Int. J. Human-Computer Studies (2001) **55,** 741–760
doi:10.1006/ijhc.2001.0501
Available online at http://www.idealibrary.com on IDEAL®

Effects of training and representational characteristics in icon design

RAVINDRA S. GOONETILLEKE, HELOISA MARTINS SHIH, HUNG KAI ON AND JULIEN FRITSCH
Department of Industrial Engineering and Engineering Management, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong. email: ravindra@ust.hk

(Received 11 January 2000, accepted in revised form 15 June 2001, and published electronically 4 September 2001)

Icons are a very important component of graphical user interfaces. However, icon design is still predominantly artistic in nature and as a result icon selection is generally based on usability evaluations after a set of alternative icons are developed. This process tends to be time-consuming and costly. In this research, we address the issues of *what* should be depicted in an icon, given the function it should represent, and how *training* affects the performance of novice users when using an iconic interface. A set of 36 concrete icons (12 functions) were selected and tested with a total of 30 participants. The experimental results indicate that complete representations are generally superior for both untrained and trained participants. Results also show that trained participants had shorter response times when compared to untrained participants. Further analysis suggests that *ambiguity*, *uniqueness* and *dominance* are three important aspects to consider when designing and developing icons. Applications of this research include the design of appropriate icons for graphical user interfaces prior to usability testing and the importance of a short training period to illustrate the composition of an icon in an effort to improve the mental model associated with each design.

KEYWORDS: icons; design; usability; training GUI; graphic representation; web; e-commerce; semiosis; semantics.

1. Introduction

With the evolution of the World Wide Web and e-commerce, imagery has become the most important communication "medium" in computer and consumer product user interfaces. There are many different guidelines for building icon-based interfaces (examples include Easterby & Zwaga, 1980; Carrol & Thomas, 1982; Hemenway, 1982; Lodding, 1983; Gittins, 1986; Shneiderman, 1987; Waterworth, Chignell & Zhai, 1993). Even though iconic interfaces are common and have distinct advantages (Paivio, 1971; Lodding, 1983; Rogers, 1989), the characteristics that make them effective and appropriate are not well known (Gittins, 1986; Rogers, 1989; Blankenberger & Hahn, 1991; Garg & Plocher, 1999; Goonetilleke, Shih & Kurniawan, 2001). Icon designs are and can be limited by implementation constraints and sometimes are restricted by those that are in existence already, thereby lacking originality in some sense (Gittins, 1986). For the most

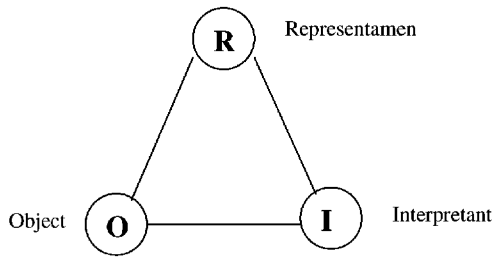


FIGURE 1. The components related to the interpretation of a sign (S) (Nadin, 1988).

part, *elegance* and *simplicity* have governed the election or selection of elements for visual design even though human factors and psychology research such as Landsdale, Simpson and Stroud (1990) has found that recall performance is related to the "semantic-fit" of a representation. The bottom-up approach proposed by Waterworth *et al.* (1993) for hypermedia systems is another that can be used for interface design based on the functions of low-level features.

Icons have evolved from the concept of signs, which are defined as "something that stands to someone for something in some respect or capacity" (Peirce, 1932, p. 135). Hence, Peirce viewed a sign (S) as a product of a three-way interaction (Figure 1) between the *representamen* (that which represents), the sign's *object* (that which is represented) and its mental *interpretant* (the process of interpretation),

$$S = S(R, O, I). \tag{1}$$

Horton (1994) indicated that icons alone are meaningless without a particular context and suggested the following relationship for icons:

$$\mathbf{Icon}_i + \mathbf{context}_j + \mathbf{viewer\ (or\ interpretant)}_k \;\Rightarrow\; \mathbf{meaning}_{ijk}. \tag{2}$$

Thus, Figure 1 may be modified to reflect context as shown in Figure 2. The aims of our study can be best understood considering Figure 2. We attempted to investigate how to enhance the O–R relationship and the R–I relationship. The I–O relationship, on the other hand, is more difficult to penetrate since it is an inherent function of the person (intepretant) or "culture". *Interpretation* also called expression (R–I relationship) is the process of understanding the meaning of a sign. It is the same as identifying the relationship between the object (O) and the representamen (R). Since each person (interpretant) is unique and possesses a certain cultural and social bias, the way in which the sign object is recalled (that is, the mental model) by the representamen can be very different for each person (Bourges-Waldegg & Scrivener, 1998; Choong & Salvendy, 1998).

The function of *representation* involves emphasizing the relationship between the representamen and the object that is represented (O–R relationship). In this relationship, the effectiveness of the representation depends on *what* is represented and *how* it is represented. Waterworth *et al.* (1993) found that good icon designers do not make good icon judges, or vice versa thereby implying that the design process is still artistic in nature.

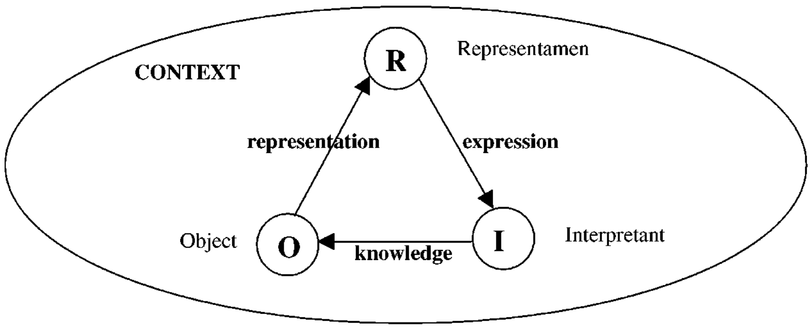


FIGURE 2. The Icon process.

There are many different classification schemes for icons (e.g. Smith, Irby, Kimball & Verplank, 1982; Lodding, 1983; Jervell & Olsen, 1985; Gittins, 1986). Historically, in the context of what is to be represented, *semiosis* (the sign process) has been limited to representing only objects. The reasons may be two-fold. Firstly, an icon has been considered to be a graphical symbol to represent objects in a computer system (Gittins, 1986). Secondly, icons have evolved from pictorial representations of objects even though, in recent times, software developers have extrapolated the concept of an icon to go beyond "object" and encompass actions (Blankenberger & Hahn, 1991; Galitz, 1996). However, actions (verbs) manipulate objects in specific ways and are not easily represented compared to their counterparts, the objects (nouns) (Apple Computer, 1996). Hence, what should be represented becomes difficult if the icon is to convey meaningful information in the simplest way so that the image possesses perceptual immediacy. In this respect, Mullet and Sano (1995) have suggested three principles for simplicity.

1. *Unified* to produce a coherent whole.
2. *Refined* to focus on the viewer's attention on the essential aspects.
3. *Fitness* or *appropriateness* of the solution to the problem.

An established way to design icons is based on unifying individual icons into a collective metaphor (Gittins, 1986). Thus, icons are meant to correspond with real objects with which the users are familiar. Limitations may arise due to a lack of a direct mapping between the real objects and the system objects. An even greater problem arises if icon designers use the same or similar metaphor in different contexts thereby causing confusion for the users. Even though the same concept can be applied at a lower level to individual icons, unification to produce a coherent whole is generally perceived to result in an increase in the GUI density and thus, simplicity has been achieved only through a minimization of parts (or elements) rather than by true unification. As pointed out by Barnard and Marcel (1984), understanding elements and compounds is an important area that needs further investigation. In this paper, we attempt to investigate the importance of true unification using elements that have a semantic-fit.

How the object is represented depends on the *semantics* (intended meaning of the sign), which addresses the direct relationship between the representamen and the sign object.

Lodding (1983) has proposed three forms for such a relationship.

1. *Representation icons* that typically represent images of the object.
2. *Abstract icons* that attempt to visualize a concept that is not far from the concrete image.
3. *Arbitrary icons* that have no obvious reference to their intended meaning.

A fourth form of representation is the use of text. Apple Computer (1996) has suggested that text should not be used in icons, as text is often confusing. However, Horton (1994) has indicated that icons and words are not enemies and are not mutually exclusive. Some researchers such as Paivio (1971) have argued that multiple modalities enhance memorability and hence text and graphics together may be more effective than pure graphics. Kacmar and Carey (1991), and Egido and Patterson (1988) have shown that text and graphics together improve performance. In addition, the text in languages such as Chinese and Japanese is somewhat graphical and its appropriateness is worthy of consideration (Choong & Salvendy, 1998; Goonetilleke *et al*., 2001). As a result, we attempted to evaluate the inclusion of Chinese text in icons.

What an icon represents may not be obvious when one encounters it the very first time even though it may be possible to guess the meaning. In this respect, a component that has somewhat been neglected in relation to icons, but could improve the R–I relationship, is user training. Somberg (1987), Anderson (1990) and many others have shown that practice does improve user performance with menu-driven systems. Training can help impart the designer mental model for ease of understanding. Thus, we hypothesize that *learning through the design process* (or making the design process visible) with a short period of training can convey the designer mental model of the icon meaning quite effectively. Galitz (1996) has indicated that getting accustomed to graphical user interfaces may require about 8 h of training while other experts estimate the learning time to be in the order of 20 or 30 h. In this study, we also attempt to evaluate the effect of a brief training session to find out whether the R–I relationship can be enhanced. We believe that the numerous icons that go unused are the result of a mismatch between the user mental model and the designer mental model (Norman, 1988). Furthermore, we hypothesize that training geared towards the design process can effectively lure even experienced users to use icons effectively.

Finally, it is hoped that the results of this study and the science of semiotics will aid the icon designer to design icons that can effectively convey their meaning.

2. Hypotheses

The two hypotheses of this study were as follows.

1. Users who are given an understanding of the icon construction scheme (visible design process) through a short training (trained group) are able to perform better than those without such an exposure (untrained group).
2. A unified (complete) representation (verb + object) of a function leads to improved performance compared to a representation with either verb or object. In other words, the "sum" of the elements has improved performance when compared to each element.

3. The participants

Thirty Hong Kong Chinese participated in the experiment. There were 20 male and 10 female participants, 11 of them were office staff and 19 were university students. The age range was 19–36 years, with an average age of 24.4 years. Participants received a maximum of HK$30 for their time. The actual remuneration depended on each individual's performance. Each incorrect answer was penalized by deducting HK$0.50 from the initial HK$30. On average, each participant received HK$29.62.

4. Selection of functions and design of icons

Horton (1994) has suggested that icons be first made in black and white, and, colour be added to make them work better. It is critical that the colour and detail used in an icon be equally effective in monochrome and colour variations due to the variations of bit depth and hue of computers and printers. Hence, all the icons used in the experiment were grey scale images.

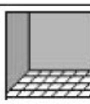
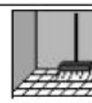
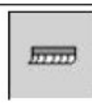
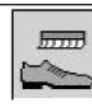
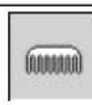
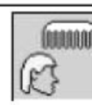
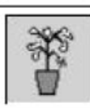
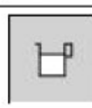
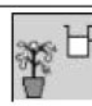
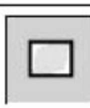
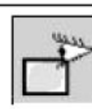
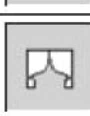
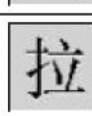
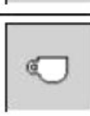
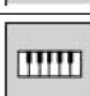
The theme (or context) of the experiment was activities in daily life. Twelve activities (that is, functions) were selected for the experiment. The functions were selected to ensure a similar level of familiarity for all participants (Table 1). All functions contained two elements: a verb or action (e.g. "to sweep") and an object (e.g. "the floor").

For each of the functions, three alternative icons were designed, in three alternative representations: representation 1 included only the *object* of the function; representation 2 included only the *verb* of the function; and representation 3, also called *complete* (or unified) representation, included both the *verb and the object*. The representations of each function are presented in Table 1.

The object part of all functions was portrayed graphically, using a pictorial representation of a concrete object (Lodding, 1983). The verbs of each function were represented either graphically or using a Chinese character corresponding to a part of the whole Chinese word for that verb. The graphical representations of the verbs were concrete images of the tools used to realize the actions. For example, a broom for "sweep", a brush to "brush", a comb for "comb" and so on. Chinese characters were used to represent the other six verbs as actions such as "play", "drink" and so forth are difficult to represent graphically. Generally, in the Chinese language, an action is written as one Chinese character and sometimes as a combination of two characters. Chinese characters take two written forms: a simplified form predominantly used in Mainland China and a more complex form used in Hong Kong and Taiwan. Simplified Chinese characters were used in this experiment. Six functions had the verb representation using Chinese characters, and the other six had graphical representation of the verb. The complete representation was a combination of representations 1 and 2 of the corresponding function. As a result, six of the complete icons consisted of a pictorial object and a Chinese verb while the remaining six consisted of a pictorial object and a pictorial verb, as may be seen from Table 1.

Each function had three associated tasks, which were three different expressions of the function. For example, the function "sweep the floor" was associated with the tasks: (1) "Sweep the floor, please"; (2) "Try sweeping the floor now"; and (3) "The floor is dirty, sweep it". Three tasks were used to eliminate any particular icon associations as a result of short-term memory.

TABLE 1
Icons used in the experiment and their corresponding functions

Function	Representation 1 (Object)			Representation 2 (Verb)			Representation 3 (Complete)		
A	1		Floor 地板	2		Sweep 打掃	3		To sweep the floor 打掃地板
B	4		Cards 紙牌	5		Play 玩	6		To play cards 玩紙牌
C	7		Shoe 鞋	8		Brush 擦	9		To brush a shoe 擦鞋
D	10		Hair 頭髮	11		Comb 梳理	12		To comb the hair 梳理頭髮
E	13		Hamburger 漢堡包	14		Eat 吃	15		To eat a hamburger 吃漢堡包
F	16		Dishes 碗碟	17		Wash 清洗	18		To wash dishes 清洗碗碟
G	19		Plant 植物	20		Water 淋	21		To water a plant 淋植物
H	22		TV 電視	23		Watch 看	24		To watch TV 看電視
I	25		Bicycle 單車	26		Ride 踏	27		To ride a bicycle 踏單車
J	28		Curtains 窗簾	29		Draw 拉	30		To draw the curtains 拉窗簾
K	31		Coffee 咖啡	32		Drink 飲	33		To drink coffee 飲咖啡
L	34		Piano 鋼琴	35		Play 彈	36		To play piano 彈鋼琴

5. Experimental apparatus and software

The experiment was programmed using HTML 4.0 and run on a Pentium 200 MHz personal computer equipped with a NEC MultiSync A700 17″ monitor, and an Accutouch touch screen (Elo Ltd.) through the university intranet. This allowed the possibility of having multiple users worldwide even though this function was not used in this experiment. The experimental configuration and the participant results were stored using MS Access 97 database. The ASP script language was used for the communication between the HTML program and the database.

6. Experimental task

A 2 (presence or absence of training) $*$ 12 (functions) $*$ 3 (representations of object, verb, complete, nested within functions) experimental design was used. The functions and their representations are presented in Table 1.

The 30 participants were divided into two groups of 15. One group received a short training before using the icons, and represented "experienced" users, while the other group was not trained, representing novice or occasional users.

Participants with training needed to perform three tasks. First, they were given a training task that showed the design rationale of each icon and how the complete representation icon (i.e. representation 3) was built (see Figure 3). As can be seen from the figure, the training was an animation performed function by function, showing the description of the function first, followed by each of the two single-element icons and the explanations of those representations, and finally the complete representation. The numbers in the figure represent the sequence of the animation. These numbers were not shown to the participants. The participants were advised to take the training seriously and were informed that there would be a quiz following the training. They were also informed that if they made more than one mistake in the quiz, they will not be allowed to further participate in the experiment.

The quiz was paper-based and the participants were asked to select the functions that corresponded to each of the 24 single-element (object only and verb only) icons. Complete icons (representation 3) were not tested, since the participants could "short-circuit" the answers of the single-element icons from their corresponding complete representations.

Following the paper-based quiz, the participants performed the experiment on a computer. Every participant was given an interface-training task before actually starting the experiment. The participants were asked to select the most appropriate icon for a given task. Each participant was given 36 tasks corresponding to the 36 icons. The tasks were blocked by representation. That is the 24 single-element representations of the functions (representations 1 and 2 in Figure 1) were performed first, followed by the 12 complete representations (representation 3) at the end. The aim of this blocking was to avoid training (or further training) of the participants, as the complete representations were formed by a combination of the single-element representations. The order of the tasks within each block was random.

Figure 4 shows an example of a task screen. The description of the task (or function) was shown first in English and in Chinese (e.g. "Ride the bicycle" as shown in the upper

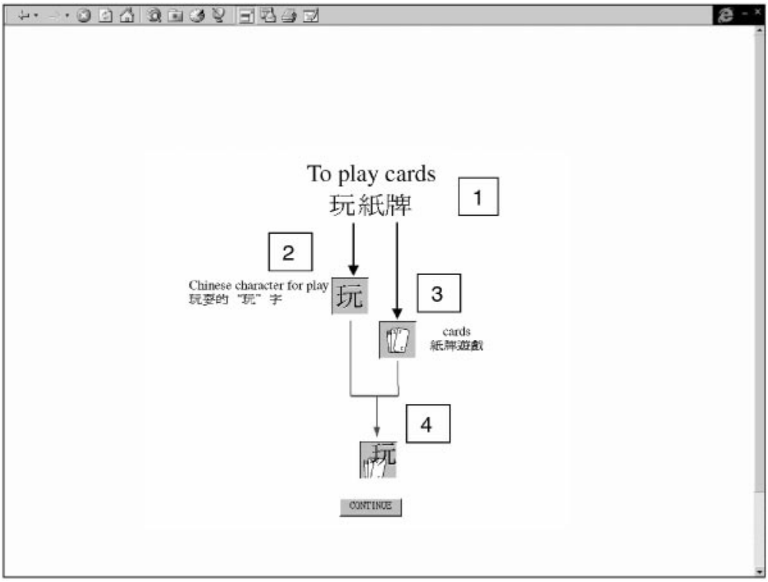


FIGURE 3. Training screen related to function "To play cards".

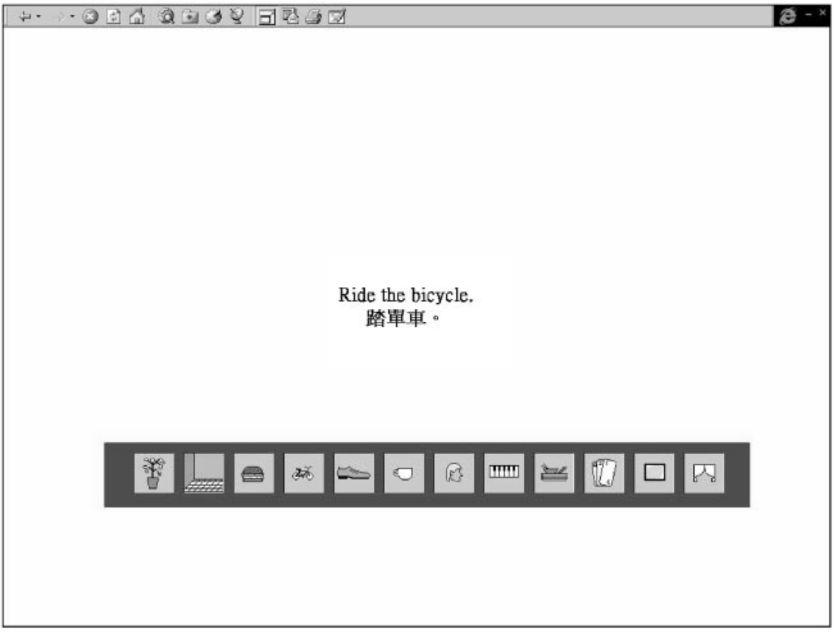


FIGURE 4. Task screen for "Ride the bicycle".

part of the figure). After the participant had read and understood this description, he/she would touch the screen. A tool bar containing 12 icons would then appear (as shown in the bottom part of Figure 4) corresponding to one type of representation (for example Figure 4 contains only object representations) among the different functions. Therefore, the 12 icons in each tool bar belonged to one single column of Table 1. The order of the icons within the tool bar was random and changed for each task.

As soon as the tool bar appeared, the timer would start. The timer stopped when the participant selected an icon by touching the screen. If the participant selected the wrong icon during the first attempt, a second attempt was allowed. The second chance was aimed at eliminating any incorrect associations of the function and icon, which could result in errors in subsequent tasks. If the participant made a mistake in the second attempt also, the experiment would continue to the next task without any additional chances.

After the participants finished the 36 tasks, their results were displayed on the screen and the participant payment was made based on the hits and errors.

The untrained participants were also required to perform the 36 tasks using the icons. The difference was that they had *no* previous knowledge of the icons that were tested. The tasks were exactly the same as the tasks for the trained group (Figure 4), with the same blocking.

By dividing the participants into trained and untrained groups, we aimed to compare the performance (time and accuracy) of "experienced" (trained) and "naïve" (untrained) users when using the 36 icons. The performance data were converted and stored in an MS Access database automatically at the end of the experiment.

All participants were also asked to rate the appropriateness of each icon. Due to length and space limitations, the appropriateness ratings and their results will not be discussed in this paper.

7. Results

Two dependent variables were used to evaluate participants' performance: *time* to select the appropriate icon and *accuracy* in the selection.

Table 2 shows the mean and standard deviation of response time and accuracy for trained and untrained groups. Accuracy for each task was 1 (for correct answer at the first attempt), 0.5 (for correct answer at the second attempt), 0 (for incorrect answers at both attempts). There were a total of 19 incorrect responses in the first attempt: 2 from the trained group and the remaining 17 from the untrained group. These responses were dropped from the response time analysis. In other words, the time reported in Table 2 includes only the hits (correct answers) in the first attempt.

The average response time per function ranged between 2.96 and 4.56 s, with standard deviations ranging between 1.47 and 4.65 s. The mean accuracy of each function varied between 0.94 and 1 with standard deviations ranging between 0 and 0.22.

The normality tests for time and accuracy showed that neither response time nor accuracy had normal distributions at $p = 0.05$ level. An inverse square root transformation on response time ($1/\sqrt{t}$) resulted in a normal distribution with $W = 0.9867$, $p = 0.4768$. Hence, all subsequent analyses were performed using this transformation. As most accuracy results are 1 or close to 1, no transformation could achieve normality for this variable.

TABLE 2
Summary data for trained/untrained groups

Group	Time (s) Mean	Time (s) S.D.	Accuracy Mean	Accuracy S.D.
Trained	2.648	1.275	0.9981	0.0304
Untrained	4.205	3.333	0.9805	0.1143
Overall	3.416	2.627	0.9893	0.0840

The analysis of variance (ANOVA) on time (inverse square root of time) and accuracy are shown in Table 3. The shaded cells show significance at the $p < 0.05$ level.

To emphasize that representations are nested within functions, the standard statistical notation of representation (function) will be used hereafter. The ANOVA for response time showed significant effects of training and representation (function) as well as of training $*$ representation (function) interaction at the $p < 0.05$ level. Function and its interaction with training had no significant effect at the $p < 0.05$ level.

Figure 5 shows the interaction plot of the effect of training $*$ representation (function) on response time. Table 4 shows the results of the *post-hoc t*-tests for each of the two training groups (note: *t*-tests were performed as it was a nested design). In general, for the untrained group, representation 1 (object only) and representation 3 (complete) had significantly shorter response times compared to representation 2. The only exceptions were functions A and D, where representation 2 (verb only) and representation 3 had significantly faster response times compared to representation 1.

Also, the trained group had results similar to the untrained group, only with lesser significant differences in response time among the three representations.

The ANOVA on accuracy (Table 3) showed significant effects of representation as well as of training $*$ representation (function) interaction at $p = 0.05$ level. No significant effects were seen for the factors training, function and their interaction.

Figure 6 shows the interaction plot of the effect of training $*$ representation (function) on accuracy. As seen from the figure, the main source of errors is representation 2 of functions K and L. Representation 2 of function K had 7 incorrect responses in the first attempt (36.8% of all incorrect selections in the first attempt) and 4 in the second attempt (100% of all mistakes in the second attempt). Representation 2 of function L had 3 incorrect selections in the first attempt, corresponding to 15.8% of the incorrect responses. As a result, the ANOVA was repeated ignoring functions K and L. The results (Table 3) show that there are no significant effects on accuracy when these two functions are dropped.

8. Discussion

The ANOVA confirms hypothesis 1—trained participants had shorter response times than untrained participants. Even though the training was very short, with a total duration of less than 2 min, the average response time was 37% shorter for the trained group.

TABLE 3

ANOVA results for time and accuracy

Independent Variables	DF	Dependent variables			
		$1/\sqrt{\text{Time}}$	Accuracy		
			All functions	(K removed)	(K, L removed)
Training	1	$F(1, 24) = 121.73$ $p = 0.0001$	$F(1, 24) = 2.60$ $p = 0.1201$	$F(1, 22) = 3.56$ $p = 0.0726$	$F(1, 20) = 2.78$ $p = 0.1112$
Function	11	$F(11, 24) = 0.373$ $p = 0.9563$	$F(11, 24) = 0.83$ $p = 0.6118$	$F(10, 22) = 1.01$ $p = 0.4626$	$F(9, 22) = 1.98$ $p = 0.979$
Representation (function)	24	$F(24, 1008) = 13.81$ $p = 0.0001$	$F(24, 1008) = 5.80$ $p = 0.0001$	$F(22, 924) = 1.21$ $p = 0.2299$	$F(20, 840) = 0.56$ $p = 0.9421$
Training * function	11	$F(11, 24) = 0.86$ $p = 0.5866$	$F(11, 24) = 0.85$ $p = 0.5920$	$F(10, 22) = 0.74$ $p = 0.6771$	$F(9, 20) = 0.80$ $p = 0.6189$
Training * representation (function)	24	$F(24, 1008) = 1.89$ $p = 0.0063$	$F(24, 1008) = 5.97$ $p = 0.0001$	$F(22, 924) = 1.56$ $p = 0.0495$	$F(20, 840) = 1.00$ $p = 0.4594$

Note: The degrees of freedom column corresponds to the complete experimental design when no cases are omitted from the analysis

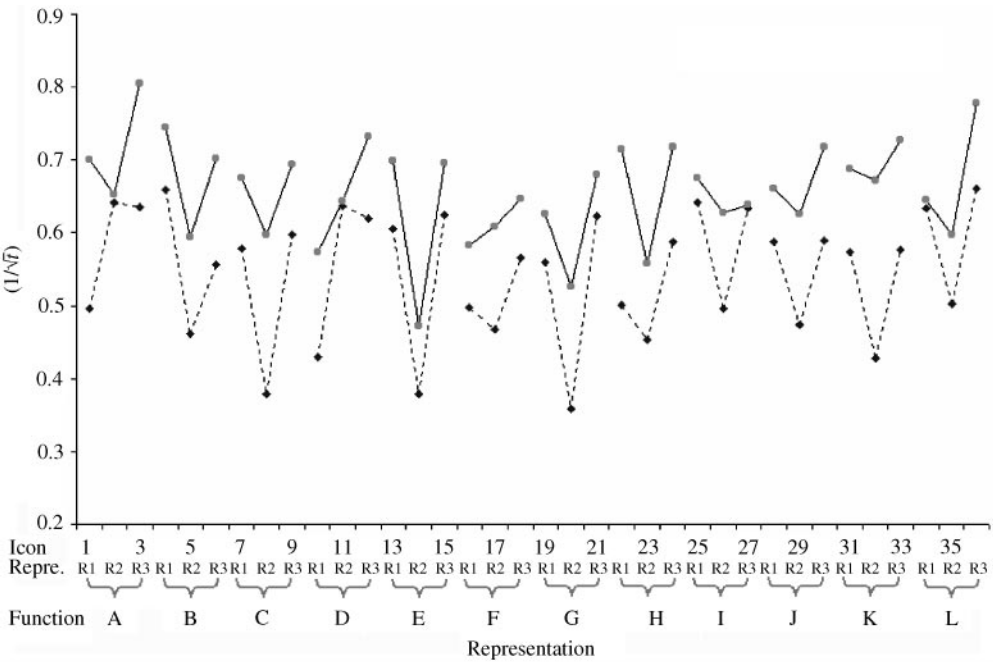


FIGURE 5. Training * representation (function) interaction plot for time: --◆--, untrained; —■—, trained.

TABLE 4
Representation mode groupings based on the post-hoc t-tests, for the dependent variable, response time

Function	Untrained Grouping	Untrained *p*-value	Trained Grouping	Trained *p*-value
A	2 3 1	0.0008	3 1 2	0.0024
B	1 3 2	0.0001	1 3 2	0.0024
C	3 1 2	0.0001	3 1 2	0.0719
D	2 3 1	0.0001	3 2 1	0.0016
E	3 1 2	0.0001	1 3 2	0.0001
F	3 1 2	0.0671	3 2 1	0.3672
G	3 1 2	0.0001	3 1 2	0.0030
H	3 1 2	0.0059	3 1 2	0.0003
I	1 3 2	0.0007	1 3 2	0.5355
J	3 1 2	0.0093	3 1 2	0.1064
K	3 1 2	0.0062	3 1 2	0.4435
L	3 1 2	0.0011	3 1 2	0.0002

For each grouping the response time increases from left to right. Underlined representations are not significantly different at the $p < 0.05$ *level.*

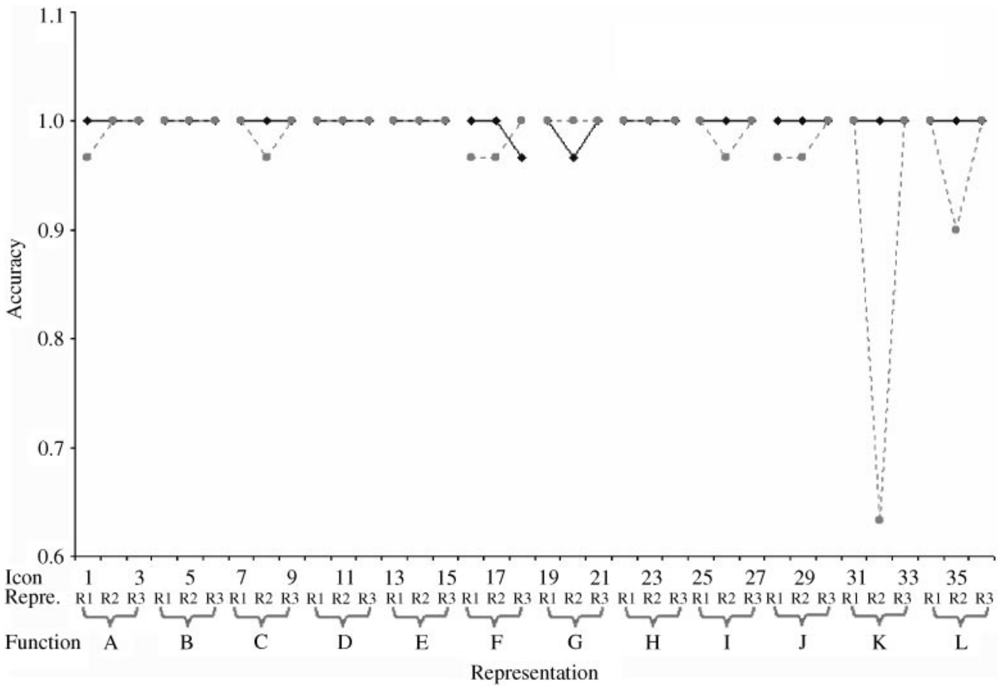


FIGURE 6. Training * representation (function) interaction plot for accuracy: —◆—, trained; ---◆--, untrained.

Hypothesis 2 was also confirmed with one exception. The only exception is function B for the untrained group, where representation 1 (object only) had a shorter response time than representation 3. We believe that this result is due to an artistic feature. Icon 6 (representation 3 of function B) poses an interesting characteristic related to the Gestalt principle of figure and ground. Even though we did not foresee this problem prior to the experimentation, we hypothesize that the identification cue for the cards is somewhat hidden behind the Chinese word. As a result, the selection time is longer and hence representation 3 is not as effective as the first representation. Therefore, in order to ensure that a design is unified, it is necessary that the perceptual qualities of the *figure* (Chinese characters in this case) are matched with those of the *ground* (cards). This result is similar to that of Knapp (1985) where it was found that subjects had difficulty identifying helicopter symbols depending on the background against which the symbol was displayed. In other words, the scale and *visual weight* (Mullet & Sano, 1995) of figure and ground have to be approximately equal to prevent any biases of perception. The elements of the function should be integrated in such a way that they complement each other to ensure harmony and balance.

The complete representation either had the shortest response time or was included in the group with the shorter response time in all other cases (Table 4). This confirms the hypothesis that icons with a truly *unified* representation of the functions have shorter response times, irrespective of user training. For the trained group, significant differences among the three representations were fewer. This result is in line with the Blankenberber

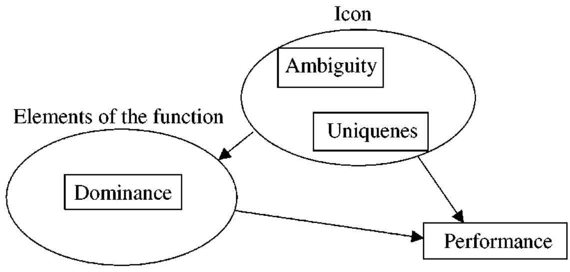


FIGURE 7. Performance model.

and Hahn (1991, p. 363) finding that " ... icon design seems to be of little influence on the performance of advanced users". The response times of some single-element icons (for example, icons 17, 26 and 32) which were relatively long in the untrained group "caught up" with the response times of the respective complete representations in the trained group. However, in some other cases, training enhanced the advantage of the complete representation over single-element representations when compared to the untrained group. Examples are functions A, D and L (Table 4), where the complete representation was equivalent to one of the single-element icons in the untrained group, but became significantly better than representations 1 and 2 in the trained group. One of the outcomes of this experiment is that the complete representation led to the shortest response time (or tied with the shortest time) for all functions for both the trained and untrained groups.

Further analysis for the untrained group shows an interesting pattern related to the single-element representations 1 and 2 (see Table 4 and Figure 5). With the exception of functions F and H, the performance for the single-element representations (1 and 2) was significantly different from each other for the other 10 functions. In functions A and D, the response times of the verb-only representations were significantly shorter than the response times of the object-only representations. In functions B, C, E, G, I, J, K, L the object-only representations led to significantly shorter response times than the verb-only representations. For the untrained group, this result is well seen in Figure 5 as a "v" shape for most functions, demonstrating that representations 1 and 3 had equivalent response times. Functions A and D, however, show an "r"-shaped pattern since representations 2 and 3 had similar response times.

Based on these results, we hypothesize that there are three important characteristics that influence performance: dominance, uniqueness and ambiguity (Figure 7).

Dominance is a characteristic intrinsic to each function and its context (the group of functions being represented). An element (verb or object) of the function is said to be dominant if the other element can be inferred from the first one. In other words, an element is dominant if the other element(s) is redundant in the description of the function. For example, in function A, given the verb "to sweep" most users would infer that the whole function is "to sweep the floor" in the given context, making the function verb-dominant. Similarly, in function B, given the object "cards" it is natural to assume the function "to play cards", making this function object-dominant. Dominance plays an important role in the performance of untrained participants when using single-element icons. The two functions A ("to sweep the floor") and D ("to comb the hair"), where representation 2 (verb-only) had shorter response time than representation 1, are clearly verb-dominant. Other functions where representation 1 (object-only) showed shorter response times are mainly object-dominant in the given context: e.g. "to play cards", "to eat a hamburger", "to ride a bicycle", "to play piano". Therefore, we may conclude that, for untrained users, single-element representations depicting the dominant element will result in shorter response times than single-element representations with the non-dominant (redundant) element.

We also hypothesize that the second factor affecting performance is *uniqueness* (i.e. design uniqueness). Uniqueness is a characteristic of the representation. If the representation and thus the function have perceptual immediacy, the representation can be said to be unique. Also, partial uniqueness may be defined as a case where one or more elements (but not all elements) of the function are unique. In Table 1, icons 2, 4, 5, 6, among others are unique, that is, most users, given the context, understand the representation as what the designer meant to represent: a broom, a set of playing cards, the Chinese word 玩 ("play") and a shoe. In contrast, icons 10, 14, 16, 20, 22 are not unique, as most untrained users would not immediately understand what they represent. In other words, each of these icons can mean different things even in the given context and as a result the function that each represents may not be clear. Uniqueness is a necessary characteristic if dominance is to influence performance. It should be noted that partial uniqueness of the required element is sufficient for dominance to play a role. An icon representing an intrinsically dominant element (verb or object) designed in a way that is not immediately understood by the user will tend to have poor performance, e.g. function H. Given the context of the experiment, even though most users could probably infer "watch TV" from the object "TV" (making the function object-dominant), the unclear (not unique) representation of TV led to the poor performance of Icon 22 (function H, representation 1) for the untrained group. In this case, the TV was intentionally designed to look like a box or a window to induce some confusion, and mimic the case of a poor O–R relationship.

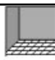
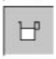
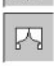
On the other hand, uniqueness might not guarantee good performance. Icons 5 and 32 depict the Chinese words 玩 ("to play") and 饮 ("to drink"), respectively, and are unique to fluent readers of Chinese. However, the response times in the use of these icons were significantly longer than the ones depicting the objects of the same functions. As the functions "to play cards" and "to drink coffee" may be considered object-dominant, the partial uniqueness of the redundant elements (the verbs) did not lead to good performance. Therefore, only concomitant uniqueness and dominance lead to shorter response times.

The third characteristic affecting performance appears to be *ambiguity*. As Horton (1994) has indicated, if multiple meanings are possible in a single context by one user, the icon is ambiguous. This can be presented as,

$$\mathbf{Icon}_i + \mathbf{context}_j + \mathbf{viewer\ (or\ interpretant)}_k \;\Rightarrow\; \mathbf{meaning}_1 + \mathbf{meaning}_2 + \cdots + \mathbf{meaning}_n.$$

In this case, the user would not be able to figure out the meaning of the icon. In other words, a representation is said to be unambiguous (i.e. NOT ambiguous) if it is associated with one and only one of the functions in the context. Icons 2, 4, 11 are examples of unambiguous icons. In this context, they can be associated with the

TABLE 5
Errors at the first/second attempt and their classification

Function/ represent.	Icon	# of errors first attempt	# of errors second attempt	Selected icon	Error classification
A/1	1 	1	0	7 	Object–object ambiguity
C/2	8 	1	0	17 洗	Verb–verb ambiguity
F/1	16 	1	0	31 	Object–object ambiguity
F/2	17 洗	1	0	20 	Verb–verb ambiguity
F/3	18 	1	0	33 	Object–object ambiguity
G/2	20 	1	0	17 洗	Verb–verb ambiguity
I/2	26 踏	1	0	5 玩	Verb–verb ambiguity†
J/1	28 	1	0	22 	Object–object ambiguity
J/2	29 拉	1	0	8 	Verb–object ambiguity
K/2	32 饮	2	3	14 	Verb–verb ambiguity
K/2	32 饮	5	1	20 	Verb–object ambiguity
L/2	35 弹	2	0	5 玩	Verb–verb ambiguity†
L/2	35 弹	1	0	8 	Verb–object ambiguity
Overall		19	4		

† Verb–verb ambiguities due to the influence of the original object.

functions they were designed to represent. Icon 14, on the other hand, is an ambiguous representation, as it may be associated with functions E ("to eat a hamburger") and K ("to drink coffee"). Ambiguity will not only lead to longer response times, but also to accuracy problems. In the past (example, Rogers, 1989), uniqueness and ambiguity have not been differentiated. But, in terms of modelling icon performance this differentiation can be very important even though an interface can account for this semantic-confusion by making some icons non-selectable given a specific context as outlined by Gittins (1986).

Analysis of the incorrect selections shows ambiguity as the sole source of errors. Table 5 shows the source of the mistakes and their classifications. Twelve (63.2%) of the incorrect selections at the first attempt were due to ambiguity within the represented element, that is, the represented verb (eight cases) or object (four cases) of another function was mistaken by the intended verb or object. Seventy-five per cent of the mistakes at the second attempt were due to verb–verb ambiguity. For example, there was verb–verb ambiguity between icons 17 (verb "to wash") and 20 (graphical representation of "to water"). There was object–object ambiguity between icons 16 (graphical representation of dishes) and 31 (graphical representation of a cup of coffee). Three of the eight cases of verb–verb ambiguity were due to the influence of the object (cases marked with '†' in Table 5). The Chinese verbs for play in "to play cards" 玩 and "to play piano" 弹 are very different, with very distinct, unique, meanings for any fluent reader of Chinese language. However, "piano" could be considered a "toy", in which case the word, 玩 (icon 5) ("to play", as in play with toys) could be linked to the object piano, thus making Icon 5 ambiguous. Two participants made this error. Similarly, one participant selected Icon 5, "to play" 玩, for the function "to ride a bicycle", probably also linking the ideas of bicycle and toy. Since the participants were all fluent readers of Chinese who saw the displays in both English and Chinese, this mistake can only be attributed to the natural dominance of the object in functions I and L. Even though the verbs depicted in the correct icons were also present in the description of the functions, once the participants understood the function, they did not go back to the description—they selected an icon that could be associated with the function as fast as possible. In these cases, the selected icon was a different verb that could be associated with the objects of the original functions.

Seven (36.8%) incorrect selections in the first attempt and one (25%) in the second attempt were due to verb–object ambiguity. In those cases, even though the icons shown to the participants were from representation 2 (*verb* only), the participants tried to find graphical representations that were similar to the *object* of the function. Six participants (five at the first attempt and one at the second attempt) selected the jug of water (Icon 20) for the function "to drink coffee", instead of selecting the Chinese verb 饮 (Icon 32), meaning "to drink". In the other two cases, Icon 8, graphical representation of "brush", was mistaken by either a curtain instead of Icon 29 depicting Chinese word 拉 for "to draw", or a piano, instead of Icon 35 that had the Chinese word 弹 for "to play". These mistakes confirm the object dominance in functions J, K and L, when the participants were actively looking for the object of the functions while the verbs were clearly and uniquely represented in the available icons for selection.

Therefore, a combination of dominance, uniqueness and lack of ambiguity can lead to single-element icons with very good performance.

Training seems to alter the balance of dominance, uniqueness and ambiguity. With training, the participants understand the design rationale, and can identify even the unclear, non-unique representations. As the training was performed at an elemental level, it helped the subjects understand the logic of the icon's construction. In other words, it helps bridge the gap between the user and designer mental model (Norman, 1988). Therefore, the R–I relationship is enhanced, and intrinsically "bad" designs, with poor O–R relationships (e.g. Icons 1 and 8), can have dramatic performance improvements with training. This fact seems to diminish the importance of dominance for the single-element icons. For the trained group only functions B, E, G and H had significantly different response times for the two single-element representations (see Table 4), as opposed to 10 out of 12 functions where dominance plays a role for the untrained group.

Function H, however, is a remarkable example that dominance is still important for the trained participants. Icon 22 is a poor representation of a TV, but well understood by the trained users as a TV, thereby making it unique. Since the function is intrinsically object-dominant, icon 22 has significantly better performance than icon 23, the unique representation of the redundant element ("to watch") for the trained participants. Most software today have "tool-tips" that pop-up the function of an icon when the mouse hovers over an icon. This technique helps improve accuracy but does not help in terms of performance time. Table 3 clearly shows the significant improvement in the performance time when the subjects received a short training. Software designers may feel that users should learn the meaning of icons by checking and seeing the results. This form of trial-and-error learning may not clearly portray the design concept as opposed to conceptual learning. Even conceptual learning can be problematic for abstract icons if ambiguity is high and uniqueness is weak.

An interesting observation about the influence of training is that there seems to be little room for performance improvements in very good icons (i.e. icons that depict the dominant element and are already unique and unambiguous for the untrained participants). Examples of such icons are Icons 2, 11 and 34, which showed little or no difference between the untrained and trained groups (Figure 5).

Even though numerous methods have been developed for icon evaluations (Hakiel & Easterby, 1987), designing icons right the first time and predicting their performance has been difficult. We have attempted to assimilate the characteristics that affect icon performance so that a quantitative model may be developed in the future.

9. Conclusions

In the past, icon design and development was more artistic by nature. The science of icons is only in selecting the appropriate icons through usability testing protocols such as ISO (1989) and ITU. The results of this experiment show that there are important considerations in addition to relevancy and simplicity. Superior performance may be achieved through the design of truly *unified* representations, depicting all elements of the function being represented. When a single-element representation is under consideration, three important aspects to consider are ambiguity, uniqueness and dominance. Even though this study was performed with a set of concerete icons, the concepts are applicable when designing any icon. Understanding ambiguity, uniqueness and dominance may improve the design and the corresponding representations for icons. In addition, to facilitate the interpretation of icons by new or occasional users, a simple and short training making the design rationale of the icons used in the interface visible may result in a considerable improvement in performance.

The R–O–I loop is closed when a "culturally biased-user" integrates the representation and expression and tries to generate specific or at least general knowledge about the object. The need for multi-cultural compatibility has been known for quite sometime (Nielson, 1990; Fernandes, 1995). Even so, icons tend to be localized as a result of their artistic nature and their geographical origin; possibly due to a lack of governing principles for icon design. Caution needs to be exercised when designing icons to take advantage of dominance as culture can dictate dominance or lack of dominance and hence the designs ought to compensate for such features.

This work has been supported by the Research Grants Council of the Hong Kong Special Administrative Region, China (Project No. HKUST800/96E). The authors would like to thank Tam Man Sin and Tang Wai Man for their contributions.

References

ANDERSON, J. (1990). *Cognitive Psychology and its Implications*. New York: W. H. Freeman Publishers.

APPLE COMPUTER, INC. (1996). *Newton 2.0 User Interface Guidelines*. Reading, MA: Addison-Wesley.

BARNARD, P. & MARCEL, A. (1984). Representation and understanding in the use of symbols and pictograms. In R. EASTERBY & H. ZWAGA, Eds. *Information Design*, pp. 37–75. New York: Wiley.

BLANKENBERGER, S. & HAHN, K. (1991). Effects of icon design on human–computer interaction. *International Journal of Man–Machine Studies*, **35,** 363–377.

BOURGES-WALDEGG, P. & SCRIVENER, S. A. R. (1998). Meaning, the central issue in cross-cultural HCI design. *Interacting with Computers*, **9,** 287–309.

CARROL, J. M. & THOMAS, J. C. (1982). Metaphor and the cognitive representation of computer systems. *IEEE Transactions on Systems, Man and Cybernetics*, SMC-**12,** 107–116.

CHOONG, Y-Y. & SALVENDY, G. (1998). Design of icons for use by Chinese in mainland China. *Interacting with Computers*, **9,** 417–430.

EASTERBY, R. S. & ZWAGA, H., Eds. (1980). Information design: the design and evaluation of signs and printed material. *Proceedings of a NATO Conference on Visual Presentation of Information*, Het Vennenbos, Netherlands, 1978. Sussex: Wiley.

EGIDO, C. & PATTERSON, J. (1988). Pictures and category labels as navigational aids for catalog browsing. *Proceedings of ACM CHI '88 Conference on Human Factors in Computing Systems*, pp. 127–132. Washington, D.C.

FERNANDES, T. (1995). *Global Interface Design*. London: Academic Press.

GALITZ, W. O. (1996). *The Essential Guide to User Interface Design: An Introduction to GUI Design Principles and Techniques*. New York: John Wiley.

GARG, C. & PLOCHER, T. (1999). The cross-cultural user interface initiative at Honeywell. *Proceedings of the 1st International Workshop on Internationalization of Products and Systems*. Rochester, NY. May 20–22, pp. 59–67.

GITTINS, D. (1986). Icon-based human-computer interaction. *International Journal of Man–Machine Studies*, **24,** 519–543.

GOONETILLEKE, R.S., SHIH, H. M. & KURNIAWAN, S. H. (2001). Appropriateness of icons. *Working paper*. Department of Industrial Engineering and Engineering Management, Hong Kong University of Science and Technology.

HAKIEL, S. R. & EASTERBY, R. S. (1987). Methods for the design and evaluation of icons for human–computer interfaces. In *Proceedings of the International Conference on Command, Control, Communications and Management Information Systems* (Conf. Publ. No. 275). IEE. London, U.K. pp. 48–53.

HEMENWAY, K. (1982). Psychological issues in the use of icons in command menus. In *Proceedings of the ACM conference on Human Factors in Computer Systems*. Gaithersburg, MD, pp. 20–24.

HORTON, W. (1994). *The Icon Book*. New York: John Wiley.

ISO (International Organization for Standardization) (1989). *ISO DIS 9186: Procedure for the Development and Test of Public Information Symbols*. Geneva: ISO.

JERVELL, H. R. & OLSEN, K. A. (1985). Icons in man–machine communications. *Behaviour and Information Technology*, **4,** 249–254.

KACMAR, C. J. & CAREY, J. M. (1991). Assessing the usability of icons in user interfaces. *Behaviour and Information Technology*, **10,** 443–457.

KNAPP, B. (1985). The precedence of global features in the perception of map symbols. *Proceedings of the Human Factors Society, 29th Annual Meeting*, Baltimore, pp. 287–291.

LANDSDALE, M. W., SIMPSON, M. & STROUD, T. R. M. (1990). A comparison of words and icons as external memory aids in an information retrieval task. *Behaviour and Information Technology*, **9,** 111–131.

LODDING, K. N. (1983). Iconic interfacing. *IEEE Computer Graphics and Applications*, **3,** 11–20.

MULLET, K. & SANO, D. (1995). *Designing Visual Interfaces: Communication Oriented Techniques*. Englewood Cliffs, NJ: SunSoft Press.

NADIN, M. (1988). Interfaces design and evaluation—semiotic implications. In H. R. HARTSON & D. HIX, Eds. *Advances in Human–Computer Interaction*, Vol. 2, pp. 45–100. Norwood, NJ: Ablex Publishing.

NIELSEN, J., Ed. (1990). *Designing User Interfaces for International Use*. Amsterdam: Elsevier.

NORMAN, D. A. (1988). *The Psychology of Everyday Things*. New York: Basic Books.

PAIVIO, A. (1971). *Imagery and Verbal Processes*. New York: Holt, Reinhart and Winston.

PEIRCE, C. S. (1932). In C. HARTSHORNE & P. WEISS, Eds. *The Collected Papers of Charles Sanders Peirce*, Vol. 2. Cambridge, MA: Harvard University Press. (Excerpted in J. Buchler, ed. *Philosophical Writings of Peirce*. New York: Dover, 1955.)

ROGERS, Y. (1989). Icons at the interface: their usefulness. *Interacting with Computers*, **1,** 105–117.

SHNEIDERMAN, B. (1987). *Designing the User Interface*. Reading, MA: Addison-Wesley.

SMITH, D. C., IRBY, C., KIMBALL, R. & VERPLANK, B. (1982). Designing the STAR interface. *Byte*, **7,** 242–282.

SOMBERG, B. (1987). A comparison of rule-based and positionally constant arrangements of computer menu items. In J. M. CARROLL & P. P. TANNER, Eds. *Human Factors in Computing Systems—IV and Graphics Interface*. Amsterdam: North-Holland, pp. 255–260.

WATERWORTH, J. A., CHIGNELL, M. H. & ZHAI, S. M. (1993). From icons to interface models: designing hypermedia from the bottom-up. *International Journal of Man–Machine Studies*, **39,** 453–472.

Paper accepted for publication by Associate Editor, Dr K. Eklundh