
What Is Good? – A Comparison Between the Quality Criteria Used in Design and Science

Christoph Bartneck

Department of Industrial Design
Eindhoven University of Technology
Den Dolch 2, 5600MB Eindhoven
The Netherlands
c.bartneck@tue.nl

Abstract

The human-computer interaction community is an umbrella for many disciplines. Conflicts occur from time to time, in particular between scientists and designers. This article compares the quality criteria used in design with those used in science, in order to gain insight into what design can contribute to the development of science. From the scientific perspective, the weakest point of design knowledge is its limited generalizability.

Keywords

Design, science, quality, criteria

ACM Classification Keywords

H5.0. Information interfaces and presentation (e.g., HCI): General.

Introduction

The human-computer interaction (HCI) community is diverse. Academics and practitioners from science, engineering, and design contribute to its lively development, but communication and cooperation between the different groups is often challenging. At times, open conflicts between the different groups emerge, in particular between scientists and designers, since they have the least common ground [2]. The Computer Human Interaction (CHI) conference of the

Bartneck, C. (2008). What Is Good? – A Comparison Between The Quality Criteria Used In Design And Science. Proceedings of the Conference on Human Factors in Computing Systems (CHI2008), Florence pp. 2485-2492. | DOI: 10.1145/1358628.1358705

Association for Computing Machinery (ACM), which is the largest and arguably one of the most important conferences in the field, is organized through the Special Interest Group Computer Human Interaction (SIGCHI). At the 2005 SIGCHI membership meeting, discussion of the CHI2006 conference ignited a shouting match between academics and practitioners [1]. At the conference itself the conflict recurred in the "Design: Creative and Historical Perspectives" session. Paul Dourish took the role of defending the science of ethnography against its degradation to a service for designers [6]. Next, Tracee Verring Wolf and Jennifer Rode defended creative design against the criticism of scientists by referring to design rigor that is as critical as scientific rigor [12]. Both groups felt the need to defend themselves, which indicated that both had the feeling of being under attack to start with. Stuart Feldman, the president of the ACM, wrote another chapter in this conflict. In his opening speech at the CHI 2007 conference he made an astonishing statement about the HCI community:

"It is also wonderful to have a group that is absolutely adherent to the classic scientific method. Not a description, I am afraid, of all the fields in computing."

However, it is obvious that the methods used by the HCI community are as diverse as its members. So, by emphasizing the classical scientific method above all other methods, Feldman was expressing the ACM's expectation of what methods the HCI community should use. This preference for the scientific method also manifests itself in the division of the CHI proceedings into "main conference proceedings" and "extended abstracts". The main proceedings are considered to be of higher quality and they include a high proportion of

scientific studies. Non-scientific studies, such as experience reports and case studies, are more often found in the extended abstracts. Furthermore, the main proceedings use the "archival format" whereas the extended abstracts do not. The omission of the term "archival" from the format of the extended abstracts suggests that these publications are not important enough to be archived. However, both types of publications are being stored in the ACM digital library, which turns this distinction into a symbolic gesture. At the risk of oversimplification, it can be observed that scientific studies are more highly regarded and hence published in the archival main proceedings, while non-scientific studies are less highly regarded and are published only in the non-archival extended abstracts. But why would the designers bother about this division? Their main focus is on improving society directly through the invention of artifacts, and not through writing papers.

Even though science is highly esteemed, Chalmer [3] argued that "there is no general account of science and scientific method to be had that applies to all sciences at all historical stages in their development". Cross, Naughton & Walker [4] even suggested that the confusing epistemology of science may be unable to function as a blueprint for the epistemology of design. Levy [8] then suggested that transformations within the epistemology of science should be seen as active growth and development, and that they should be considered as providing an opportunity for design to participate in its ongoing improvement. As a matter of fact, any person can contribute to the growth of science. It is an old rule of logic that the competence of a speaker has no relevance to the truth of what he says. The world's biggest fool can say the sun is shining, but that doesn't

Bartneck, C. (2008). What Is Good? – A Comparison Between The Quality Criteria Used In Design And Science. Proceedings of the Conference on Human Factors in Computing Systems (CHI2008), Florence pp. 2485-2492. | DOI: 10.1145/1358628.1358705

make it dark outside [9]. Designers and engineers can discover new knowledge without applying the classic scientific method or becoming a scientist. The more important question is how valuable this new knowledge is and how efficient their methods are in finding it. In this paper I would therefore like to discuss criteria that serve to assess the quality of knowledge. If design wants to make a contribution to science, then its insights must be judged against these criteria. By comparing the quality criteria of science with those of traditional design, the similarities and differences of the respective communities will become apparent. This comparison may also provide insights into the direction in which design methods have to evolve to become more scientific.

This comparison of quality criteria does not imply that design should use the classical scientific method. Cross provided an excellent historical review of the developments in the various design methodologies [5]. He attested to a healthy growth in the field during the 1980s. The design community may continue to define its own method to turn itself into design science, as was attempted at the CHI2007 workshops on "Converging on a science of design through the synthesis of design methodologies" and on "Exploring design as a research activity".

Before diving into these topics, it appears necessary to clarify the terminology of this paper. The different interpretations of the word "research" alone account for considerable friction between designers and scientists. Scientists can barely resist pointing out that designers' research does not provide reliable and valid knowledge. It follows that design decisions made on this basis are also in doubt.

First, we need to distinguish between the verb research and the noun research. When designers research they predominantly collect relevant information. For scientists, "to research" describes the activity of conducting science, and the noun research is used as a synonym for science. Since there is no verb form of science, it appears necessary to continue to use the verb research for it. It follows that the activities of designers to collect information must be labeled with a different term and "to explore" appears a good choice. A design science project that does not use the classical scientific method can then be described as an exploration. Having clarified this important term, we may now proceed to discuss the quality criteria. The scientific reader may well be familiar with them and hence there is a danger of preaching to the converted. However, the comparison with related criteria in design may still be enlightening.

Quality Criteria For Science And Design

The generalizability of scientific knowledge is one of the most important criteria. It describes the degree to which general statements can be derived from a particular statement. The more general statements that can be derived, the better the particular statement. Newton's law of gravity was not only able to describe the behavior of the apple that inspired him, but also all other apples, fruits, organic materials, and inorganic materials. Even the motion of the stars could be described by it. His law is therefore of high value. If, on the other hand, a statement depends on the individual researcher then its generalizability is low. If I state "bugs are awful" then this may hold true only for people who share my paranoia about small creatures with many legs. Objectivity is therefore a good method for increasing the generalizability of a statement. Generalizability is also related to the repeatability of an experiment. If the

Bartneck, C. (2008). What Is Good? – A Comparison Between The Quality Criteria Used In Design And Science. Proceedings of the Conference on Human Factors in Computing Systems (CHI2008), Florence pp. 2485-2492. | DOI: 10.1145/1358628.1358705

results of an experiment are objective, meaning that they are not dependent on the experimenter, then others should be able to repeat the experiment with exactly the same results. Furthermore, time itself should not matter. Repeating the experiment at a later point in time should yield the same results. For the design community time does matter and hence the CHI conference has the section "Contemporary Trends". Like fashion, the results of design work are expected to change over time, which makes them less generalizable. However, some design classics, such as the Tizio lamp, appear to be timeless.

Designers know a similar concept: universality. It describes the degree to which general problems can be solved by a particular solution. The more universal a solution is, the better. A hammer, for example, is more universal than a pair of horseshoe pliers, and hence more valuable. However, there is usually a tradeoff between effectiveness and universality. Specific solutions usually work better than general solutions at the price of having to create a solution for each problem. The challenge is to find the right balance between universality and effectiveness. Science, on the other hand, strives towards the highest level of generalizability.

The knowledge that designers typically create in their design projects suffers from its lack of generalizability. The solutions found for a given problem are limited to the scope of the problem, and cannot be applied easily, if at all, to different problems. Also, the solutions are dependent on the individual designer. A different designer might have come to a different solution.

Falsifiability is another important criterion that is known to both scientists and designers. Originally proposed by Karl Popper (2002), falsifiability describes the property of statements that they must admit of logical and empirical counterexamples. The latter refers to the condition that it must be possible, at least in principle, to make an observation that would show the statement to be wrong, even if that observation is not actually made. The statement "all swans are white" is in principle falsifiable by observing a black swan. The higher the number of logical and empirical counterexamples that a statement withstands, the higher its value.

The use of falsifiability in design is very similar. A solution must admit of logical and empirical counterexamples. If, for example, a certain device is intended to continuously increase one's karma, then its function is impossible to falsify. Such a device could not be considered a design. Falsifiability plays a less important role in design in comparison with science, since it often deals with concrete and well-defined problems. The effects of a solution are usually easy to observe, and this criterion overlaps the criterion of effectiveness that will be discussed later.

Truth is a key criterion in science, and it also plays an important role in design. However, multiple definitions of truth exist. The Wikipedia lists many theories of truth including correspondence, coherence, constructivist, consensus, pragmatic, performative, semantic, and Kripke's theory. The correspondence and coherence theories are probably the most acknowledged, and hence this study focuses on them. In the coherence theory, truth is primarily a property of a whole system of statements. The truth of a single statement can be derived only from its accordance with all the other

Bartneck, C. (2008). What Is Good? – A Comparison Between The Quality Criteria Used In Design And Science. Proceedings of the Conference on Human Factors in Computing Systems (CHI2008), Florence pp. 2485-2492. | DOI: 10.1145/1358628.1358705

statements. If a new statement contradicts an existing statement, then both statements need to be reconsidered. In the previously used example of swans, one of the statements must be false. Either not all swans are white or the particular swan is not black. The equivalent concept in design is known as compatibility. If a new component is introduced to an existing system then it should not prevent any existing component from operating correctly. For example, the installation of new software on a computer can lead to incompatibilities in which previous functions cease to operate.

The correspondence theory of truth deals with the relationship between statements and reality. If theories correspond to observations in reality then they are considered to be true. This direction in the relationship between truth and reality is usually attributed to science. The other direction can be attributed to design. If an artifact corresponds to theory then it is considered true. Our understanding of the physical world makes it difficult to invent artifacts that could not be explained fully by existing theories of physics. Many attempts have been made to invent a perpetual motion machine, and patents have even been filed, but no working model has been built. The United States Patent and Trademark Office (USPTO) has made an official policy of refusing to grant patents for perpetual motion machines without a working model:

"With the exception of cases involving perpetual motion, a model is not ordinarily required by the Office to demonstrate the operability of a device. - 608.03 Models, Exhibits, Specimens [R-3]"

However, solutions have often been used without full theoretical understanding. The Bayer Company patented

aspirin as early as 1899, and has successfully marketed it ever since. Its pain relieving effect was understood only in 1971. In 1982, John Robert Vane received the Nobel Prize in the Physiology of Medicine for this discovery.

Another important quality criterion for scientific knowledge is novelty. Rediscovering Newton's laws has little value. But newness in itself is not sufficient. A novel scientific theory does not only need to be different from existing theories, but it also has to explain more than existing theories. Galileo's theories extended Aristotle's, Newton's law extended Galileo's, and Einstein's extended Newton's. In design, the same principle is known as innovation. Novelty, in its pure 'newness' definition, is even a requirement for patents. Moreover new artifacts are expected to work not only differently, but also better. Modern PCs are currently even powerful enough to completely simulate older computers, for example, simulating the Commodore 64 using the VICE emulator. Modern PCs can do everything that older ones can, and more.

The criterion of parsimony, also known as Occam's razor, is the preference for the least complex statement to explain a fact. A good example can be found in the field of Astronomy. The Copernican model is said to have been chosen over the Ptolemaic due to its greater simplicity. The Ptolemaic model, in order to explain the apparent retrograde motion of Mercury relative to Venus, posited the existence of epicycles within the orbit of Mercury. The Copernican model (as expanded by Kepler) was able to account for this motion by displacing the Earth from the center of the solar system and replacing it with the Sun as the orbital focus of planetary motions, while simultaneously replacing the circular orbits of the

Bartneck, C. (2008). What Is Good? – A Comparison Between The Quality Criteria Used In Design And Science. Proceedings of the Conference on Human Factors in Computing Systems (CHI2008), Florence pp. 2485-2492. | DOI: 10.1145/1358628.1358705

Ptolemaic model with elliptical ones. In addition, the Copernican model excluded any mention of the crystalline spheres that the planets were thought to be embedded in according to the Ptolemaic model. At a single stroke, the Copernican model reduced the complexity of Astronomy by a factor of two.

In design, simplicity plays a similar role. Simplicity is the preference for the least complex solution to achieve a given goal. Just 20 years ago, the only way to print a photo required a complete photochemical process that involved various toxic chemicals and sophisticated machines. These days, everybody can print his own pictures with cheap inkjet printers.

Lastly, the scientific criteria of accuracy, precision, and efficiency are discussed, together with their counterparts in design: effectiveness, reliability, and efficiency.

Accuracy refers to the degree to which a statement or theory predicts the facts it is intended to predict, while precision refers to the degree to which a statement or theory predicts the exact same facts. The analogy of bullets shot at a target is useful to explain the difference between these two related concepts and at the same time to show the similarity between design and science criteria.

In this analogy, a gun firing at a target (design) parallels a theory predicting observations (science). The effectiveness of the gun describes the closeness of the bullets to the center of the target (see Figure 1 left). Bullets that strike closer to the center are considered more effective. The parallel is that the closer the observations concur with the predictions of the theory, the more accurate the theory.



Figure 1: High effectiveness but low reliability (left), high reliability but low effectiveness (middle) and high reliability and high effectiveness (right).

To continue the analogy, the reliability of the gun refers to the spread of the bullets. The closer together the bullets strike, the higher the reliability (see Figure 1 middle). In science, the closer the observations are to each other, the more precise the theory. The bullets do not necessarily need to be close to the center for this. The bullets (or observations) can be reliable (precise) without being effective (accurate). However, for bullets (and observations) to be perfectly effective (accurate), they also need to be reliable (precise) (see Figure 1 right).

For science, efficiency refers to the resources expended in relation to the precision and accuracy of the observations predicted, and for design, efficiency refers to the resources expended in relation to the effectiveness and reliability of the goals achieved.

So far only those quality criteria of design that have a direct relation to the quality criteria of science have been discussed. Of course, design also has criteria that are of less relevance to science. Conformity to social customs, popularity, ego satisfaction, reputation, pleasure, and commercial success are examples. It is difficult to define general design criteria, since each design can be judged only in its specific context of use. The Hummer sport

Bartneck, C. (2008). What Is Good? – A Comparison Between The Quality Criteria Used In Design And Science. Proceedings of the Conference on Human Factors in Computing Systems (CHI2008), Florence pp. 2485-2492. | DOI: 10.1145/1358628.1358705

utility vehicle (SUV), for example, is a car that is not intended to be environmental friendly and hence it should not be judged by the fuel consumption criteria. The Hummer SUVs are not designed for driving fuel-efficiently from A to B.

Conclusion

Science has established several criteria for assessing the quality of the knowledge it produces. Some of these criteria overlap or relate to criteria that are used in design. Design methods are not yet optimized for the creation of scientific knowledge, and therefore they generally produce knowledge that is of lesser scientific quality. Often they are not even interested in it. Jon Kolko, editor of ACM's <interactions> magazine, rejected this very manuscript based on its academic format:

"However, the submission is in a very academic format, while Interactions Magazine is intended to read in a more approachable and casual manner - specifically, it is intended to be of worth to practitioners, who may not be familiar with or interested in the very specific and grounded citations and discourse you have provided."
[7]

If design wants to contribute to the growth of scientific knowledge, then it will primarily have to improve the generalizability of its results. Most of all, to guarantee objectivity, its results need to become independent of the designer. Pitt claimed [11] that such a method would lead to knowledge that is "far more reliable, secure, and trustworthy than scientific knowledge". Currently, designers who want to work as scientists have to become either engineers or psychologists. Since they often lack training in these disciplines they have a natural disadvantage.

Until considerable progress has been made in defining a suitable epistemology for design, we shall have to take small steps forward using current methods and policies. Design has to acknowledge that the knowledge it produces is, from a scientific perspective, not very generalizable, and hence of lesser value. Scientists, on the other hand, need to acknowledge that the highly general knowledge they produce is often too abstract to improve society. It requires a skilled designer to translate this knowledge to a specific context of use.

As for the CHI conference, it would be wise to follow Confucius' recommendation to "rectify the names". Labeling only one section "archival" when both sections will be stored in the ACM Digital Library is confusing at best. Storing a document in the ACM Digital Library means, by definition, that it has been archived. This contradiction becomes dramatically clear when reading the CHI2008 extended abstracts style guide:

"The publication is not considered an archival publication; however, it does go into the ACM Digital Library."

Also, the labels "main proceedings" and "extended abstracts" are ambiguous. Pirsig's static quality patterns [10] appear suitable for defining the sections, but the terms "intellectual" and "social" carry different meanings in the various sub-communities, and hence may cause misunderstandings. Maybe the sections could be called "Discovery" and "Invention". The latter would collect contributions that are aimed at improving society. The discovery section would gather contributions that present scientific insights. Whatever principle is used to divide the proceedings, it should be made explicit.

Bartneck, C. (2008). What Is Good? – A Comparison Between The Quality Criteria Used In Design And Science. Proceedings of the Conference on Human Factors in Computing Systems (CHI2008), Florence pp. 2485-2492. | DOI: 10.1145/1358628.1358705

The use of 'best paper' awards is another ranking method. Excellence should be rewarded. However, rankings should not be used to discriminate between communities. Excellence can be found in design papers as well as in scientific papers. The factors that influence paper rankings should be made explicit. This would require the agreement of the community on the factors used. The CHI community is diverse, and it may be difficult to reach agreement. But nothing worthwhile is ever easy. As long as no shared quality criteria are defined for the community as a whole, it will remain a trans-disciplinary rather than a multi-disciplinary community. The sub-communities of design, education, engineering, management, research, and usability will co-exist, but future shouting matches cannot be excluded.

References

- [1] Arnowitz, J., & Dykstra-Erickson, E. (2005). CHI and the Practitioner Dilemma. *Interactions*, 12(4), 5-9. | DOI: 10.1145/1070960.1070964
- [2] Bartneck, C., & Rauterberg, M. (2007). HCI Reality - An Unreal Tournament. *International Journal of Human Computer Studies*, 65(8), 737-743. | DOI: 10.1016/j.ijhcs.2007.03.003
- [3] Chalmers, A. F. (1999). What is this thing called science? (3rd ed.). Indianapolis: Hackett.
- [4] Cross, N. (1993). Science and design methodology: A review. *Research in Engineering Design*, 5(2), 63-69. | DOI: 10.1007/BF02032575
- [5] Cross, N., Science and design methodology: A review. *Research in Engineering Design*, 1993. 5(2): p. 63-69.
- [6] Dourish, P. (2006). Implications for design. *Proceedings of the SIGCHI conference on Human Factors in computing systems*, Montreal, Quebec, Canada. | DOI: 10.1145/1124772.1124855
- [7] Kolko, J. (2007). Email from Jon Kolko to Christoph Bartneck on August 21st, 2007.
- [8] Levy, R. (1985). Science, technology and design. *Design Studies*, 6(2), 66-72. | DOI: 10.1016/0142-694X(85)90016-X
- [9] Pirsig, R. M. (1974). Zen and the art of motorcycle maintenance: an inquiry into values. New York: Morrow.
- [10] Pirsig, R. M. (1991). Lila : an inquiry into morals. New York: Bantam Books.
- [11] Pitt, J., C. (2001). What Engineers Know. *Techne*, 5(3), 17-30.
- [12] Wolf, T. V., Rode, J., A., Sussman, J., & Kellogg, W., A. (2006). Dispelling "design" as the black art of CHI. *Proceedings of the SIGCHI conference on Human Factors in computing systems (CHI)*, Montreal, Quebec, Canada. | DOI: 10.1145/1124772.1124853