prior experimentation is often published. The point here, however, is that as a community we are not encouraged to first replicate and then extend. Instead, we are encouraged to differentiate into novel spaces, design novel interfaces, and run novel experiments that create new insights into human behavior.

[1] The theme of CHI2008 was 'art.science.balance' for example.

In medical science, it is also common for methods to be recreated and subsequently modified. Medical studies, however, are often carefully controlled because of their danger to human life. The consequence of this danger is that studies are planned, performed, and analyzed with extensive rigor. Further, results are only accepted with highly significant evidence, and are often modified by long-term experiences of the community thereafter. In HCI, we often do not have this same *critical* impact on human life, and are not bound by the same rigor.[2]

It is not only other major disciplines that utilize and reward replication. Other areas of computer science leverage the concepts of replication in research. One notable example is the evaluation framework provided by the Cranfield and TREC models [2] in the Information Retrieval (IR) community. TREC provides an environment with set example tasks with known optimal answers, for a single type of search. In TREC, therefore, it is easy for a researcher to check that they have implemented a prior algorithm correctly, before they devise a new one and compare them (again: replicate and extend). In fact, all new algorithms, or algorithmic examples become comparable, when using the same TREC dataset. Unfortunately, the interactive stream of TREC has had limited success, as it is very difficult to model human behavior in a consistent way.

Replication in HCI

When thinking about the concept of replicating HCI research, there are many questions that arise. 1) What exactly is 'replication of research?' 2) Have we been replicating research (in some form on the spectrum) in the past? 3) If not, do we have an equivalent process? 4) Can we even 'replicate' research in HCI? 5) Should we be addressing the ideas of replication as our community matures? And 6) If so, how should we facilitate and support the replication of research in our community's future?

The Potential Advantages of Replication in HCI

It is not the aim of this panel to create a venue where we shoot each other down or criticize methodological choices. Instead, replication of research has many advantages towards strengthening out community outputs. These include:

1. Checking whether results generalize past a sample
2. Checking assumptions made about prior work
3. Facilitating better provision and sharing of resources
4. Improving the quality of method reports in papers
5. Learning from and designing better methodologies
6. Training new researchers in HCI techniques

Examples of Replication in HCI

Despite the questions above about replication in our community as it currently stands, there *are* examples of replication in our history. The Touchstone system, for example, was designed to make studies shareable and repeatable [6]. Before that, some early methods like the cognitive walkthrough and heuristic evaluation, for example, were included in a range of different user studies in different scenarios or cases (e.g. [4, 5]). Although these studies rarely used the exact same methodology, or involved the same user interfaces, they did continue to produce similar results and novel insights into the methods.

As noted in other disciplines, we also see examples of multiple teams around the world studying the same

[2] As the majority of HCI does not study critical systems, we do not have statutory structures for safe rigorous HCI.

problem. Faceted browsing, for example, has been studied many times and from different institutions around the work (e.g. [1, 3, 7]). These projects perhaps replicate the principle of faceted browsing in many forms, by supporting or challenging prior results. To some extent, these examples do show that we produce some form of replication in our community, although rarely do we see direct replication of a study. Below we discuss some of the challenges that have hindered replication in our community.

Challenges for Replication in HCI

Aside from not rewarding or facilitating replicated research, there are many challenges to engaging in replication in our community. These include:

1. Replication often means the sharing of, or redevelopment of, software prototypes.
2. Much of the research in our community is design, qualitative, anthropological, and ethnographic, which can be very hard to replicate.
3. Such qualitative research is often extremely lengthy, expensive, and contextually dependent.
4. Methods and Study descriptions are not always complete in the limited space conference papers.

Conclusions

In this panel, we propose that the community, as it matures, should actively engage in a discussion of whether and how replication of research should become a part of our research methods. We believe that the community should move towards a model shown in many other sciences where we first replicate and then extend research. One aim of this panel is to make a recommendation to the CHI2012 and CHI2013 organizers (the panel contains a representative from both) that we, for example, provide some form of venue that facilitates the discussion of replicated research, so that we can discuss, strengthen, learn from, and improve the study of Human-Computer Interaction. In the long run, however, we hope that our community can begin to inherently acknowledge, reward, and value the replication of prior findings.

Acknowledgements

Thanks to those at CHI2010, and subsequent related events, and to Dr Crossley and Dr Seisenberger, for discussions of replication in HCI and other disciplines.

References

[1] Capra, R., Marchionini, G., Oh, J.S., Stutzman, F. and Zhang, Y., Effects of structure and interaction style on distinct search tasks. In *Proc. JCDL 2007*, 442-451. 2007

[2] Harman, D.K., The TREC conferences. *Morgan Kaufmann Multimedia Information And Systems Series*247-256. 1997.

[3] Hearst, M.A., Next generation web search: setting our sites. *IEEE Data Engineering Bulletin: Special Issue on Next Generation Web Search*, *23*(3), 38-48. 2000.

[4] Jeffries, R., Miller, J.R., Wharton, C. and Uyeda, K., User interface evaluation in the real world: a comparison of four techniques. In *Proc. CHI 1991*, 119-124. 1991

[5] Karat, C., Campbell, R. and Fiegel, T., Comparison of empirical testing and walkthrough methods in user interface evaluation. In *Comparison of empirical testing and walkthrough methods in user interface evaluation*, 397-404. 1992

[6] Mackay, W.E., Appert, C., Beaudouin-Lafon, M., Chapuis, O., Du, Y., Fekete, J.-D. and Guiard, Y., Touchstone: exploratory design of experiments. In *Proc. CHI 2007*, 1425-1434. 2007

[7] schraefel, m.c., Wilson, M.L., Russell, A. and Smith, D.A., mSpace: improving information access to multimedia domains with multimodal exploratory search. *Communications of the ACM*, *49*(4), 47-49. 2006.