

Call for Papers

Workshop on Methodology and/of Evaluation of IVAs “Replicating the replication crisis”

Merijn Bruijnes, University of Twente, the Netherlands, ✉ m.bruijnes@utwente.nl

Ulysses Bernardet, Aston University, United Kingdom, u.bernardet@aston.ac.uk

Willem-Paul Brinkman, Delft University, the Netherlands, w.p.brinkman@tudelft.nl

Deborah Richards, Macquarie University, Australia, deborah.richards@mq.edu.au

Website: <https://iva2018methodologyworkshop.wordpress.com/>

The aim of the workshop is to critically but constructively discuss the empirical methods that are used in Human Computer Interaction, specifically in the area of Intelligent Virtual Agents. The social and life sciences are in a crisis of methodology as the results of many scientific studies are difficult or impossible to replicate in subsequent investigation (e.g. Pashler & Wagenmakers, 2012). The Open Science Collaboration (2015) observed, for example, that the effect size of replications was about half of the reported original effect size and that where 97% of the original studies had significant result, only 39% of the replication studies had significant results. In fact it has been suggested that more than 50% of psychological research results are likely false (i.e. theories hold no or very low verisimilitude) (Ioannidis, 2005). Many of the methods employed by HCI researchers come from the fields that are currently in a replication crisis. Hence, do our studies have similar issues?

Long before the replication crisis hit psychology, Meehl (1990) suggested ten obfuscating factors that make that research on psychological theories are often uninterpretable. Viewing these factors gives us an idea of the scope of the problems that our research methodology might face:

1. Loose derivation chain: Very few derivation chains running from the theoretical premises to the predicted observational relation are deductively tight;
2. Problematic auxiliary theories: each auxiliary theory is itself nearly as problematic as the main theory we are testing;
3. Problematic ceteris paribus clause;
4. Experimenter error;
5. Inadequate statistical power;
6. Crud factor: everything correlates with everything;
7. Pilot studies: A true pilot study is a main study in the small. But these are often not published which can lead to line of research being dropped;
8. Selective bias in submitting report;
9. Selective editorial bias;

10. Detached validation claim for psychometric instruments: claiming a measure is 'valid' without further consideration;

A variety of ideas to improve research practices have been proposed and it is likely these ideas can be beneficial to the methods used in the field of HCI. Some actionable points leading to open and reproducible science are pre-registration of experiments, replication of findings, collaboration and education of researchers. The replication crisis needs our attention and as we reflect on our methods it makes sense to discuss in general our scientific methods and practises.

In the past years there has been limited activity on this topic at IVA conferences. A workshop aimed at improving the quality of IVA research and methods should be welcomed by all IVA researchers. During the workshop we will discuss the methodological challenges identified in other fields and how they relate to the methods we use in our field. Additionally, we will discuss the proposed remedies and whether these are applicable for the research we conduct. We will discuss whether questions such as those posed above are relevant and, if so, how to go about answering them. This workshop is intended as a starting point and it will be the first of a series of workshops (at IVA and other conferences in the field) on this topic.

The goal is to embrace a positive, proactive approach that is sustainable and will lead to better science (no naming and shaming). The idea is to foster discussion and one way to achieve this is by having provocative statements to respond to. We invite participants to submit thought provoking statements about the methodology in HCI and/or respond to statements that we propose. Participants can posit their statements and/or discussions in an **extended abstract (max 3 pages, excluding references)**.

Papers can be submitted via e-mail to: m.bruijnes@utwente.nl

Important dates:

- 1 October – Submission deadline
- 15 October – Acceptance notification
- 5 November – Workshop

Ten provocative statements to start the discussion:

1. HCI research is too much novelty focussed.
2. Sample size estimation is impossible to do when evaluating new technology.
3. Experimental design/methodology are seen as necessary evil & boring by HCI researchers.
4. Theory building is difficult because of technical implementation of auxiliary hypotheses.
5. Knowledge of theories and concepts is insufficient in HCI (e.g. Basic emotions) and failures do not lead re-evaluation of assumptions.
6. HCI relies on small corpora and ground truth does not exist.

7. Technology focus creates legacy problems.
8. Custom/proprietary technology prevents accurate replication.
9. Open science is prevented by novelty focus, technical one-off solutions, and conference schedule.
10. Valorisation & entrepreneurialism, for which HCI is a key field, are at odds with proper conduct of science.

Participation is encouraged for all who are interested in good science. Contributions are welcome discussing methodology in HCI and/or related to the following topics:

- Replicability of studies;
 - Methodological pitfalls specific to HCI;
 - Tools and procedure that can improve the replicability;
- Validity of HCI research;
 - What are we investigating (are the definitions clear)?
 - Do we agree on definitions and what we are investigating?
 - Are we asking the right questions?
 - What are the answers worth?
- Generalisability of results;
 - From theoretical background to concrete predictions;
 - Relating data from HCI experiments back to theory (e.g. Gratch, 2017);

Tentative Schedule

09.00 – 09.30 Conference registration and welcome coffee

09.30 – 09.45 Introduction by Organizers

09.45 – 10.45 Keynote: Alex Holcombe (University of Sydney)

Methodological issues and solutions

10:45 – 11:00 Coffee break

11:00 – 12:00 Keynote: TBA

Tentative: Measurement instrument creation

12:10 – 13:30 Lunch break

13.30 – 15.00 Presentations of accepted abstracts (15 x 5 min)

15.00 – 16.00 Panel discussion on the directions to take towards better methodology

16.00 – 16.15 Coffee break

16.15 – 17.00 Breakout Session on 'measurement instrument creation'

17.00 – 17.30 presentations of breakout sessions & finishing notes by organizers

17.30 – 20.30 Social

Organizing Committee

Merijn Bruijnes (University of Twente)

Merijn Bruijnes is a Post-Doc researcher at the Human Media Interaction group at the University of Twente. He holds a masters in cognitive psychology and ergonomics and a doctorate in human media interaction. His research interests include, but are not limited to, artificial social agents, dialogue systems, and the effect technology has on humans.



Ulysses Bernardet (Aston University)

Ulysses Bernardet is a Lecturer in Computer Science at Aston University, Birmingham, UK. He holds a doctorate in psychology from the University of Zurich and has a background in psychology, computer science and neurobiology. In this research, Ulysses develops biologically grounded computational models of cognition and emotion that are used to control the behaviour of virtual and physical agent.



Willem-Paul Brinkman (Delft University)

In 2003 Willem-Paul Brinkman (1970) received his PhD degree in the area of human computer interaction from Eindhoven University of Technology, The Netherlands. His primary research interests are human-computer interaction, behavior change support systems, specifically eHealth systems including virtual reality therapy systems, and virtual health agents. He is fascinated by eHealth systems that include conversational agents that offer psychological support.



Deborah Richards (Macquarie University)

Deborah Richards is a Professor in the Department of Computing at Macquarie University. She joined academia in 1999 after nearly 20 years in the ICT industry during which she completed a BBus, MAppSc and a PhD in knowledge acquisition and reuse at UNSW. As an applied researcher and human-centred practitioner, her work is multidisciplinary and draws on theories from the cognitive science, biology, sociology, learning sciences, medicine, psychology and computer science to build and evaluate intelligent virtual agents to improve human learning and well-being.



Program committee

Douglas W. Cunningham (Cottbus)

Bibliography

Gratch, J. (2017). Understanding the mind by simulating the body: virtual humans as a tool for cognitive science research. (S. E. F. Chipman, Ed.) *The Oxford Handbook of Cognitive Science (Vol. 1)*. Oxford University Press.

Ioannidis, J. P. (2005). Why most published research findings are false. *PLoS medicine*, 2(8), e124.

Meehl, P. E. (1990). Why summaries of research on psychological theories are often uninterpretable. *Psychological reports*, 66(1), 195-244.

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251), aac4716.

Pashler, H., & Wagenmakers, E. J. (2012). Editors' introduction to the special section on replicability in psychological science: A crisis of confidence?. *Perspectives on Psychological Science*, 7(6), 528-530.