



# Usability Studies and the Hawthorne Effect

## Ritch Macefield

Department of Applied Computing

Staffordshire University

College Road

Stoke on Trent, ST4 2DE UK

[ritchie@bulldoghome.com](mailto:ritchie@bulldoghome.com)

[www.ritchmacefield.info](http://www.ritchmacefield.info)

## Abstract

This paper provides a brief review of the Hawthorne effect, a discussion of how this effect relates to usability studies, and help for practitioners in defending their studies against criticisms made on the basis of this effect.

## Keywords

Usability study

Hawthorne effect

Experimenter effects

## Introduction

As usability practitioners, we are well familiar with many issues relating to how well a study reflects real system usage, for example, how accurately the task scenario mirrors a real context of use and how well the study participants represent the target user group. However, it seems that there is one issue with which our discipline is less familiar – the Hawthorne effect.

The popular understanding of the Hawthorne effect is that it is an experimenter effect whereby participants, in *any* human-centered study, may exhibit atypically high levels of performance simply because they are aware that they are being studied.

Usability studies are examples of such human-centered studies and, if this view of the Hawthorne effect is the

whole story, there is good reason to have serious concerns about the validity of many findings.

The aim of this paper is three-fold. First, it reviews the Hawthorne effect and discusses whether this popular view is indeed the whole story. Second, it specifically relates this effect to usability studies. Finally, it establishes a number of ways in which usability practitioners can defend their studies against criticisms made on the basis of this effect.

### Review of the Hawthorne Effect

The Hawthorne effect was identified from a series of studies conducted in the 1920s at the Hawthorne Works, just outside Chicago. The company produced electrical equipment. These studies were a major effort to explore, in a systematic way, the impact of environmental factors on the productivity of the workforce. For the purpose of this review, we will focus on the studies that explored the impact of changing the lighting levels, as reported in Snow (1927).

In this set of studies, the workshop lighting levels were systematically varied for workers performing repetitive tasks. In one study, the impact of three different lighting levels on the productivity of the workers was evaluated. In another study, a control group was given stable light levels while an experimental group was given progressively increasing levels of light. To the surprise of the researchers, all of these conditions led to an increase in productivity. Even when lighting levels were decreased, productivity continued to increase. Similarly, productivity also increased in the control group.

It was clear from these studies that there *were* changes in the productivity of the test participants; however, these changes were *not* due to the lighting levels. The researchers realized that there was some other reason for the productivity to generally increase.

Mayo (1933), a key member of the research team at the Hawthorne Works, argued that the performance improvements came about because the test participants *believed* that the changes would improve their performance and were *flattered* and *motivated* by the attention they were being given during the study.

The day-to-day work of these participants was quite monotonous, and they normally had no contact with management or outside experts, so the conditions of the experiment were very unusual for them. Mayo argued that they had responded to this novelty and increase in attention by making extra effort no matter what the experimental conditions. Further, Mayo suggested that this effect was generalizable to many other contexts.

In the decades that followed the Hawthorne Studies, Mayo's account of the Hawthorne effect became very popular, and Draper points out that:

when we read about "the Hawthorne effect" in most textbooks and papers, what the author really means is Mayo's interpretation of the Hawthorne effect. (Draper, 2006).

As pointed out by Rice, in keeping with the suggestions of Mayo (1933), application of the effect also became more generalized. As a result:

Proponents of the Hawthorne effect say that people who are singled out for a study of any kind may improve their performance or behavior not because of any specific condition being

tested, but simply because of all the attention they receive. (Rice, 1982, p. 1).

Because Mayo's interpretation of the Hawthorne effect had become so prominent, many authors went back to the detailed accounts of the studies and produced a number of different interpretations of what was happening (e.g., Carey 1967; Gillespie, 1991; Olsen et al., 1994; Draper, 2006). Indeed, the Hawthorne effect became the subject of wide and extensive debate.

The first critique that gained widespread support came from Parsons (1974), who performed a detailed analysis of the Hawthorne studies. This included interviewing many of the experimenters and participants in the original studies.

Parsons concluded that the performance improvements were easily explainable by mechanisms other than those proposed by Mayo. Some of these mechanisms were quite obvious. For example, Parsons cited the fact that one performance improvement occurred when the participants were taking shorter rest breaks. When the participants returned to taking their normal (longer) rest breaks the performance went back to the previous level.

Overall, Parsons concluded that the Hawthorne effect was better explained by *learning* and *feedback* mechanisms. In their normal work, the operators got no feedback on their productivity from day to day. However, in the study situation the experimenters were collecting data and telling them how well they were doing. It is easy to envisage that the operators got involved in setting themselves higher and higher targets, and got considerable satisfaction from attaining these targets. On this basis, Parsons redefined the Hawthorne effect as:

the confounding that occurs if experimenters fail to realize how the consequences of subjects' performance affect what subjects do" i.e., learning effects, both permanent skill improvement and feedback-enabled adjustments to suit current goals. (Parsons, 1974, abstract).

Further, Parsons argued that these learning and feedback mechanisms were well understood before the Hawthorne studies, and that there was no need for the term "Hawthorne effect" at all.

Parsons made significant attempts to correct what he saw as Mayo's mistaken interpretation, but ultimately recognized that this interpretation was so entrenched that it had become part of the accepted wisdom within the research community. When asked by Rice (1982, p. 3) why authors of current textbooks continued to include unquestioningly Mayo's interpretation, Parsons simply replied with "They're lazy.". This stance was later supported by Adair (1984) who warned that most secondary publications on the Hawthorne effect contain gross factual errors.

Parsons' interpretation of the Hawthorne effect was subsequently challenged by Sonnenfeld (1983), although this challenge was only partial. Sonnenfeld agreed with Parsons that the experimenters' *feedback* was a critical factor in the effect; however, Sonnenfeld questioned whether the workers at Hawthorne were really *learning* anything new about their tasks during the study.

Rice (1982) also argued against Mayo's interpretation of the Hawthorne effect, concluding that it had become one of the 'scientific myths' that get perpetuated because authors recycle the generalizations of others rather than going back to original sources. He stated that:

Like a number of other once widely held but faulty theories in psychology, such as the belief in a racial basis for intelligence, the Hawthorne effect has a life of its own that seems to defy attempts to correct the record. (Rice, 1982, p. 1).

Rice (1982) also reported that, over the last 50 years, there have been many attempts to validate Mayo's interpretation of the Hawthorne studies, including many studies carried out at the Hawthorne Works itself. Rice argued that none of these studies had found in favor of the Mayo interpretation and many found against Mayo. Rather, Rice provided close support for the interpretation of the effect proposed by Parsons (1974).

Other authors have claimed that there is some empirical evidence to support Mayo's interpretation. These include Draper (2006), who has made an extensive analysis of the Hawthorne effect. However, Draper cites little of this evidence. He also points out that the evidence is sporadic and, on the whole, inconclusive. Further, Draper uses the sporadic nature of this evidence to conclude that the Hawthorne effect is not generalizable, as Mayo had suggested.

Despite the criticisms of Mayo's interpretation, it remains widespread today. Indeed, its refusal to die still receives the attention of contemporary authors. For example, in an interview with Kolata (1998), Ross calls it one of the "Scientific Myths That Are Too Good to Die".

In summary, the interpretation of the Hawthorne effect originally presented by Mayo (1933) remains widespread. However, those who have researched this effect are aware that there are many interpretations of the Hawthorne effect, of which Mayo's is just one. Further, the balance of informed opinion seems to have

moved strongly away from Mayo's interpretation and more toward that proposed by Parsons (1974). Similarly, contrary to Mayo's suggestion, it would not seem safe to view the Hawthorne effect as generalizable to all human-centered studies.

### **Reliability of the Hawthorne Effect to Usability Studies**

Given that the Hawthorne effect should not be viewed as a generalizable idea, the question now is – how (well) do the studies at Hawthorne Works relate to usability studies?

The first point to consider is that there are some very significant differences between the context of the Hawthorne studies and the context of a typical usability study.

#### *Longitudinal nature of the Hawthorne studies*

The Hawthorne studies were all *longitudinal* in nature; that is, the idea was to improve the performance of the *same task over a period of time*. Sometimes, this is also the case with usability studies; however, usability studies more often involve a *once-off testing*.

#### *Expertise of the test participants in the Hawthorne studies*

The participants in the Hawthorne studies were all *experts* in the task being measured. Typically, they had repeated the task many thousands of times, over a period of years. Again, usability studies are sometimes concerned with experts in the task. However, many usability studies involve *novice* users. Further, even when expert users are involved, the combination of task, system and user is nearly always *novel*.

#### *Metrics used in the Hawthorne studies*

The Hawthorne studies were primarily concerned with *efficiency*, that is, pieces completed/hour. Indeed, efficiency improvement was the whole driver for the studies. This is somewhat different from usability studies, which generally adopt the definition of usability defined in ISO 9241-11 and the CIF. In other words, it is generally held that *effectiveness*, *efficiency*, and *satisfaction* are held in equal (or at least similar) regard.

#### *Novelty of the study for participants in the Hawthorne studies*

The workers at Hawthorne Works had *monotonous* jobs and, as such, the studies may have been a welcome novelty and subject of interest for the study participants. In turn, this may have affected their motivation and subsequent performance. We cannot assume this to be the case with participants in a usability study. Taking part in a usability study may constitute an unwanted interruption for some participants. Indeed, they may only be taking part in the study because they are being paid.

#### *Participants' relationship with management in the Hawthorne studies*

At Hawthorne, the studies represented a point of contact between the workers and (senior) management. As such, it can be conjectured that the workers may have had a vested interest in a successful outcome for a study, perhaps thinking that this would reflect well on them as individuals in the eyes of their management.

Alternatively, there is no such point of contact in a typical usability study. The participants in such studies

are not typically employed by the system vendor or organization conducting the study. Indeed, we often take steps to avoid such conditions. As such, participants in a usability study typically have no vested interest in a particular outcome for the study in terms of (improving) how they are perceived by their management.

These differences suggest that we need to be careful when relating the experiences at Hawthorne Works to the context of a usability study. However, if there are doubts that the performance of participants is being influenced by the Hawthorne effect then there are some other points to consider.

#### *Application of Parsons' interpretation of the Hawthorne effect to Usability Studies*

If we accept the interpretation of the Hawthorne effect proposed by Parsons (1974), then an interesting paradox becomes apparent when we apply this thinking in the context of a usability study.

Parsons argued that the effect comes about mainly due to the participants gaining feedback during the study, then learning from this feedback to improve their performance. The provision of feedback is generally a key design feature in most modern interface designs (e.g., Norman, 1988). Similarly, it is generally accepted that participants in a usability study will *always* learn something about the system (e.g., Norman, 1983; Raskin, 1994). On this basis, it can be argued that the Hawthorne effect is an *inevitable* part of all usability studies and is a necessary mechanism for bringing about any performance improvements.

This view of how Parsons' interpretation of the Hawthorne effect relates to usability studies is correct in the widest sense. However, the feedback that

Parsons drew attention to in the Hawthorne Studies is qualitatively different from that referred to above in relation to the usability studies. In usability studies, the feedback is *intrinsic* to the use of the system and, therefore, would also take place outside of a study context. Alternatively, the feedback in the Hawthorne studies was *extrinsic* because it was provided by the experimenters, so it only occurred as a result of the study context.

The message for usability practitioners here is that they should avoid making performance data available to participants or giving feedback during the study. This relates to the well-known problem of what help to give study participants when they are having difficulties with a task.

*Proof and explanation of the causation mechanisms in usability studies*

Those who adopt Mayo's interpretation of the Hawthorne effect stress the need to explain the causation mechanism that brought about the effects measured in the study. If we can explain *exactly* how the performance of the users is a result of system attributes, then it is possible to argue that the effect was not due to the Hawthorne effect.

For example, suppose some users had been reporting usability problems with an interface. The interface is re-designed and a usability study performed to test if the re-design has solved the problems. To defend against Mayo's interpretation of the Hawthorne effect, it would then be necessary to explain exactly how the new design solved the usability problems. Further, it would not be enough to simply cite the changes made to the old interface, for example, fonts were made larger. It

would be necessary to explain how the test participants had benefited from the changes.

The problem here is that, with usability studies, understanding causation mechanisms logically involves understanding what mental model users have of the interface and how this affected their actions. Unfortunately, there are many problems with eliciting and understanding users' mental models (Johnson-Laird, 1983; Norman, 1983; Sasse, 1997). Often, it is simply not possible to prove or explain the causation mechanisms that led to a particular set of findings.

Despite this, there are some *qualitative* techniques available to practitioners that can provide *indicators* of causation mechanisms. The most common of these are *verbal protocols* (e.g., think-aloud protocols) used during the testing, and pre- and post-test *semi-structured interviewing* of the participants.

*Use of controls*

Those who adopt Mayo's interpretation of the Hawthorne effect also advocate the use of *controls*. Regardless of what happened in the Hawthorne studies, it remains possible that the expectations of experimenters and study participants might influence the results that are obtained, and controls are often used in experimental science to avoid these contaminations.

For example, in the scenario presented earlier, whereby an interface has been re-designed to solve some identified usability problems, a usability study could be designed with two participant groups. The 'test group' could use the new design and the 'control group' could use the old design. Any Hawthorne effect should act across both groups (as was found in the original Hawthorne studies), and therefore, it can be argued

that any improvement in the test group's performance would not be due to the Hawthorne effect.

The argument is strengthened if it is ensured that the two groups are matched in user characteristics, neither have previous knowledge of the interface and that the participants are 'blind'; that is, they do not know which interface is expected to give the better performance.

Independent of the Hawthorne effect, the use of (blind) controls is clearly an example of best practice in usability studies, and therefore, is always advocated. However, the use of controls is not applicable in many usability studies, for example, when a completely new system is being tested. We must also recognize that commercial realities may prohibit the use of such best practices, but that this does not inevitably render a study invalid or not useful.

### **Defenses Against Criticisms Based on the Hawthorne Effect**

Based on the previous discussions, it seems likely that most usability practitioners will ultimately find themselves in the position of having to defend a study against criticisms that the results are contaminated by the Hawthorne effect.

The most fundamental issue in mounting any defense is that the practitioner is aware of the whole story regarding the Hawthorne effect.

Reference to the Hawthorne effect in the literature is mixed. Some prominent texts, for example, Preece, et al. (1994, p. 617) cite the effect. However, other prominent literature does not, for example, the Common Industry Format for usability testing (CIF, v2.02). When the effect is cited, it is typically the

popular view that originated with Mayo (1933) that is presented, for example, Kahn Research Group (2000).

Therefore, we probably have more to do in terms of raising awareness of the Hawthorne effect across our discipline.

Another issue is that, when the effect is mentioned in usability literature, the usual advice given is not very specific; it generally takes the form of 'watch out for this effect'. In other words, there is very little (specific) advice available to practitioners as to how they can mount any defense.

From the analysis presented here, we can identify a number of possible defenses.

#### *The Rice-Ross defense*

The most aggressive defense is to present the views of Rice (1982) and Ross, in their interview with Kolata (1998) – that the Hawthorne effect, as it is usually understood, is nothing more than a popular myth that should not be used as the basis on which to question the validity of *any* experimental study.

#### *The controversy defense*

A similar, but less aggressive, defense is to argue that the Hawthorne effect is a highly controversial topic and that the original phenomenon has been subject to many different interpretations. Therefore, it is unsafe to criticize an experimental study on the basis of this effect.

#### *The relatability defense*

A more specific defense is to point out that there are many ways in which usability studies are significantly different from the original Hawthorne studies.

Therefore, it is unsafe to relate the Hawthorne effect to a usability study.

#### *The Parsons defense*

It is also possible to mount a defense based on the interpretation of the Hawthorne effect proposed by Parsons (1974); that is, that performance increases come about because of the learning and feedback that takes place during the experiment. Therefore, in the widest sense, any performance benefits identified in a study actually *rely* on the Hawthorne effect.

However, with this defense, the study team should also ensure that extrinsic feedback from the experimental condition is eliminated or, at least minimized as far as possible.

#### *Defense against the Mayo interpretation*

From the earlier discussions, it seems likely that most criticisms of usability studies that are based on the Hawthorne effect will be grounded in the popular understanding of this effect, which originated with Mayo (1933). It is also likely that the challenger will not be aware of the whole story about this effect. As such, the challenger may well retreat when fully informed of the whole story here, thereby opening the way for one of the defenses presented above. However, in cases where Mayo's interpretation of the Hawthorne effect must be addressed (perhaps as a result of client pressure), then the study team's best defense lies in the design of the study.

First, the study design should maximize the use of *qualitative techniques* designed to elicit the participants' thinking during the study. Such techniques will typically include verbal protocols, and pre- and post-test semi-structured interviewing of the

participants. The aim of using these techniques would be to gather evidence as to the causation mechanisms which brought about any significant findings (benefits) identified from the study, for example, how the use of large fonts made an interface simple(r) to use. This evidence can then be used to argue that the findings were *not* due to the Hawthorne effect, but to some other causation mechanism.

Second, the study should use (blind) *controls* wherever this is possible. In such a design, any Hawthorne effect would act across both the 'test group' and the control group. Therefore, it would be safe to argue that any differences between the test group and control group were not due to the Hawthorne effect.

In these cases, the argument can be strengthened by pointing out any *similarities* between the results from the two groups. For example, suppose a study showed that the test group exhibited significantly higher efficiency when using the "search" facility on an e-Commerce web site, but that the efficiency was similar across both groups when they used the "checkout" facility. This is strong evidence against the presence of any Hawthorne effect. If the Hawthorne effect was contaminating the results by acting to improve the performance of the test group (over and above that of the control group), we would expect the test group to be more efficient in *all* elements of the study. In other words, we would expect any Hawthorne effect to be *global* in nature and the absence of this condition is strong evidence against any contamination by the effect.

### Practitioner's Take Away

- The Hawthorne effect can be (mis)used as a basis on which to criticize the validity of human-centered studies, including usability studies. Therefore, it is important that practitioners are able to defend themselves against such criticism. A wide variety of defenses are possible; depending on which interpretation of the Hawthorne effect is adopted. To make an informed decision as to which interpretation to adopt, practitioners should be aware of the *whole story* regarding this effect.
- A precursor to any defense should be pointing out that there are many significant differences between the studies carried out at Hawthorne Works and typical usability studies. Therefore, care must be taken when relating any interpretation of this effect to our discipline.
- Most criticisms will be founded on the interpretation of the effect proposed by Mayo (1933). However, despite its popularity, this interpretation has been largely debunked over the last few decades. At worst, it can be considered as nothing more than a *popular myth* that has no place in any serious research thinking. At best, it can be considered as a controversial idea that has highly questionable reliability to our discipline. Therefore, a defense against this interpretation is not likely to be required once all the stakeholders in a study understand this fact. However, should such a defense be required, the study should maximize the use of (blind) *controls* and gather evidence of the *causation mechanisms* that resulted in any significant findings (benefits). This is probably best achieved through the use of *qualitative techniques* such as *verbal protocols* and pre- and post-test *semi-structured interviewing*.
- If the interpretation of the Hawthorne effect proposed by Parsons (1974) is adopted, then ensure that *extrinsic performance feedback* to participants in a study is eliminated, or minimized as far as is reasonably possible.
- Since some defenses affect the study design and its execution, it is clearly important that the study team agree with the client what position will be taken in relation to the Hawthorne effect *in advance* of the study. This position should be then be published in an appendix to the study report.

### References

- Adair, G. (1984). The Hawthorne effect: A reconsideration of the methodological artifact. *Journal of Applied Psychology*, 69(2), 334-5.
- Carey, A. (1967). The Hawthorne Studies: A radical criticism. *American Sociological Review*, 32, 403-16.
- CIF v2.02 (2002). *Common Industry Format for usability test reports*. Retrieved June 22, 2002, from <http://zing.ncsl.nist.gov/iusr/overview.html>.
- Draper, S. (2006). *The Hawthorne, Pygmalion, placebo and other effects of expectation: some notes*. Retrieved January 1, 2007, from <http://www.psy.gla.ac.uk/~steve/hawth.html>.
- Gillespie, R. (1991). *Manufacturing knowledge: a history of the Hawthorne experiments*. Cambridge: Cambridge University Press.
- International Standards Organization (1998). *Definition: 9241-11:1998*. Retrieved April 27, 2002, from <http://www.iso.ch>.
- Johnson-Laird, P. N. (1983). *Mental models*. Cambridge: Cambridge University Press.

- Khan Research Group (2000). *Validity 101 – What it is, why it is essential, where others fail, and how we succeed*. Retrieved January 18, 2007, from <http://www.webehavior.com/validity101.htm>.
- Kolata, G. (1998). Scientific Myths That Are Too Good to Die. *New York Times*. Retrieved January 18, 2007, from <http://www.nytimes.com/library/review/120698science-myths-review.html>.
- Mayo, E. (1933). *The human problems of an industrial civilization*. New York: MacMillan.
- Norman, D. A. (1983). Some observations on mental models. In: Gentner, D. & Stevens, A., L. *Mental Models*. London: Erlbaum.
- Norman, D. A. (1988). *The psychology of everyday things*. New York: Basic Books.
- Parsons, H. M. (1974). What happened at Hawthorne? *Science*, 183(4128), 922-32.
- Preece, J., Rogers, Y., Sharp, H., Benyon, D., Holland, S. & Carey, T. (1994). *Human-computer interaction*. Wokingham: Addison Wesley.
- Raskin, J. (1994). Intuitive equals familiar. *Communications of the ACM*, 37(9), 17.
- Rice, B. (1982). The Hawthorne defect: Persistence of a flawed theory. *Psychology Today*, 16(2), 70-4.
- Sasse, M. A. (1997). Eliciting and Describing Users' Models of Computer Systems. PhD Thesis. Birmingham University, UK.
- Snow, C. E. (1927). Research on industrial illumination: A discussion of the relation of illumination intensity to productive efficiency. *The Tech Engineering News*, 257-82.
- Sonnenfeld, J. (1983) Commentary: Academic Learning, Worker Learning, and the Hawthorne Studies, *Social Forces*, 61(3), 904-9.

## Acknowledgements

I thank Professor Ken Eason for acting as a sounding board, and for his assistance with editing this article.

## Author Bio



**Ritch Macefield** is a Senior Lecturer in Information Technology at Staffordshire University and an HCI Consultant. Prior to this, he was Head of Usability Engineering and User Experience Thought Leader at Differentis Ltd, and a research fellow at Loughborough

University's Human Sciences and Advanced Technology research centre (HUSAT). Ritch has recently completed doctoral research at Loughborough University into how the provision of online conceptual models of (web-based) IT systems can aid their usability.