The Weak Science of Human-Computer Interaction.

Saul Greenberg
Harold Thimbleby

1992

Cite as:

Greenberg, S. and Thimbleby, H. (1992) “The weak science of human-computer interaction.” In *CHI '92 Research Symposium on Human Computer Interaction*, Monterey, California, May.

An earlier working paper was published as Research Report 91/459/43, Department of Computer Science, University of Calgary, Calgary, Canada (1991).

The Weak Science of Human-Computer Interaction

Saul Greenberg
Department of Computer Science, University of Calgary, Canada T2N 1N4

Harold Thimbleby
Department of Computer Science, University of Stirling, Stirling, Scotland

Abstract

This article is about science and the discipline of human-computer interaction (HCI). Science in HCI is merely one component of a wider agenda; alone science is not sufficient for 'good' HCI (whatever that is). We argue that science is necessary, but the way that science is undertaken—or purported to be undertaken—in HCI is inadequate. Failures are due to the sparsity of theories and risky hypotheses, the pragmatic difficulty of substantiating experiments through replication, and the over-generalization of experimental results.

1. Introduction

The discipline of human-computer interaction is concerned with the study of human factors in computing systems. As it now stands, it is mainly an *engineering* discipline that concerns the construction of 'better' interfaces. A listing of the areas covered by HCI journals, conference proceedings, and books support this statement. There is a creative component pushing newer innovative inventions, as evident in the proliferation of novel interaction techniques (eg virtual environments). Interface techniques are refined and contrasted to see which leads to improved human performance (eg broad versus deep menu hierarchies). Computer algorithms are developed for making our machines smarter and better partners (eg intelligent agents). Tools for building interfaces are becoming more sophisticated (eg user interface management systems and interface builders). Techniques are presented for improving the interface design process (eg participatory design). Practitioners tell folklore and anecdotes of their experiences so that all can learn from their successes and failures.

Human-computer interaction is also a *scientific* discipline. Mathematical theories predict human performance when interacting with computers (eg Fitts Law, the keystroke and GOMS models). Theories of cognitive processes are abstracted (eg mental models, software comprehension). Humans as a diverse population are studied (eg individual differences). Methodologies furthering a scientific approach are presented (eg Wizard of Oz).

Science is clearly necessary in HCI. With good theories, we could do away with much of the trial and error now required in HCI engineering. Yet good theories are hard to come by and hard to evaluate. Certainly, human-human interaction is one of the most complex phenomena we humans know. Human-computer interaction is at once easier and harder than human-human interaction. It is easier because one agent is merely a computer; it is harder in that the computer imposes a delay between the design of the interaction and the actual interaction, but furthermore the computer appears to act autonomously and becomes a confounding component in the system.

This article is about the science in HCI. It should be clear that science in HCI is merely one component of a wider agenda; alone science is not sufficient for 'good' HCI. We argue that science is necessary, but the way that science is now undertaken—or purported to be undertaken—in HCI is inadequate.

2. What is science?

Science is a method of investigation, of reasoning and of communication: human-world interaction (investigation), intra-human (reasoning), and inter-human (communication). Science is one of several approaches to improve our understanding of the world in which we find ourselves. "Improving our understanding" involves forming theories that explain natural phenomena and allow us to predict future acts in the world: in short, science defines and clarifies how we *trust* the world.

2.1. Communication and replication

Science has been developed as a systematic and rigorous method for handling the complexity in the natural world and for refining our ability to predict its behavior through theories. Methods for examining theories include hypothesis testing, experimental replication, even anarchic ways of generating new theories (paradigm shifts). Science involves communication since scientists benefit from each others' work only to the extent it is communicated freely: this assumes that scientific information (both theories and evidential matter) can be replicated, otherwise there is no point communicating it. The Proceedings of the National Academy of Sciences USA is an exemplary organ of scientific communication. It requires that writers of publications relating to bacteria, antibodies, DNA or computer programs make their work available to all qualified investigators. The Proceedings of the Royal Society, and Nature make similar demands on its authors.

Is replication really necessary? First, if a paper is to say something worth communicating, it must be novel or unusual, and unusual claims should be confirmed. Popper and Eccles (1977), for example, insist that a good scientific idea should be falsifiable and forbid as much as possible—confirmations of the idea should only count if they are 'risky' ie that they would refute the idea (Thimbleby 1990, ch 9). Unfortunately, the easiest way to produce an unusual result is by fabrication or by an experimental flaw. The journals above therefore require potential substantiation, where the community of scientists are able to check claimed results. While replications are rarely performed for 'supportive' claims of an idea, extraordinary claims demand extraordinary testing (a good example is the recent controversy over cold fusion). Second, scientific papers often generalize a result beyond the experiment; predicting the world is, after all, the main purpose of science. However, other scientists must be able to replicate the experiment under different experimental factors, for they must attempt to falsify the theory under changed or new (and risky!) conditions.

2.2. Science and HCI design

Since computers are in the world, they and HCI are an object of scientific interest. But why worry about the quality of science in HCI? Formalism, rigor, replication and statistical significance seem remote from the rush to push products out the door, and from the aesthetic experience of using systems. This is true, but that is not all there is.

When we design we are making laws about the future behavior and use of a system. The problem of design is to determine what laws would be best to embed in an interactive system. If 'science' is an attempt to understand the real world, design is a 'synthetic science' that attempts to build a world (the system) that can be understood (by the user as a scientist). Indeed, systems are now built to encourage their users to learn and understand them through exploration and experimentation. By better understanding science, designers as scientists can search the literature for theories that they can apply to their problem, and when necessary they can use the scientific methods to design experiments that test the applicability of theories or other experimental results to their particular cases.

There is also a moral need for science. If a user chooses not excitement but trust then they have a right to expect that the best methods have been used to ensure that the product purchased will behave well. An office worker who has paid for a word processor assumes that the expensive package will fulfil the claims made for it, which gave it the apparent worth. If the package turns out to be capricious and unpredictable to the user, it fails a contractual commitment—however much the software vendors may attempt to deny it in their so-called software warranty. Our point is that sometimes (rather often in our view) science is not an option but a necessity in HCI. Because science is concerned with predictability, a human-computer system founded on science should, in principle, exhibit more predictable behavior than one founded on intuition or engineering principles alone.

3. Weak science?

We have argued in the previous section that science is a necessary component of HCI. We do not expect this to be hotly contested; indeed most HCI authors and publishers consider their papers to be scientific ones. Yet we believe that much of the purported science is weak.

3.1. The problem of theories and generalization

Many HCI papers report on experiments that gather facts and/or test very simple hypotheses. They appear scientific. They employ scientific methods for data collection and hypothesis testing; they use experimental design and quantitative methodologies; they are written in a style familiar to scientists. Yet most papers do not present an underlying theory, thus the hypotheses reported and the data gathered do not really have any context outside of the experiment. While the experimenter will often try to generalize the results through extrapolation, there is little scientific basis to believe that the generalizations are useful as a broad theory. As mentioned previously, confirmations of a good scientific idea (a generalization) should only count if they are risky. In contrast, generalizations made from experiments are usually self-promoting (ie "we expect the experimental results should also apply under the following conditions").

Without the underlying theories, how can a reader safely generalize the results of the experiments? Given the huge differences between interfaces, computers, people, and tasks, could an experimental result from one set of conditions really apply to another? The pragmatics of interface design and the rush to deliver products force us to make these leaps of faith. While the consequences are often positive, we should recognize that what we are doing is not scientific. Rather, we are just replacing pure introspective intuition with a more informed style of intuition.

What can we do? First, we should recognize that the best HCI papers will present a theory that is falsifiable and forbid as much as possible, and that any experiments will make best efforts to try to refute the theory. This means that the HCI community should favor research (and publication) of theory-based works. While data gathering and hypothesis testing for its own sake is certainly important for helping us make short-term human factors decisions on system design, it should be considered a lesser contribution than theoretical papers. Second, we as experimenters should be self-critical. If we do try to make generalizations from experiments that do not have an underlying theory, we should include not only supportive claims, but adverse ones as well. The experimenter (and the community) should propose and pursue risky confirmations to the experiment. In plain words, we should resist self-adoration and stop being so nice to each other. The most important feature of a scientific idea is that it is prepared for, indeed encourages, its own criticism, testing, refutation, and eventual replacement.

3.2. The problem of replication

The importance of replication in science was raised in

Section 2.1. Yet many papers in HCI, while accurately describing results, keep the methods by which they were obtained inaccessible to the reader. True replication is rarely possible.

Theory-based experiment does not require exact replication, for the theory should explain how experiments will behave under different conditions. Because most HCI papers report on experiments without an underlying theory, a replication can only occur if the exact experiment were repeated, for there is no real means to predict the effect (and the relevance) of even minor experimental variations. This exacerbates the problems of replicating scientific work in HCI. First, the way results are obtained often depend on circumstances beyond the reader's ability to replicate. The experiment may insist upon possession of a proprietary operating system, software package, or computer; it may use some unavailable prototype system that was developed by the researcher. Second, even if the software used is publicly available, it often has a very short life-span. It may not run on newer releases of an operating system; its characteristics may change in unfavorable ways (eg response time); source may not be available to customize it to local conditions. Third, there is rarely enough information presented in HCI publications to allow replication; page limitations and the author's (and audience's) desire for a snappy presentation make the experimental description a prime candidate for pruning. Also, it is extremely difficult to describe interactive systems in words, making the task of 'reconstructing the interface' a difficult one. Fourth, only summary data or interpretation are presented in papers, and raw data collected by the experimenter is rarely available. Given that it is difficult to regenerate the data through replication, the research community should at least have the opportunity to verify that the data was processed correctly. In summary, the activity of HCI, to the extent that the underpinning science cannot be replicated, is unsound

What can we do? First we should make sure that our experimental circumstances and artifacts are accessible to the community. Unless it is a critical factor in the experiment, we should use stable and widely available hardware and software platforms. If a home-grown prototype system is being used, it should be available to serious researchers wishing to replicate results. Given the existence of the INTERNET and file transfer programs (ftp) in the computer science community, this is now physically easy to do. At the very least, we should be willing to release our software under restricted (even licensed) conditions. Second, editors and referees of archival publications should insist that papers provide enough detail for replication, or that clear reference is made to sources of the gory details (eg technical reports and ftp addresses). Third, descriptions of interactive systems critical to the experiment should be available by videotape or similar media. While there are recognized video publications (such as the SIGGRAPH Video Review), it is a simple matter to produce a low-quality video of a screen and make it available as a technical report. Even a poor quality video is better than none. Fourth, any data collected should be kept and made available. If the data is useful beyond the scope of the current experiment, it should be disseminated in its own right (again, ftp is a good vehicle).

4. Summary

The study of HCI has two fundamental components: engineering and science. We do not champion one approach over the other; rather we recognize the contributions of each. However, we do insist that science is necessary in HCI, and that much of what is now purported to be science in the HCI literature is weak on several grounds: the sparsity of theories and risky hypotheses, the pragmatic difficulty of substantiating experiments through replication, and the over-generalization of experimental results. If it is true that there is little science in HCI and that the science present is weak, we might do well to encourage it.

References

Popper, K. R. and Eccles, J. C. (1977) *The self and its brain*, Springer International.

Thimbleby, H. (1990) *User interface design*. Frontier Series, ACM Press, Addison-Wesley Publishing Company, New York.