
Is replication important for HCI?

Christian Greiffenhagen
Loughborough University
c.greiffenhagen@lboro.ac.uk

Stuart Reeves
University of Nottingham
stuart@tropic.org.uk

Abstract

Replication is emerging as a key concern within subsections of the HCI community. In this paper, we explore the relevance of science and technology studies (STS), which has addressed replication in various ways. Informed by this literature, we examine HCI's current relationship to replication and provide a set of recommendations and points of clarification that a replication agenda in HCI should concern itself with.

Author Keywords

Replication; psychology; science and technology studies; philosophy of science.

ACM Classification Keywords

H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

Introduction

Replication is emerging as a concern within subsections of the HCI community. A key motivation for this is a feeling that HCI emphasises *novelty* over *consolidation* of research; consolidation that can be achieved *via replication*. In response, we advocate the relevance to HCI of understandings of 'replication' emerging from the philosophy and sociology of science and technology. This paper highlights a collection of rejoinders to the ways in which this programme for replication is currently conceptualised within HCI. In doing so we intend to help the development of an endogenous

understanding of replication as a practice that can be a) motivated, b) mature and c) fit for the purposes of HCI.

Replication: Lessons from STS

We believe that debate on replication in HCI can be enriched by STS and philosophy and sociology of science. In this section we review some of the findings of this literature and their pertinence to HCI.

One of the motivations for replication within HCI is the wish to make HCI more scientific by modelling HCI on other sciences (e.g., "psychology, physics and medicine" [11]). While there is nothing problematic in asking for a field to involve more replication, to frame this in terms of making it more 'scientific' is possibly based on a mythical view of 'good science' of which "[r]eplication of research is a cornerstone" [11]. This view suggests this 'science' may be a homogenous practice, possibly even based around a particular method, 'the' scientific method. It also tends to think about replication more from the perspective of the *philosophy* of science, rather than the *practice* of different sciences.

In contrast, philosophical and sociological studies have shown that 'science' refers to a fragile structure of multiple disciplines and multiple methods linked by 'family resemblances' only [9, 3]. Not all empirical sciences work with experiments, and the role of experiments may differ between different fields.

Complicating this picture is the separation between these varied and autochthonous scientific practices and their rendering into literature. Scientific literature is written in such a way as to promise replicability, emerging from Boyle's attempts to create scientific records that were

publicly accountable and would let 'anyone' replicate experimental practices [10]. However, the nature of instructions is such that they are always incomplete [4], thus scientific instructions must be 'filled in' by competent members of the target scientific community in order to enact them as replications. This is one of the reasons why Medawar characterised the scientific paper, somewhat misleadingly, as a 'fraud' [8].

STS reports an alternative view on the nature of replication in the natural sciences to the surface view of scientific replication where scientific articles (in particular: their 'method sections') provide an adequate instruction manual for replication work. Specifically it problematises the notion of a 'decisive experiment' or by extension a 'decisive replication'. At the heart of this problem is what Collins calls the "experimenter's regress" [1], that is, a circular relation between experimental findings and the instruments used to produce them. Reliable experimental findings themselves rely upon reliable instruments and vice versa. As a result, a key difficulty of replication and the experimenter's regress is that, particularly for contested science, there is not necessarily any standard for what is to be considered a valid replication. This raises a principle problem, since it is not clear whether a 'failed' replication is due to a problem with the original experiment or the subsequent replication ("it is often hard to tell whether an inability to replicate a result is due to a group's failings or a flaw in the original paper" [5, p. 345]).

Further to this, when we consider the track record of replication in the natural sciences, STS literature argues that replication in the (natural) sciences employs replication for specific, highly motivated and reasoned

ends. Thus we find a marked absence of large amounts of replication in the sciences unless we focus on particular issues [1, pp. 210-211]. For instance, Collins' tracing of the construction of gravitational wave detectors during the 1970s reveals the relevance of replication as an activity for working through what was a contested, controversial domain [2]. In short, 'doing replication' is not always seen as a fundamental prerequisite for valid scientific practice, since a vast number of results go unreplicated: instead it emerges as the result of pragmatic action for specific contested cases.

In summary, then, our cursory examination of STS and its related literature highlights that: a) there is no singular form of science or scientific method upon which to model; b) there is no 'algorithmic' method for replicating directly from scientific literature (indeed, this is not its purpose); c) 'absolute' security of results is problematic in light of the experimenter's regress; and d) sciences often do not involve replication as a 'matter of course', it being difficult and of little value unless motivated (typically via contestation of results).

Replication within HCI

This issue of replication has become a centre of discussion within HCI. In light of STS's view on replication, we seek to ask what is at stake in this discussion. Why replicate? Or: What are the (different) *aims and motivations* for replication?

Within HCI, it has been acknowledged that there is not just one kind of replication. For example, Wilson et al. distinguish between four forms: "direct replication" ("driven by the aspirations of strong science"), "conceptual replication" (replication via "alternative methods"), "replicate & extend" (building on prior

studies incrementally) and "applied case studies" (replication through application of prior work) [11].

Nuancing this view, we want to start with introducing two different kinds of distinction to help us to think about replication.

The first distinction is between what we characterise as *textbook replication* and *frontier replication*. By 'textbook replication' we refer to replications of well-known studies that are conducted from HCI textbooks, typically as part of undergraduate or graduate education. For instance, these could be replications of well-known usability studies. We distinguish this from 'frontier replication' by which we mean replications of 'ongoing' or 'recent' studies. We see these forms as conceptually and practically incommensurate, as opposed to integral facets (e.g., see position in [11] on "Benefits of Replication"). Thus, while the primary aim and motivation of textbook replications is *learning*, the point of frontier replication is often a form of 'checking' (which may even be done during the review process). As such we argue that the activities at this 'frontier' becomes the main issue for replication rather than what is happening 'in the textbook'.

A second distinction has to do with what may be *replicable* and what is actually *replicated*, in which the aims for each are quite different. 'Being replicated' concerns the 'factual' question of whether a particular study has, actually, been replicated by other researchers or not. We say 'factual' since subsequent studies may or may not be seen as valid replications, as in Collins' study of gravitational wave detectors [2]. We also note again that a lack of actual replications may be related to matters such as experiments being too costly, too time

consuming or lacking in providing the experimenter any obvious credit.

In contrast 'being replicable' is motivated by the 'in principle' possibility of some other researcher being able to replicate an empirical study. This is often cited as one of the differences between 'quantitative' and 'qualitative' methods (very problematic descriptions themselves), where the former supposedly produce results that *could* be replicated (again, 'in principle'), while the latter are not. For instance, ethnographic research is often said to be too reliant on the 'subjective' insights of the ethnographer, resulting in non-generic and non-replicable findings.

What's at stake in this distinction? We would argue that the issue of 'being replicable' concerns a foundational question, in particular, whether HCI is a science and its preferences for particular methods over others. These questions are not new: psychology—which has strongly informed HCI's development—has repeatedly foregrounded replication as an explicit agenda, such as in response to perceived experimental biases (e.g., being too 'WEIRD' [6]), as well as intentional and unintentional misconduct [12]. In this sense, 'being replicated' is probably more common in psychology than many other sciences because of this explicit concern (now displayed in HCI) for the lack of actual replicated studies (or those 'seen as' validly replicated).

Psychology's own debates around its status as a science are also consonant with these foundational concerns of 'being replicable', and in the replication agenda we see HCI grasping towards key epistemological themes which arise in the natural sciences: alongside 'observation', 'measurement',

'description' and 'reasoning' is, of course, 'replication'. If we take HCI as a scientific endeavour (e.g., [11]) then it follows that its concern for replication would thus be informed by this particular picture of 'normal science'; or 'doing what scientists do'. However, this assumes coherence of 'science' as monolithic practice as well as mythologising that practice.

In contrast, 'being replicated' is a more pragmatic question, which concerns what we can *learn* from replications and, for example, whether it would be worthwhile to publish more papers based on replication.

In order to focus the discussion of replication in HCI, it would be very helpful if one could gather more examples from different disciplines, from biology to physics, to see whether and how replications are valued in these. Thus we hazard a conjecture: that replication enjoys a special status within psychology (and the debate of replication in HCI is thus a reflection of the influence of psychology, rather than, say, biology, in HCI). But why might that be?

One issue is with the *scale* of the question to be answered through experiment. Some sciences tackle very detailed and small questions through extremely detailed experiments. In other words, there exist a very tight relationship between the data gathered through the experiments and the derived conclusions. Other sciences (e.g., social science) tackle bigger questions and consequently involve a looser relation between data and conclusion.

We would argue that there is a 'scale' tension in psychology—and thus HCI—between tackling 'big' and 'minute' questions, questions that can, or can't be

settled through experiments. One possible reason for more replication in psychology is that studies can be questioned more (i.e., findings are more open to interpretation).

Discussion

We have raised some broad issues in the relationship between replication and HCI, and informed this debate through recourse to existing work in STS that has explored replication in the natural sciences.

Firstly we argue for the importance of the increased consultation of literatures normally foreign to HCI such as that of STS. This is particularly the case for situations where knowledge within the field is out of step with more recent advances in understandings of scientific knowledge. For instance, our discussions on replication (and science) within HCI are largely Popperian or pre-Popperian in form, such as appeals to ideals such as falsificationism. While we would not argue against such ideals, we contend that understanding benefits from expansion, thus as well as citing Collins, we might also refer to developments by Kuhn, Feyerabend or Lynch that, for instance, encapsulate empirical investigations into practical mundane scientific action [7].

A fundamental question for the desire for replication in HCI is that of the motivation to perform replication in the first place. We need to ask ourselves *why we might bother* with replication in the first place and whether there is any value gained from pursuing a replication agenda as a distinctive activity within HCI (which is the position of the workshop call [11]). As we have seen from STS literature, if we feel the need to derive HCI's programme from the methods and epistemological

topics of the natural sciences (e.g., via psychology), then we must do so knowingly in light of findings from STS. Thus we argue for different understandings of replication: a) as an unstable and negotiated practice; b) as a highly motivated activity rather than as an end of itself; and c) as playing an important role in the resolution of scientific controversies. Moving forwards we would draw attention to the judicious motivated application of replication—and the need for 'just why' and 'just how' it is to be pursued. So, we must be clear about the purposes and motivations of any given replication beyond abstractly "validating and understanding contributions" [11].

Finally, we have argued that a mythological view of science tends to be implicit in HCI regarding its status as scientific. This leads us to question the value in positioning HCI as a scientific endeavour. Thus we recommend that it would be helpful to separate the 'foundational' question (whether HCI is a science) from the above 'pragmatic' question (about the specific benefits of replication for HCI).

Acknowledgements

This work is supported by Horizon Digital Economy Research, RCUK grant EP/G065802/1.

References

- [1] Collins, H. M., *Changing Order: Replication and Induction in Scientific Practice*, Beverly Hills & London: Sage, 1985.
- [2] Collins, H. M. The seven sexes: A study in the sociology of a phenomenon, or the replication of experiments in physics. *Sociology*, 9(2):205-224, 1975.
- [3] Dupre, J. The disunity of science. *Mind* 112, 321-346, 1983.

- [4] Garfinkel, H. *Studies in Ethnomethodology*. Prentice-Hall, 1967.
- [5] Giles, J. The trouble with replication. *Nature*, 442:344-347, July 2006.
- [6] Henrich, J., Heine, S. J. and Norenzayan, A. The weirdest people in the world? *Behavioral and Brain Sciences*, 33, pp. 61-83, 2010.
- [7] Lynch, M. *Scientific Practice and Ordinary Action*. Cambridge University Press, 1993.
- [8] Medawar, P. B. Is the scientific paper a fraud? *The Listener*, 70 (12 September): 377–378, 1963.
- [9] Putnam, H. The idea of science. *Midwest Studies In Philosophy*, 15(1):57-64, 1990.
- [10] Shapin, S. and Schaffer, S. *Leviathan and the Air-Pump: Hobbes, Boyle, and the Experimental Life*. Princeton University Press, 1989.
- [11] Wilson, M. L., Resnick, P., Coyle, D. and Chi, E. H. RepliCHI—The Workshop. In *CHI '13 Extended Abstracts (CHI EA '13)*. ACM, New York, NY, USA, 2013.
- [12] Yong, E., Replication studies: Bad copy. *Nature*, 485(7398):298-300, 2012.