A formal framework for interpreting the links between usability and utility

André Tricot

Work and Cognition Laboratory, CNRS, and

Hypermedia and Learning, Toulouse University for Teachers' Education

56 avenue de l'URSS, 31 078 Toulouse cedex - France

andre.tricot@toulouse.iufm.fr

phone : + 33 5 62 25 22 23

fax : + 33 5 62 25 20 18

A formal framework for interpreting the links between usability and utility

Abstract

When an interface, an information system, a piece of software, and more generally a goal-directed object is evaluated, its implementation (its utilisation) and its results (the goals that can be attained with it and their conformity to the intended goals) have to be evaluated. Implementation and results are described by variables: the variables of usability and utility respectively. Different models of relations between usability and utility can be found in human computer interaction literature, but they don't allow the reader to interpret the relations between empirical results in usability and utility tests. In this paper a formal framework for the interpretation of fifteen types of relations between these two variables is proposed. This framework permits an evaluation of the impact of usability improvement and utility improvement on the interface's global quality.

Keywords: evaluation, usability, utility, interface, ergonomics

1. Introduction

When an interface, an information system, a piece of software, and more generally a goal-directed object is evaluated, its implementation (its utilisation) and its results (the goals that can be attained with it and their conformity to the intended goals) have to be evaluated. Implementation and results are described by variables: the variables of usability and utility respectively. Grudin (1992) has discussed these two concepts, described the research fields and practices that they cover and emphasized the distance between them and how difficult this distance is to span. He expresses his desire for a convergence between the research fields and hopes that the two dimensions will be taken into account in design practices. In this paper it is argued that as far as evaluation is concerned, progress must first be made in the interpretation of the relations between these two variables. For example, a question such as “If the usability of this interface were improved, would it have an impact on the interface’s utility and on the interface’s global quality?” is very difficult to answer if the utility and the relation between the utility and the usability of this interface are unknown. Consequently a formal framework for interpreting the links between these two variables is proposed in this paper.

The concept of utility refers to the satisfaction of user’s needs. It is in many ways analogous to the concept of relevance, a concept which is itself central to information science. Within this discipline, the two concepts have practically the

same value and can even be taken as synonymous : useful information is relevant information and vice-versa. The value of considering them as being synonymous within the field of information science lies in the possibility of using the vast body of work that has been devoted to the concept of relevance. Mizzaro (1998) has conducted a review of the literature involving nearly 160 articles about relevance. For this author, relevance lies in the relation between two groups: one group contains a document, descriptors of this document and information (which is received by the user); the other group contains the user's problem, the user's need for information (his or her representation of the problem in informational terms), and the user's "natural" or "formalized" query. Each of these entities can be broken down into three registers: the field of content, the way the user intends to employ the information and the environment (e.g. working, learning, research). Relevance lies in the compatibility between each of the entities in the two groups (the document on the one hand and the user on the other) for each of the three components (field, employment, environment). It is possible to adopt this conception of relevance and extend it to the concept of the utility of interfaces. In this paper the following definition is used: the utility of an interface is the compatibility between the purpose of the interface and the user's goals for a given domain, employment and environment; the user's goal is to meet its need.

The evaluation of the usability of interfaces has been a factor of some importance for several years. What Eason (1984) and Nielsen (1993) began has been explored in thousands of further publications. This field of research and ergonomic engineering is primarily concerned with information systems that

exploit new electronic media such as Web sites (Buckingham Shum and McKnight, 1997). The five criteria of usability proposed by Nielsen (1993) include: learnability (i.e., How easy is it for users to accomplish basic tasks the first time they encounter the design?); efficiency (i.e., Once users have learned the design, how quickly can they perform tasks?); memorability (i.e., When users return to the design after a period of not using it, how easily can they re-establish proficiency?); errors (i.e., How many errors do users make, how severe are these errors, and how easily can they recover from the errors?); user satisfaction (i.e., How pleasant is it to use the design?). In this paper the following definition is used: the usability of interface can be seen as the possibility of using this interface; the user's goal is to use the interface.

Different models of relations between usability and utility can be found in human computer interaction literature. There are examined below.

1.1 Usability and utility are equally important

One of the most frequent positions is represented by Nielsen (2003) who wrote: "Usability is a quality attribute that assesses how easy user interfaces are to use (…) There are many other important quality attributes. A key one is utility, which refers to the design's functionality: Does it do what users need? Usability and utility are equally important: It matters little that something is easy if it's not what you want. It's also no good if the system can hypothetically do what you want, but you can't make it happen because the user interface is too difficult". In a panel of the IEEE Visualization Conference (Grinstein et al. 2003), Stasko wrote: "If the question is, "Which comes first, utility or usability?" my answer is "yes".

Both notions are vitally important (…) and they are just two sides of the same coin". Many empirical studies in human computer interaction adopt this position (e.g., Höök, 1997). The ergonomic criteria proposed by Scapin and Bastien (1997) are example of the rigorous definition of criteria for the ergonomic evaluation of software which integrates usability criteria and one utility criterion (i.e., the adaptability criterion).

1.2 Usability is more important, or Utility is a dimension of usability

Most of the empirical studies in human computer interaction evaluate usability but not utility (e.g., Blandford and Wong, 2004), or usability and to a lesser extent utility (e.g., Diamadis and Polyzos, 2004). The success of the concept of usability has been such that it appears that usability is gradually becoming synonymous with the general quality of an interface. For example, Park (2000) conducted a comparative study of two types of information system in which she measures usability, users' preferences, the effectiveness of the systems and the behaviour of the users. The author used three criteria to evaluate usability: users' opinions concerning the simplicity of utilization, their level of satisfaction with the results of their information searches and their judgement of the time taken to perform their search (adequate or not). She distinguished these criteria of usability from user satisfaction and system effectiveness which was measured using the classic criterion of recall. Finally, she characterized users' behaviour on the basis of a series of thirteen criteria, among which the time taken to search for the information, i.e., an unweighted measurement of efficiency. Some experts (e.g., Meads and Stubbs, 2004) even go so far as to incorporate the criterion of

utility among the criteria of usability. In short, it seems that not only does usability function as a criterion for the global evaluation of the quality of interfaces but also that there is a certain confusion surrounding the concept of usability and its links with utility.

1.3 There is one more dimension: usability, utility and acceptance

This last position is well represented by the Technology Acceptance Model (TAM) and its discussion by Dillon and Morris (1996): "Given a usable and utilitarian tool, an important question is whether people think that they should use the tool or not. This is the domain of perception, where TAM most naturally fits. Thus, we suggest that one's subjective evaluation of the system is a third important determinant of use". Usability and utility key dimensions, but their respective perception is the third dimension to be taken into account.

An original view is presented by Le Coadic (1995) who considers that usability "can act as a causal antecedent of utility" which, for him, seems to mean that non-usability implies non-utility and therefore that utility implies usability. The author does not indicate how the results relating to utility and usability can be used. Le Coadic also does not consider intermediate cases in which, for example, an interface is moderately useful and fairly usable. In sum, different kind of relations between usability and utility are more or less precisely described in the literature. This paper sets out the logical description of the set of possible relations between these two variables for certain standard values of the variables. This framework permits the interpretation of any data as far as they concern usability and utility.

2. Interpretation of the relations between the variables of usability and utility

Let us consider UT et US as variables, the first one measuring the utility of an interface, the second one its usability, for example in an empirical evaluation of the interface with n participants. These variables can be multidimensional or not, continuous, discrete or binary. Let us consider for this first step of the presentation, that both are binary variables.

The relations between the variables of usability and utility are described by the frequency f of co-occurrences of the states of UT and US for n set of pairs of values 0 or 1 (see Table 1). The sum of these frequencies is obviously equal to 1.

[Insert table 1 about here]

In logic, truth table and many other symbolic tools (e.g. Carroll, 1896), permits the description of the relations between two binary variables. This furnishes a logical interpretation of the nature of the relation. For example, in Table 2 the third column corresponds to an implication from UT_1 to US_1 ($UT_1 \Rightarrow US_1$). This column describes interfaces where the users attain their goal (i.e., meet their need) when they succeed in using the interface (line 1), can succeed in using the interface but not attain their goal (line 2), cannot attain their goal without succeeding in using the interface (line 3), and can not attain their goal and not succeed in using the interface (line 4). $UT_1 \Rightarrow US_1$ means that using interface is necessary but not sufficient to the intended goal. If the usability is good then the

interface will be useful. Improving usability will probably have a positive effect on utility.

[Insert table 2 about here]

On the basis of the distribution of the frequencies in a contingency table, it is possible to infer the logical relation between these two variables. It is therefore simply necessary to treat a 2x2 contingency table as if it were a line in a truth table: frequencies = 0 correspond to "false" (i.e., impossible) states and frequencies > 0 to "true" (i.e., possible) states (Table 3). Let us consider that the sum = 1 of the frequencies > 0 corresponds to a uniform distribution of these frequencies. For example, Table 4 represent the contingency table corresponding to the third column $\underline{UT_1 \Rightarrow US_1}$.

[Insert table 3 about here]

[Insert table 4 about here]

Below are a number of examples of interpretations of the relations between these two variables for standard frequency values.

$\underline{\text{Inclusive OR, } UT_1 \text{ v } US_1}$. Interface is usable, useful or both: it is neither necessary nor sufficient (not particularly specific) to the intended goal.

Implication UT_1 => US_1. Utility implies usability: using interface is necessary but not sufficient to the intended goal. If the usability is good then the interface will be useful. Improving usability will probably have a positive effect on utility.

Independence, US_1. Interface is usable but it is only moderately useful for the intended goal.

Implication US_1 => UT_1. Usability implies utility: using interface is sufficient but not necessary for the intended goal. If the usefulness is good then the interface will be usable. Improving utility will probably have a positive effect on usability.

Independence, UT_1. Interface is useful, whatever the usability is: good or bad. The interface can be used or not, the goal is attained. The interface is a kind of placebo. Improving utility will probably have no result.

Equivalence UT_1 <=> US_1. Utility and usability are equivalent. Interface is necessary and sufficient (i.e. specific), for the intended goal. Improving usability will probably have a positive effect on utility .

Conjunction $UT_1 \wedge US_1$. Interface is perfect, usable and useful for the intended goal.

NAND (not And), US_1 / UT_1. Incompatibility. Interface hinders the intended goal.

Exclusive OR, US_1 w UT_1. Interface hinders the intended goal.

Independence, UT_0. Interface useless for the intended goal even though moderately usable.

Implication UT_0 => US_1. Interface is usable but useless for the intended goal.

Independence, US_0. Interface unusable for the intended goal.

Implication $UT_1 \Rightarrow US_0$. Interface paradoxical.

NOR (not Or) UT_1 ? US_1. Interface bad or unsuitable for the intended goal.

Tautology. All the relations between use and goal are true.

3. Processing

A very simple way of processing the data is proposed. First of all, a $?^2$ (Chi Square) calculation can be performed to check whether the distribution of the values in the contingency table is significantly different from a uniform distribution (table corresponding to the tautology: All the relations between use and goal are true). If the observed distribution is significantly different from the uniform distribution then the procedure consists of identifying the two or three distributions of the model that are the closest to the observed distribution. Then, for each expected sample size $S_{expected}$ (i.e. the frequency defined by the model multiplied by the total sample size) the sum d of the differences for the observed sample size S_{obs} is calculated.

$d = ? \ | S_{obs} - S_{expected} |$

This difference is thus the measure of the distance between the different distributions tested and the observed distribution. All that needs to be done is to compare the values d. The smallest value of d corresponds to the theoretical reference distribution.

We might, for example, want to evaluate an “interactive terminal” type interface the purpose of which is to allow users to buy transport tickets and the usability of which is evaluated by means of a single criterion, namely “utilization

time less than 5 minutes" (let us imagine, for example, that these 5 minutes represent the mean time required to attain the goal with the "reference tool", i.e., the ticket office).The evaluation consists of asking a representative group of users of this mode of transport (let us say 50 individuals) to buy a ticket for a destination A, B, C or D (where each of the destinations is representative of the set of destinations accessible via this mode of transport). UT is measured by checking that the user actually purchased his or her ticket for the intended destination (UT=1 if purchase completed successfully, UT=0 otherwise). US is measured on the basis of whether the time is greater than (US=0) or less than (US=1) 5 minutes. The evaluation simply consists of indicating the number of users out of the 50 in each field of the contingency table (e.g. data in Table 5).

[Insert table 5 about here]

We now check that the distribution is different from a uniform distribution ($?^2$ = 20.2; $p < 0.01$). We can now ask whether this distribution corresponds more closely to equivalence UT1 <=> US1 (Table 6) or to implication UT1 => US1 (Table 7).

[Insert table 6 about here]

[Insert table 7 about here]

In the first case, $\underline{d} = 0+3+10+13 = 26$

In the second case, $\underline{d} = 8.5+3+6.5+4.5 = 22.5$

We can therefore conclude that the interface is necessary but not sufficient to the intended goal. Improving usability will probably have a positive effect on utility. We would be especially troubled by the fact that 22 users out of 50 did not succeed in purchasing a ticket and, in particular, by the fact that 10 of these 22 users spent less than 5 minutes at the terminal (they gave up more quickly than the mean waiting time at the ticket office).

While a contingency table is always easy to interpret by calculating $\underline{d}$, it is often difficult to interpret a set of points. Main principal component analyses and other techniques of data analysis can provide "summaries" (straight lines, curves) which do not fall within our interpretational framework. On the other hand, techniques like the implicative analysis (e.g. Bernard, 1999; Gras et al., 1999) makes it possible to process in the same way any set of discrete or continuous non-binary variables. For example, it is possible to process usability as a five dimensions variable, to describe causal links between these five sub-variables, and to describe the causal links between this causal tree and utility variable(s).

4. Application: Utility and usability of an encyclopaedic dictionary for pupils

An interest of our framework undoubtedly lies in the ability to compare interfaces and to interpret a series of different tasks processing results. For example, in a recent study (Tricot and Lafontaine, 2002) we explored the links between learning the meaning of new words and the use of dictionary. We performed an experiment involving children in the 4^{th}, 5^{th} and 6^{th} grades (the

children being 8 years old at the start of the 4th grade and 11 years old at the end of the 6th grade). We asked pupils to write down words definitions either with or without the aid of an encyclopaedic dictionary. The words were either known to the pupils (in which case, the dictionary was, a priori, useless to them since there was nothing to learn), or unknown to them (in which case, the dictionary was, a priori, useful to them since there were words to learn).

Forty nine pupils took part in the experiment. They completed the task individually. The pupils were given a 6-page notebook with one word to be defined on each page. The words were: Epitaph, Bruise, Didascalie, Snail, Bottle, Staircase. In French, these words consist of either three or four syllables and occur at the beginning of the alphabet. Three of them were known to the pupils and three of them were not (this fact was checked in our experiment). The word presentation order was rotated so that, overall, each word appeared with the same frequency on each page of the notebook (i.e., 5 to 6 times).

The following instructions were given: “I’m going to give you a list of words. You know some of them and some you don’t. You must try to write a definition for each word. So, two things are possible. You might think you need a dictionary so that you can look for the definition. But, watch out. I’m only going to give you one and a half minutes to find the definition. Then I’ll ask you to tell me the definition. Or you might think that you don’t need the dictionary and we’ll go on to the next word”. After the initial definition, the pupils were asked whether they were certain, not very certain or totally uncertain about the definition of each word.

The dictionary was closed after 1 minute 30 seconds or when the definition had been found thus making it impossible for the pupils to recopy the definitions.

The results show that the pupils in the 4th grade did not succeed in using the provided dictionary within the time allowed. When they were allowed to take their time, they needed between 8 and 18 minutes to find the definitions. We therefore decided not to continue after the first 4 pupils since they had been disconcerted by this excessively difficult task.

The tool appears to be unsuitable since there is something to be learnt yet it is impossible to use the material. To learn a new word from a dictionary, it is necessary to have learnt how to use the dictionary.

For the 45 pupils in the 5th and 6th grades who looked up 6 words (i.e., 270 definitions), we observed that the dictionary was almost a "perfect" tool for the unknown words (see Table 7). It was 80% useful and useable. To use Gras' terminology (Gras et al., 1999), there was a quasi-conjunction relation of .80 between the learning of the new words and the use of the dictionary. In contrast, in the case of the known words, even though the dictionary is useable (and used) it is useless. The pupils appeared to overuse the dictionary.

[Insert table 8 about here]

5. Discussion

In this paper, we have proposed a formal framework for the interpretation of fifteen possible relations between two variables: usability and utility. We have

considered the utility of an interface is the compatibility between the purpose of the interface and the user's goals, but it is also possible to consider that an interface is useful when it just meet user's goals and not the purpose of the interface. This framework makes it possible to interpret the results obtained in tests of interfaces and, more generally, tools developed for specific purposes provided that the tests measure usability and utility independently of one another. Furthermore, in experimental research into the use of the interface of this framework suggests we can describe users' activities by incorporating utilization variables and variables relating to goal attainment levels.

We have presented the way in which the results are processed in only a rudimentary fashion. In most situations, and in particular during the design phase, this is more than adequate. We believe that a qualitative analysis of the contingency tables is more interesting than a quantitative approach. However, in other cases it may be of interest to be in a position to draw statistical inferences concerning the results.

6. Bibliography

Bernard, J.M., 1999. Implicative analysis for multivariate binary data using an imprecise Dirichlet model. 1st International Symposium on Imprecise Probabilities and Their Applications, Ghent, Belgium, 29 June – 2 July.

Blandford, A., Wong, B.L.W., 2004. Situation awareness in emergency medical dispatch. International Journal of Human-Computer Studies, 61, 421-452

Buckingham Shum, S., McKinght, C., (Eds.), 1997. Web usability. International Journal of Human Computer Studies, special issue, 47, 1-222.

Carroll, L., 1896. Symbolic Logic. Part I. Elementary. Macmillan, Londres.

Diamadis, E.T., Polyzos, G.C., 2004. Efficient cooperative searching on the Web: system design and evaluation. International Journal of Human-Computer Studies, 61, 699-724.

Dillon, A., Morris, M., 1996. User acceptance of information technology : theories and models. Annual Review of Information Science and Technology, 31, 3-32.

Eason, K.D., 1984. Towards the experimental study of usability. Behavior and Information Technology, 3, 133-143.

Gras R., Briand, H., Peter, P., Philippé J., 1997. Implicative statistical analysis, in: Proceedings of the Fifth Conference of the International Federation of Classification Societies, Kobe, Springer-Verlag, Tokyo, pp. 412-419.

Grinstein, G., Kobsa, A., Plaisant, C., Shneiderman, B., Stasko, J.T., 2003. Which comes first, usability or utility?, in: 14th IEEE Visualization Conference, October, pp. 112.

Grudin, J., 1992. Utility and usability: Research issues and development contexts. Interacting with Computers, 4, 209-217.

Höök, K., 1997. Evaluating the utility and usability of an adaptive hypermedia system, in: Proceedings of the 2nd international conference on Intelligent user interfaces, January, pp. 179-186.

Keates, S., Clarkson, P.J., 2001. Combining utility, usability and accessibility methods for universal access, in: ACM CHI Workshop on Universal Design, Seattle, [available on line : http://www.ics.forth.gr/proj/at-hci/chi2001/files/keates.pdf]

Le Coadic, Y.F., 1997. Usages et usagers de l'information. Nathan, Paris.

Meads, J., Stubbs, D.D., 2004. Usability Architects, [available on line : http://www.usabilityarchitects.com]

Mizzaro, S., 1998. Relevance, the whole history, in: Bellardo Hahn T., Buckland M. (Eds.), Historical studies in information science, Journal of the American Society for Information Science, special issue, 221-243.

Nielsen, J., 1993. Usability engineering, Academic Press, Boston.

Nielsen, J., 2003. Usability 101: Introduction to Usability, Jakob Nielsen's Alertbox, Aug 25, [available on line : http://www.useit.com].

Park, S., 2000. Usability, user preferences, effectiveness and user behaviors when searching individual and integrated full text databases: implications for digital libraries. Journal of the American Society for Information Science, 51, 456-468.

Scapin, D.L., Bastien, J.M.C., 1997. Ergonomic criteria for evaluating the ergonomic quality of interactive systems. Behavior and Information Technology, 17, 220-231.

Tricot, A., Lafontaine, J., 2002. Une méthode pour évaluer ensemble l'utilisation un outil multimédia et l'apprentissage réalisé avec celui-ci. Le Français dans le Monde, numéro spécial, Janvier, 41-52.

List of tables

	US = 1 (usable)	US = 0 (not usable)
UT = 1 (useful)	$\underline{f}\ UT_1US_1$	$\underline{f}\ UT_1US_o$
UT = 0 (not useful)	$\underline{f}\ UT_0US_1$	$\underline{f}\ UT_0US_0$

Table 1

Contingency table

UT_1US_1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	1
UT_0US_1	1	1	1	0	0	0	0	1	1	1	1	0	0	0	1
UT_1US_0	1	0	0	1	1	0	0	1	1	0	0	1	1	0	1
UT_0US_0	0	1	0	1	0	1	0	1	0	1	0	1	0	1	1
Interpretation	v	=>	US_1	<=	UT_1	<=>	∧	/	w	UT_0	UT_0=>US_1	US_0	UT_1=>US_0	↓	T

Table 2

Truth table

UT_1US_1	.33	.33	.5	.33	.5	.5	1	0	0	0	0	0	0	0	.25
UT_0US_1	.33	.33	.5	0	0	0	0	.33	.5	.5	1	0	0	0	.25
UT_1US_0	.33	0	0	.33	.5	0	0	.33	.5	0	0	.5	1	0	.25
UT_0US_0	0	.33	0	.33	0	.5	0	.33	0	.5	0	.5	0	1	.25
Interpretation	v	=>	US_1	<=	UT_1	<=>	∧	/	w	UT_0	UT_0=>US_1	US_0	UT_1=>US_0	↓	T

Table 3

Truth table with real values replaced by probabilities

	US = 1 (usable)	US = 0 (not usable)
UT = 1 (useful)	.33	0
UT = 0 (not useful)	.33	.33

Table 4.

Contingency table corresponding to $\underline{UT_1 \Rightarrow US_1}$

	US = 1 (usable)	US = 0 (not usable)
UT = 1 (useful)	25	3
UT = 0 (not useful)	10	12

Table 5.

Observed data

	US = 1 (usable)	US = 0 (not usable)
UT = 1 (useful)	25 (f =.5)	0
UT = 0 (not useful)	0	25 (f =.5)

Table 6.

Equivalence UT_1 <=> US_1. Utility and usability are equivalent. Interface is necessary and sufficient (i.e. specific), for the intended goal.

	US = 1	US = 0
UT = 1	16.5 (f =.33)	0
UT = 0	16.5 (f =.33)	16.5 (f =.33)

Table 7.

Implication UT_1 => US_1. Utility implies usability: using interface is necessary but not sufficient to the intended goal.

	Known words		Unknown words	
	Successful use	Unsuccessful use or Not used	Successful use	Unsuccessful use or Not used
Learnt	3 (f =.02)	6 (f =.04)	108 (f=.80)	0 (f =.00)
Not learnt	90 (f =.67)	36 (f =.27)	22 (f =.16)	5 (f =.04)

Table 8.

Relations between learning and use as a function of whether the words are known or unknown