JOURNAL OF

Educational Computing Research

Editor: Dr. Robert H. Seidman

Volume 14, Number 1—1996

Relative Perceptibility of HyperCard Buttons Using Pictorial Symbols and Text Labels

Kira S. King, Elizabeth Boling, Janet Annelli, Marty Bray, Dulce Cardenas, Theodore Frick

Baywood Publishing Company, Inc.
26 Austin Ave., Amityville, NY 11701
call (516) 691-1270 **fax** (516) 691-1770 **orderline** (800) 638-7819
e-mail: baywood@baywood.com **web site:** http://baywood.com

J. EDUCATIONAL COMPUTING RESEARCH, Vol. 14(1) 67-81, 1996

RELATIVE PERCEPTIBILITY OF HYPERCARD BUTTONS USING PICTORIAL SYMBOLS AND TEXT LABELS

KIRA S. KING, ELIZABETH BOLING,
JANET ANNELLI, MARTY BRAY,
DULCE CARDENAS, THEODORE FRICK
Indiana University

ABSTRACT

The purpose of this experimental study was to investigate the comparative perceptibility of hypertext navigation buttons in three configurations: buttons with both pictorial symbols and text labels, with text labels only, and with pictorial symbols only. An instructional HyperCard stack was created in three versions, each differing only in the type of buttons used. Subjects were given typical situated tasks which required them to interpret navigational functions of various buttons. Findings indicated that buttons with both pictorial symbols and text labels resulted in significantly less user confusion than did buttons with pictorial symbols only. Buttons with text labels only also produced significantly less confusion, compared to those with pictorial symbols only. These findings have practical implications for hypertext designers. Many extant stacks typically use buttons with pictorial symbols only, which may create user confusion during stack navigation.

NAVIGATION BUTTONS AND HYPERCARD™

In order to use instructional software created with HyperCard (stacks) effectively, learners must be able to navigate, or find their way through, the software. McKnight, Dillon, and Richardson identify "landmark knowledge," or use of conspicuous and relatively stable features of an environment, as the type of knowledge most people are observed to use in finding their way around hypertexts in a process analogous to physical wayfinding [1]. Since HyperCard stacks do not have tangible presence in the sense that learners may literally walk around in them, visual interface elements must function as "physical movers." These are mechanisms by which users act on their landmark knowledge to locate information, to move through a sequential presentation, to connect to a related segment of content, or to go back for another look at something.

The primary visual interface elements HyperCard employs for navigation are buttons (a subset of computer icons). These buttons are rectangular "hot spots" on the screen that can be programmed to carry out different actions when a user clicks on them. Creators of stacks, including the non-professional designers that HyperCard is intended to support [2], need to provide navigation buttons that convey their functions easily and accurately to users. While HyperCard buttons can be presented in many ways, there are three major classes of buttons available. This authoring system provides options for creating rectangular buttons that contain a text label only, a pictorial symbol only (creators choose from a supplied library), or both a pictorial symbol and a label. Stack creators must decide, consciously or not, which of these options—pictorial symbol, text label, or both—will result in the most effective navigation buttons for their users. We state our research question as follows: do people with some HyperCard experience make fewer errors interpreting the function of navigation buttons when those buttons contain pictorial symbols only, pictorial symbols and text labels, or text labels only (see Figure 1)?

SINGLE MODALITY ICONS AND MIXED MODALITY ICONS

Pictorial Icons versus Verbal (Text) Icons

Several studies have been conducted to determine the comparative effectiveness of single modality icons—those comprised of either a pictorial symbol or a text label [3-5]. Results strongly suggest that single modality pictorial icons are not de facto superior to single modality text icons. Neither Rohr and Keppel [5] nor Benbasat and Todd [3] found a significant difference in time to complete tasks

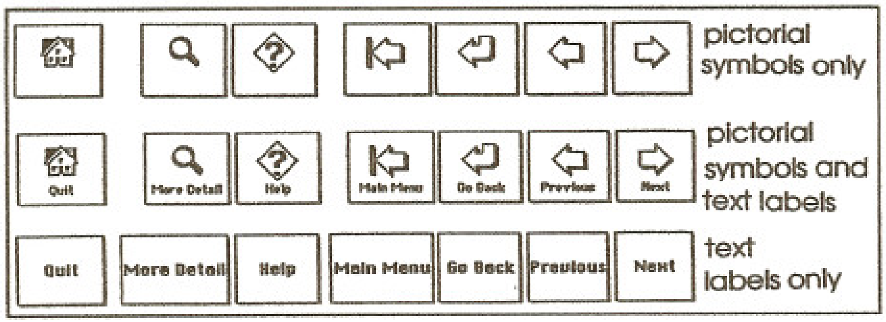

Figure 1. HyperCard buttons showing three standard configurations.

or in the number of faults made by subjects using software which employed only pictorial icons or text-only icons. Lindgaard, Chessari, and Ihsen studied icons which represented complex telephone functions and found that subjects identified functions more accurately from text than from pictorial symbols [4]. However, when subjects were provided with the rule-set governing the design of the pictorial symbols while they were making their identifications, they could identify functions as accurately from pictorial symbols as from text.

Mixed Modality Icons versus Single Modality Icons

Guastello, Traut, and Korienek summarize several concepts suggesting the possible advantages of mixed modality icons (those containing both pictorial symbols and text labels) over single modality icons [6]. The first concept, derived from the parallel processing theory of cognition, holds that the brain processes stimuli of mixed verbal and spatial content in both hemispheres simultaneously, and that mixed modality stimuli give individuals the chance to use their dominant hemisphere to process that stimuli. The second concept, redundancy, states that in human information processing redundant information perceived through two channels simultaneously is processed faster and more reliably than information perceived through a single channel only. Researchers also speculate that either labels serve to clarify the meaning of pictorial symbols [7-9] or that pictorial symbols help specify the meaning of text labels [10].

Mixed modality icons have been tested against both types of single modality icons, with inconsistent overall results. Brems and Whitten [11], Edigo and Patterson [10], and Guastello, Traut, and Korienek [6], found that mixed modality icons were superior to single modality icons as measured, respectively, by ease of learning and preference, time to complete tasks and selection errors, and meaningfulness. Bewley, Roberts, Schroit, and Verplank measured preference, accurate identification and recognition time for icons from several sets [7]. They concluded that mixed modality icons were preferred by subjects while they were learning the icons, but that thereafter, there was no significant difference between mixed modality and single modality pictorial icons. Kacmar and Carey found optimal performance in subjects matching verbal descriptions of functions to one of fifteen icons for mixed modality and single modality text icons, but not for single modality pictorial icons [12]. In a study similar to the icon studies, Chambers, Alexander, Howard, Andrew, O'Boyle, Eastman, and Motoyama compared mixed modality and single modality symbols for communicating photocopier functions [13]. They concluded that text labels were sufficient for identification of familiar functions, and that mixed modalilty symbols helped communicate understanding of unfamiliar functions. The findings from these studies are summarized in Table 1.

Table 1. Summary of Measures and Findings Comparing Mixed Modality and Single Modality Icons

Study	Measures	Preferred Modality
Brems and Whitten, 1987 [11]	Ease of learning and preference	Mixed modality
Guastello, Traut, and Korienek, 1989 [6]	Meaningfulness	Mixed modality
Edigo and Patterson, 1988 [10]	Time to complete tasks and selection errors	Mixed modality
Bewley, Roberts, Schroit, and Verplank, 1983 [7]	Accurate identification, preference, and recognition time	Mixed modality preferred for learning; thereafter no preference over single modality pictorial
Kacmar and Carey, 1991 [12]	Accurate identification	Mixed modality and single modality text
Chambers et al., 1992 [13]	Accurate identification	Single modality text for familiar functions; mixed modality for unfamiliar functions

RELATIONSHIP OF SYMBOL STRUCTURE AND METHODS USED TO STUDY ICONS

Interdependent Components of Symbols

Easterby provides a useful and much-cited description of symbols as having five basic components: 1) pragmatic, the context in which the symbol will be used; 2) semantic, the symbol's relationship to its referent; 3) syntactic, the symbol's relationship to others in the set, 4) visibility, how well the symbol can be seen; and 5) discriminability, how well the symbol is differentiated from others [14]. Each of the five elements affects the other four and contributes to the symbol's overall perceptibility. No one component alone determines a symbol's effectiveness.

Easterby further suggests that determining a symbol's perceptibility can only be done when all five of its components are intact. If one component is missing from the symbol during testing, results may not be adequate to ensure that the symbol actually is perceptible. Rogers [15] supports Easterby's view with respect to a symbol's pragmatic component (context), and syntactic component (relation to the entire symbol set of which it is a part).

Methods Used to Study Icons

Of the studies we summarize in Table 1, only Edigo and Patterson [10] used actual tasks to test icons. The rest of those studies—along with several others we reviewed which were not looking specifically at mixed modality versus single modality icons—followed a similar methodology. In the Edigo and Patterson study, subjects were shown a single icon or a symbol projected on an overhead projector or printed on an index card. After viewing the symbol for a short time (25-30 seconds), or in some cases for as long as desired, the subjects were required to demonstrate how well they understood the symbol's meaning. In addition, methods for subject response varied, but in most studies the subjects were asked to write or state the meaning of the symbol in their own words, or to choose among a given set of answers [16-19]. While several of the more recent studies we reviewed examined icons within their icon sets [12, 20-22], they still presented the icons out of context.

None of the studies we reviewed used the ISO (International Organization for Standardization) procedure for testing public information symbols which Easterby was instrumental in developing [23, 24]. The procedure has been criticized, interestingly enough, for not including contextual tests, measures for ease of learning, or measures for behavioral effectiveness [25].

Design of the Study

We wanted to determine whether people with some HyperCard experience make fewer errors interpreting the function of navigation buttons when those buttons contain pictorial symbols only, pictorial symbols and text labels, or text labels only. We decided to study navigation buttons in the context of a HyperCard stack and in the context of realistic user actions, rather than testing individual icons. We expected to find results compatible with the studies previously cited which found that buttons with pictorial symbols and text labels (mixed modality icons) would be more perceptible than either of the other two groups: buttons with pictorial symbols alone or buttons with text labels alone (single modality icons). Since we were using buttons from HyperCard's supplied library, we also expected that subjects who had prior experience with HyperCard would already be familiar with the pictorial symbols we were using.

METHODOLOGY

We studied three types of HyperCard buttons in the context of an actual stack: buttons with pictorial symbols only, buttons with pictorial symbols and text labels, and buttons with labels only. The subjects for the study were students attending classes in the School of Education at Indiana University (IU). All of the subjects had some experience using HyperCard.

Preliminary Survey

A preliminary questionnaire was created and administered to twelve graduate students in the department of Instructional Systems Technology. Subjects were asked to identify, out of context, the meanings of seven basic pictorial symbols which are available within HyperCard for use on buttons. We conducted this preliminary survey to discover whether these symbols were either so perceptible, or so commonly used and therefore recognized, that they were not suitable for more formal study. Our results indicated that further research was warranted.

Study Design

We developed a model stack to ensure that our subjects would be exposed to a HyperCard stack they had never seen, and would, therefore, be attempting to perceive the functions of buttons without prior learning in that context. After creating the six-card stack, we considered two alternative approaches to measurement: on-line testing vs. a paper-based survey. In the first approach, subjects are presented with a fully functioning HyperCard stack and a list of tasks to complete. A camera or a software program is used to record every key stroke the subjects make while proceeding through the test. In the paper-based survey, each card of the stack is printed out onto a sheet of paper, and the subject answers questions regarding the buttons on each card.

While the on-line test is the optimal approach with regard to replicating real-world conditions, it does have several critical drawbacks. First, while using the stack, the subjects may learn the functions of buttons before we have the opportunity to test for perceptibility. That is, if users chose an incorrect button for one task, they would learn the correct function of the button which they had incorrectly selected. This prior knowledge would then be used as they encountered that button in subsequent tasks, and would confound our results. Second, if the subjects became lost in the stack, their frustration may negatively affect their performance on the test, and may limit their progress through the on-line test. Therefore, we concluded that the risks associated with an on-line test outweighed the benefits, and chose to create a paper-based survey.

We did not impose a time limit on the tasks. Subjects were allowed to spend as much time as desired viewing the HyperCard printouts, and answering the questions. As discussed previously, this decision runs counter to many of the studies we reviewed, most of which allowed subjects to view the symbols for only twenty-five to thirty seconds. These studies, however, were measuring other components besides perceptibility, as can be seen in Table 1. Two studies [7, 10] were specifically measuring speed of recognition. Since we were attempting to measure perceptibility and not time, we saw no need to impose a time limitation. One possible effect of imposing a time limit on subjects would have been to reduce the time they had available to attend to the contextual clues we expected them to use in making their choices. Furthermore, we felt that allowing as much time as subjects wanted was congruent with real-world conditions in that users probably do want to make their selections quickly, but they are not ordinarily constrained to making those selections in twenty-five to thirty seconds.

The survey packets contained two sections: the background information questionnaire, used to identify the subjects' native language and relative experience using HyperCard, and the test section. The test section of the packet consisted of a series of screens printed out from the model HyperCard stack. Each of the screens included seven buttons representing commonly-used functions. The functions were: Quit, More Detail, Help, Main Menu, Go Back, Previous, and Next. Each screen printout was presented on a separate page. At the bottom of each page was a brief statement of a realistic circumstance the subjects might find themselves in, a specific task to be completed, and the instruction to circle the part of the screen that would help perform that task. For example, "You've opened this stack and discovered it isn't the one you want. Circle the part of the screen that you would click in order to leave the stack" (see Figure 2). Buttons that would not be functioning on a particular screen according to the stack design were represented in gray on the printout.

Three versions of the survey packet were used: Group A packets showed buttons with pictorial symbols only, Group B packets showed buttons with pictorial symbols and text labels, and Group C showed buttons with text labels only.

Subjects

Subject volunteers were recruited from the following sources: undergraduate students in teacher education, graduate students enrolled in a writing course, and graduate students attending a weekly lecture series in Instructional Systems Technology. Since several students were present in both the writing class and the weekly lecture, care was taken to ensure that each volunteer participated in the study only once.

Survey Procedure

The packets were collated into stacks containing equal numbers of Group A, Group B, and Group C packets; additionally, within each stack, the packets were

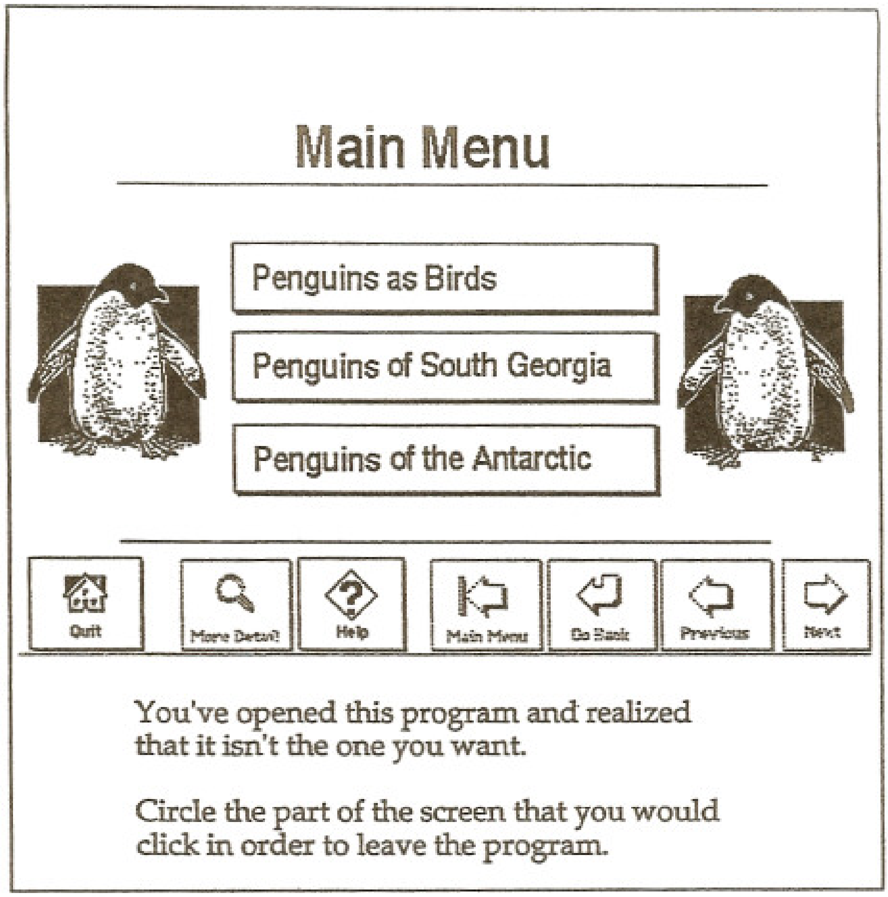

Figure 2. Sample page from one test packet. The HyperCard screen is printed at the top of the page and the task description is printed below.

sequenced so that the first packet came from Group A, the second from Group B, and the third from Group C. The facilitator then randomly distributed the packets to the volunteers. The subjects were not informed that there was any difference between the packets.

Before the subjects opened the packets, the facilitator gave them brief scripted instructions on how to fill out the survey and requested that they complete their packets individually. Since subjects with different packets sat close to each other, we hoped this last instruction would prevent them from collaborating. As discussed previously, subjects were given as much time as they needed to complete their packets; however, from our casual observation they typically spent about five to ten minutes.

Measures

Facilitators collected a total of forty-one packets. Two of those packets were randomly discarded in order to keep the number of Group A, Group B, and Group C packets equal.

To score each packet, one facilitator read the correct answers for each sheet while a second facilitator marked the subject's response. A third facilitator simultaneously checked the marking. The responses were tallied and double-checked for each packet, then totaled for each of the three groups.

Results

Mean number of errors in identifying navigation buttons in each of the three treatment groups are reported in Table 2. A one-way ANOVA resulted in a significant F ratio of 12.72 ($df = 2{,}36$, $p < .0001$) (see Table 3). Tukey's HSD procedure indicated significant differences ($p < .05$) between Groups A and B and between Groups A and C. As can be seen in Table 2, the group which used the buttons with pictorial symbols only (mean = 8.23) made about *twice* as many errors in identifying navigation buttons as did the group that used buttons with text labels only (mean = 3.77). The group which used buttons with pictorial symbols only (mean = 8.23) made about *three times* as many errors as did the group which used buttons with both pictorial symbols and text labels (mean = 2.62). The difference between means for Groups B and C was not significant at the 0.05 level. The strength of the effect of the treatment was 0.375 using omega-squared [26, p. 162]. Over one-third of the variance in errors in identifying navigation buttons can be attributed to the form of representation of navigational functions.

As to whether prior experience with HyperCard stacks was related to frequency of errors, no significant relationship was found. Prior experience was grouped into two categories based on questionnaire responses: those who had little or no HyperCard experience vs. those with some or much experience. The ANOVA resulted in an $F = 1.09$, $df = 1{,}37$, $p > .303$.

Table 2. Mean Number of Errors for Group A, Group B, and Group C

Group[a]	Mean	Standard Deviation
Group A—pictorial symbols only	8.2308	2.7127
Group B—pictorial symbols and text labels	2.6154	2.7549
Group C—text labels only	3.7692	3.4678

[a] $n = 13$ for each group.

Table 3. ANOVA Source Table for Navigation Errors in the Three Treatment Groups

Source	D.F.	Sum of Squares	Mean Squares	F Ratio	F Probability
Between groups	2	228.6667	114.333	12.7158	.0001
Within groups	36	323.6923	8.9915		
Total	38	552.3590			

Almost a third of our subjects were non-native English speakers. An ANOVA resulted in no significant difference ($F = 0.59$, $df = 1{,}37$, $p > .449$) in native language (native vs. non-native speaker of English) relative to frequency of errors.

When looking at the number of errors made with specific navigation buttons, we noticed that two buttons, "Go Back" and "Main Menu," accounted for 74 percent of the total errors in the study.

With regard to non-normality of distribution, our data set is positively skewed. However, research indicates that ANOVA remains robust when the normality assumption is violated *if the sample or cell sizes are equal*—as they are in our study [26]. Furthermore, as can be seen from examining the standard deviation in Table 2, the assumption of homogeneity of variance appears to have been met. Even if it had not been met, ANOVA remains robust when cell sizes are equal [26].

DISCUSSION

Mixed Modality Buttons and Single Modality Text Buttons More Perceptible than Single Modality Pictorial Buttons

The results of this study corroborate previously cited icon studies: buttons containing pictorial symbols and text labels are more perceptible than buttons containing pictorial symbols alone, as shown by the relative accuracy with which subjects interpret their functions.

Our findings suggest further, in contrast to our expectations and in contrast to most of the studies we reviewed, that buttons containing text labels are as perceptible as buttons containing pictorial symbols in addition to text labels. Chambers et al [13] report similar results to ours from their study of symbols representing photocopier functions, although they establish the strongest effect for the text-only condition, stating that mixed modality symbols were helpful for subjects unfamiliar with the copier functions. Similar results are also reported in the first of two studies conducted by Edigo and Patterson [10]. However, after they revised their text-labels-only interface to eliminate an extra pictorial aid available through a text button, they concluded that mixed modality icons resulted in better performance than either single modality condition.

Nearly identical results to ours are reported by Kacmar and Carey [12]. They compare their results to those of Edigo and Patterson [10] also, and conclude that the difference is due to subjects mapping a concept phrase to the symbol set from which they made their selection, instead of viewing a symbol and then matching it with a described goal. Although we presented subjects with complete HyperCard screens from which to choose the correct button, and since our subjects' task was analogous to that used by Kacmar and Carey [12], the same explanation may account for our results.

High Error Rates for "Go Back" and "Main Menu" Buttons

The comparatively high error rates seen on the "Go Back" and "Main Menu" buttons may be related in part to the similar, and perhaps ambiguous, pictorial symbols contained in those buttons for Groups B and C. The underlying functions represented by these two buttons are also similar (both move *backwards* in the stack to a point already visited by the user). "Main Menu" returns the user to a fixed and named location (the main menu) while "Go Back" returns to a variable, unnamed location (the beginning of the current section). These differences in functionality may not be apparent to the user, or may be ambiguous enough to cause confusion. We suspect both these issues are involved, and see them as potential areas for further study.

Native versus Non-Native Speakers of English

We were interested to note that there was no significant difference in the mean error rates between groups for native and non-native speakers of English. This suggests that reliance on pictorial symbols alone to create buttons that are perceptible across cultures may not be a useful strategy.

Key Areas for Future Research

As discussed previously, we chose to utilize a paper-based survey rather than an on-line test because we anticipated that our results would be confounded by subject learning and frustration. While we did avoid these two risks, our results may have been affected by the use of a paper-based survey. Future research is needed to determine if there is a difference in subject error rates for these two alternative measurement formats. It is interesting to note that a recent evaluation conducted by Frick, Corry, Hansen, and Maynes found little difference between the errors subjects made on paper and the errors made on a computer-based test

[27]. However, since this was not an experimental study, further research is needed.

Additionally, since our study only examined seven basic pictorial symbols from the HyperCard set, future research needs to test a wider variety of symbols. Even more critical, is the need to measure the perceptibility of single and mixed modality icons in learning products created in different authoring environments, including Authorware™, Toolbook™, and the World Wide Web.

Boling, Beriswill, Huang, Kaufman, Xaver, Frick, and Chuang recently found similar difficulty with the perceptibility of buttons representing backward navigation ("Main Menu" and "Go Back" buttons) [28]. Further research needs to be conducted in order to determine the optimal icon types for these tasks. We hypothesize that abstract pictorial symbols are not effective for backward navigation; rather, we must use unique features of the desired location. That is, in physical wayfinding we may be able to describe a place we have been and to which we would like to return, but we may be unable to recall exactly how many steps backward it will take to return to that place. A similar situation may be occurring in HyperCard. We know that we want to get back to a certain card and have a mental image of that card; yet, we have no idea how far back that is in the stack. Current pictorial symbols attempt to use an abstract depiction of the distance needed to go backwards in order to reach the intended location. A better strategy may be to create a miniature symbol that resembles the actual destination, i.e., the "Main Menu."

Finally, while we contend that evaluating icons within context is critical to determining their actual perceptibility, further research is required to determine the extent to which results obtained in a highly specific context like the one we have used will match, or fail to match, results obtained through more traditional means.

SUMMARY

In our view, the most important feature of this study is that we placed the navigation buttons being evaluated into the context of a HyperCard stack and provided our subjects with situated task descriptions to respond to. Seeing the buttons in context allowed the subjects to view them as a set, compare them, and discriminate between them for the one that most closely matched a function they presumed to be related to a task. We believe that the availability of information beyond the buttons themselves (for example, other text and titles on the screen and buttons appearing in a "grayed," or unavailable, state at appropriate times) makes the subjects' performance in this test realistic, and therefore increases the usefulness of our findings as a basis for informed design.

The primary implication of our results applies to those who design HyperCard stacks; if buttons including pictorial symbols only are used in a stack, even people who have experienced HyperCard may be confused about the functions of some of those buttons. To increase the perceptibility of HyperCard navigation buttons, use both pictorial symbols and text labels.

ACKNOWLEDGMENT

Technical assistance was provided to the research team by Dr. Richard Pugh, Indiana University.

REFERENCES

1. C. McKnight, A. Dillon, and J. Richardson (eds.), *Hypertext: A Psychological Perspective,* Ellis Horwood, New York, 1993.
2. Apple Computer, Inc., *HyperCard™ Stack Design Guidelines,* Addison-Wesley, Reading, Massachusetts, 1989.
3. I. Benbasat and P. Todd, An Experimental Investigation of Interface Design Alternatives: Icon vs. Text and Direct Manipulation vs. Menus, *International Journal of Man-Machine Studies, 38*:1, pp. 369-402, 1993.
4. G. Lindgaard, J. Chessari, and E. Ihsen, Icons in Telecommunications: What Makes Pictorial Information Comprehensible to the User? *Australian Telecommunication Research, 21*:2, pp. 17-29, 1987.
5. G. Rohr and E. Keppel, Iconic Interfaces: Where to Use and How to Construct? *Human Factors in Organizational Design and Management: Proceedings of First Symposium,* Elsevier Science Publishers, North-Holland, pp. 269-275, 1984.
6. S. Guastello, M. Traut, and G. Korienek, Verbal versus Pictorial Representations of Objects in a Human-Computer Interface, *International Journal of Man-Machine Studies, 31*:1, pp. 99-120, 1989.
7. W. Bewley, T. L. Roberts, D. Schroit, and W. L. Verplank, Human Factors Testing in the Design of Xerox's 8010 "Star" Office Workstation, *CHI '83 Conference Proceedings,* Association for Computing Machinery, New York, pp. 72-77, 1983.
8. D. Gittens, Icon-Based Human-Computer Interaction, *International Journal of Man-Machine Studies, 24,* pp. 519-543, 1986.
9. R. Krull, If Icon, Why Can't You? in *Effective Documentation: What We Have Learned from Research,* S. Doheny-Farina (ed.), MIT Press, Boston, Massachusetts, 1988.
10. C. Edigo and J. Patterson, Pictures and Category Labels as Navigational Aids for Catalog Browsing, *Human Factors in Computing Systems: CHI '88 Conference Proceedings,* Association for Computing Machinery, New York, pp. 127-132, 1988.
11. D. Brems and W. Whitten, Learning and Preference for Icon-Based Interface, *Proceedings of the Human Factors Society, 31st Annual Meeting,* The Society, Santa Monica, California, pp. 125-129, 1987.
12. C. Kacmar and J. Carey, Assessing the Usability of Icons in User Interfaces, *Behaviour & Information Technology, 10*:6, pp. 443-457, 1991.
13. S. Chambers, J. Alexander, A. T. Howard, M. O'Boyle, V. Eastman, and T. Motoyama, Symbol and Word Effectiveness for Conveying Photocopier Functions

in Experienced Photocopier Users, *Perceptual and Motor Skills, 74,* pp. 1203-1215, 1992.

14. R. S. Easterby, The Perception of Symbols for Machine Displays, *Ergonomics, 13*:1, pp. 149-158, 1970.
15. Y. Rogers, Icons at the Interface: Their Usefulness, *Interacting with Computers, 1*:1, pp. 105-117, 1989.
16. R. Geiselman, B. Landee, and F. Christen, Perceptual Discriminability as a Basis for Selecting Graphic Symbols, *Human Factors, 24*:3, pp. 329-337, 1982.
17. A. M. Mackie, *National Survey of the Knowledge of the New Traffic Signs* (Report No. 51), Road Research Laboratory, Harmondsworth, England, 1966.
18. A. Pejtersen and L. P. Goodstein, Beyond the Desktop Metaphor: Information Retrieval with an Icon-Based Interface, *Visualization in Human-Computer Interaction: 7th Interdisciplinary Workshop on Informatics and Psychology,* Springer-Verlag, Scharding, Austria, pp. 149-182, 1988.
19. Y. Rogers and D. J. Oborne, Pictorial Communication of Abstract Verbs in Relation to Human-Computer Interaction, *British Journal of Psychology, 78,* pp. 99-112, 1987.
20. R. B. Stammers and J. Hoffman, Transfer between Icon Sets and Ratings of Icon Concreteness and Appropriateness, *Proceedings of the Human Factors Society 35th Annual Meeting,* The Society, Santa Monica, California, pp. 354-358, 1991.
21. R. Magyar, Assessing Icon Appropriateness and Icon Discriminability with a Paired-Comparison Testing Procedure, *Proceedings of the Human Factors Society 34th Annual Meeting,* The Society, Santa Monica, California, pp. 1204-1208, 1990.
22. A. Salasoo, Towards Usable Icon Sets: A Case Study from Telecommunications Engineering, *Proceedings of the Human Factors Society 34th Annual Meeting,* The Society, Santa Monica, California, pp. 203-207, 1990.
23. R. S. Easterby and I. R. Graydon, *Evaluation of Public Information Symbols, ISO Tests: 1979/80 Series, Part I: Appropriateness Ranking Tests* (AP Report 99), Applied Psychology Department, University of Aston, Birmingham, England, 1981.
24. R. S. Easterby and I. R. Graydon, *Evaluation of Public Information Symbols, ISO Tests: 1979/80 Series, Part II: Comprehension/Recognition Tests* (AP Report 100), Applied Psychology Department, University of Aston, Birmingham, England, 1981.
25. C. Brugger, Advances in the International Standardization of Public Information Symbols, *Information Design Journal, 6*:1, pp. 79-88, 1990.
26. R. Kirk, Experimental Design (2nd Edition), Brooks/Cole, Belmont, California, pp. 75, 77, 1982.
27. T. Frick, M. Corry, L. Hansen, and B. Maynes, Design-Research for the Indiana University Bloomington World Wide Web Home Page, unpublished report, Indiana University, School of Education, URL address: http://education.indiana.edu/ist/faculty/iuwebrep.html, 1995.

28. E. Boling, J. Beriswill, J. Huang, D. Kaufman, R. Xaver, T. Frick, and W. Chuang, *Out-of-Context Testing of Relative Perceptibility of HyperCard Buttons using Pictorial Symbols and Text Labels,* unpublished manuscript, Indiana University, 1995.

Direct reprint requests to:

Dr. Elizabeth Boling
Education 2224
201 N. Rose Ave.
Instructional Systems Technology
Indiana University
Bloomington, IN 47405