

CAFE OF EVE: a method for designing and evaluating interfaces

Neville Stanton
Department of Psychology
University of Southampton
Highfield
Southampton
SO17 1BJ

Tony Gale
Department of Psychology
University of Portsmouth
King Charles Street
Portsmouth
PO1 2ER

INTRODUCTION

Few would doubt the need for good interface design. When we use software over an extended period we may come to appreciate some aspects of the interface whilst abhorring others. We can also appreciate that some software packages are better than others although they perform essentially the same function. Intuitively, we are aware that good interface design enhances the usability of software and makes its functions more accessible to the user. The often quoted knock-on benefits include; greater productivity, fewer errors, and greater user satisfaction. What is needed however, is a mechanism for ensuring that newly designed software encapsulates the positive aspects of interface design whilst minimising the negative. This undertaking requires us to understand what is meant by usability.

The idea of usability is becoming common parlance in interface design. This is a welcome shift in emphasis towards 'ease-of-use'. In many respects, the fundamental tenet of usability is that software should be easy to use. This heightening of interest does not mean that usability (AKA *user friendly*, AKA *ergonomically designed*, AKA *user-centred design*, AKA *consumer-oriented product development*) is a new concept. Ergonomists have been beating this particular drum for the past fifty years or so. While one can point to the consequences of not considering usability, there is much debate as to what the term actually means. Stanton & Baber (1995) argue that one of the main problems with the term 'usability' is that it means different things to different people. Some may suggest that usability is simply another attempt to introduce 'user friendliness' back into product design jargon: usability is simply new wine in old bottles. Others argue that the issues surrounding usability have already been dealt with in 'user-centred design'. Baber (1993) points out, using the analogy of the soupstone, that the term usability takes on individual meaning to each person involved in evaluation to describe whatever they are doing: the individual adds their own ingredients. The trouble with this approach is how can we determine if one product is better than another, or indeed if the product has achieved some acceptable benchmark? Clearly this matter needs to be resolved as a matter of urgency, particularly in the light of recent legislation which makes usability a legal requirement in some products! This is a rather ridiculous situation given the debate and controversy surrounding the concept of usability (Stanton & Baber, 1992).

WHAT IS USABILITY?

While it is possible to indicate the necessity for usability in software development, as a concept it has proved remarkably resilient to definition; we all know what it is, but have difficulty reaching an agreed, coherent definition (it is our own personal soupstone) which will allow recommendations to be made concerning how best to make something more 'usable'. This is the first, and perhaps most important, stumbling block in determining methods appropriate to evaluation. If we cannot agree on what usability is, how can we hope to measure it? It is likely that different definitions of the concept will lead people to measure different aspects of product use. This suggests that a usability evaluation may not have a common standard between individuals. If usability is to be more than an ephemeral concept, we must agree on its

constituent ingredients. Stanton & Baber (1992) draw upon a decade of work represented by Shackel (1981), Eason (1984) and Booth (1989) to suggest the factors above serve to shape the concept of usability and define its scope. These are as follows.

1. **Learnability:** a system should allow users to reach acceptable performance levels within a specified time;
2. **Effectiveness:** acceptable performance should be achieved by a defined proportion of the user population, over a specified range of tasks and in a specified range of environments;
3. **Attitude:** acceptable performance should be achieved within acceptable human costs, in terms of fatigue, stress, frustration, discomfort and satisfaction;
4. **Flexibility:** the product should be able to deal with a range of tasks beyond those first specified.
5. **The perceived usefulness or utility of the product.** Eason (1984) has argued that ".....the major indicator of usability is whether a ... [product] is used" As Booth (1989) points out, it may be possible to design a product which rates high on the LEAF precepts, but which is simply not used.
6. **Task match:** In addition to the LEAF precepts set out above, a 'usable' product should exhibit an acceptable match between the functions provided by the system and the needs and requirements of the user.
7. **Task characteristics:** The frequency with which a task can be performed and the degree to which the task can be modified, e.g. in terms of variability of information requirements.
8. **User characteristics:** Another section which should be included in a definition of usability concerns the knowledge, skills and motivation of the user population.

Whilst we may argue over the relative merits of different ingredients and the labels we give them, this rarely becomes more than an exercise in semantics. ISO 9241 goes some way toward incorporating the above factors, but we feel that it falls short of a comprehensive definition in an important way. From reading ISO 9241 (at the time of writing this was still unreleased), we feel that usability has been defined by what can be measured: usability is what usability evaluations do. This appears to largely concentrate on the LEAF precepts mentioned above (Learnability, Effectiveness, Attitude and Flexibility). Stanton & Baber (1995) believe that haste in producing *the definition* of usability should be chastened by rather more circumspect consideration about what is meant by usability.

The reader will not be surprised to learn that each of the various factors which make up usability has spawned particular approaches to usability evaluation. In this section we present approaches related to aspects of product development. We are particularly concerned that reliance upon one approach exclusively or a very narrow definition of usability, which could lead an individual to perform a limited usability evaluation.

In an attempt to design better interfaces, some organisations are turning to Graphical User Interfaces (GUIs) in preference to the traditional Character-based User Interfaces (CUIs) with the inherent notion that the former is more *usable* than the latter. Unfortunately, it is not that simple! In a study comparing performance of people using GUIs and CUIs, Baber, Hoyes and Stanton (1993) found no difference in task completion times on two equivalent interfaces. The main differences being in learning times and the number of hypothesised mental operations for the two interfaces (both in favour of the GUI). However, for highly skilled performers using routine actions, a CUI may be preferable. Even users of GUI may revert to CUI-like interaction for very routine actions, such as CUT and PASTE in word processing applications because it is more efficient than selecting items from menus. Despite these subtle differences between the two types of interface, adopting a GUI does not mean that it will be a good GUI (Baber et al 1993). This leaves the question of: what is a good interface and how may it be achieved?

ISO 9241 (Draft at September 1994)

Notwithstanding the critique of ISO 9241 above (i.e. the measures of usability), the emphasis of the document is upon design and evaluation process and, in particular, upon the context within which the product is to be used. Contextual factors include; the user, the task, the equipment and the environment. Within ISO 9241 a quality plan is proposed for designing products that may be incorporated into an ISO 9001 quality systems containing four stages; identify context of use, select measures of criteria and context, evaluate usability and redesign the product. Whilst we have no argument with these processes in principle, we are sceptical of current efforts of implementing them. As an alternative, we propose that the CAFE OF EVE methodology has come of age and would enable organisations to both implement ISO 9241 and design good interfaces for their software products.

CAFE OF EVE

In 1984, Gale presented an internal report to the Human Factors Technology Centre at ITT Europe, in which he proposed a new research strategy for assessing the impact of new technology and for guiding design. In 1987, Gale and Christie set out a detailed blueprint for the approach. The project was called the CAFE OF EVE - a Controlled Adaptive and Flexible Experimental Office of the Future in an Ecologically Valid Environment.

In reviewing the literature on the impact of new office technology it was apparent that two research approaches were dominant: the use of surveys and questionnaires and the simulation of office environments in specialist laboratories. Neither was seen to offer predictive validity. Many of the early questionnaire studies involved selected samples of employees, often at a time of introducing new technology, under conditions of task restructuring, ill-trained for the technology in question, fearful of redundancy and motivated to a particular set of response biases. Some of the early literature on the stress imposed by continuous VDU exposure are contaminated by such extraneous factors (Salvendy and Smith, 1981). Laboratory studies, which offered more control, often involved relatively brief samples of behaviour generated by unrepresentative samples of volunteer participants with little sense of continuity of employment or of the personal significance of the tasks they were required to perform.

The aim of the CAFE OF EVE project was to combine the advantages of both laboratory and field studies, while minimising the disadvantages. The proposal involved taking over a major operating unit within a company in way which allowed for day-to-day operational function, combined with a parallel set of research studies. Staff operating within the selected office would be recruited with a job description which included the research function. The researchers would share some of the office functions with the aim of understanding the meaning of events and activities for participants. At the same time, through daily exchanges with the permanent employees, the barriers between participant and experimenter would break down. Thus the researchers would take on the role of participant observer as developed in anthropology (Vetere and Gale, 1987) living and working within the human system in question but also recording daily events. Researchers and participants share a social world. As the boundaries between researcher and participant become more permeable, participants feel more free to express their opinions and reactions about their working environment. In daily debriefing sessions, participants interact with researchers, with the goal of identifying problems from the participants' perspectives. Thus, the research questions which are generated are not dictated by existing theories but by the actual perceived experience of operators.

Attached to the working environment is an experimental environment for simulation, where those studies which cannot be carried out in vivo are set up in vitro. So far as possible, however, video observation and analysis, diary keeping, interactive recording of subjective responses, or even physiological monitoring would be carried out in the real office and integrated with everyday task functions. The aim of the CAFE OF EVE is to use a longitudinal and developmental technique to capture real experience and to shape new technological

developments. The research questions are not imposed by prior conceptions but emerge from the working context and the views and analyses of participants. It involves a partnership in exploration in which researcher and participant have equal status. Objectivity is retained because the researcher is still apart, but ecological validity is ensured by drawing on the participants' day to day experiences.

CONCLUSIONS

Thus, the CAFE OF EVE project seeks to draw together a normal working context and a controlled laboratory to create a special human factors environment, capitalising on the benefits of ecological validity and experimental control, while seeking to avoid the disadvantages of the two contrasting approaches. In so doing, the research benefits should surpass the benefits typically yielded by either approach taken separately or sequentially. What we are proposing and its emergent properties could constitute a minor revolution in human factors research.

REFERENCES

- Baber, C. (1993) Usability is useless. *Interfaces*. 21, 24-25
- Baber, C., Hoyes, T.W. & Stanton, N.A. (1993) Comparison of GUT's and CUI's: Appropriate ranges of action and ease of use. *Displays* 14 (4) pp. 207-215.
- Booth, P. (1989) *An Introduction to Human Computer Interaction* LEA: London.
- Eason, K.D. (1984) Towards the experimental study of usability *Behaviour and Information Technology* 3(2) 133-143.
- Gale, A. (1984). *Prelude to the Cafe of Eve*. Report to the Human Factors Technology Centre, ITT Europe (ESC Research Centre). Harlow: ITT Industries Ltd.
- Gale, A. and Christie, B. (1987). *Psychophysiology and the electronic workplace: the future*. In A. Gale and B. Christie (Eds). *Psychophysiology and the Electronic Workplace*. Chichester: John Wiley and Sons.
- Shackel, B. (1981) The concept of usability *Proceedings of the IBM Software and Information Usability Symposium* 1-30 IBM: New York.
- Stanton, N. A. & Baber, C. (1992) Usability and EC Directive 90/270. *Displays*, 13 (3) 151-160.
- Stanton, N. A. & Baber, C. (1995) Factors affecting the selection of methods prior to conducting a usability evaluation. In: Jordan, P. (Ed) *Usability Evaluation in Industry*. Taylor & Francis: London.
- Salvendy, G. and Smith, M.J. (Eds). *Machine Pacing and Occupational Stress*. London: Taylor and Francis.
- Vetere, A. and Gale, A. (1987). *Ecological Studies of Family Life*. Chichester: John Wiley and Sons.