Using Icons to Find Documents: Simplicity Is Critical

Michael D. Byrne

School of Psychology and
Graphics, Visualization, and Usability Center
Georgia Institute of Technology
Atlanta, GA 30332-0170
byrne@cc.gatech.edu

ABSTRACT

A common task at almost any computer interface is that of searching for documents, which GUIs typically represent with icons. Oddly, little research has been done on the processes underlying icon search. This paper outlines the factors involved in icon search and proposes a model of the process. An experiment was conducted which suggests that the proposed model is sound, and that the most important factor in searching for files is the type of icons used. In general, simple icons (those discriminable based on a few features) seem to help users, while complex icons are no better than simple rectangles.

KEYWORDS: screen design: icons, empirical evaluation, formal models of the user

INTRODUCTION

Despite the prevalence of icons in graphical user interfaces (GUIs), the basic processes underlying the interaction of humans and icons are not well understood. Icons are used in a wide variety of ways in GUIs, one of the more common of which is the representation of objects recognized by the operating system, such as files and directories. Searching for documents which are represented by icons will henceforth be referred to as "icon search." Presumably, icons are used because it is believed that they make it easier to locate such items.

Unfortunately, there is little empirical data that either supports or refutes this claim. By and large, the assumption that has driven the use of icons in the GUI is that icons seem to work better than having to recall and type file names, as is typical with a command-line interface. While this is likely to be the case, it not an adequate description of what processes the user may actually bring to bear, nor does it provide clear guidelines for the use of icons.

This research attempts to address many of the questions surrounding icon search.

TASK ANALYSIS

Icon search is a complex task involving a number of factors. Several factors have been identified, which are divided into three classes: general factors, visual search factors, and semantic search factors.

General factors

Mixed search. Icon search involves two kinds of search which may or may not be related to each other. This is because icons (or, at least, document icons) have two parts: the icon picture and a label, in this case a file name. Thus, search is both over the icon pictures (referred to as "visual search") and over the strings attached to the icons, termed "semantic search."

Task mapping. Icon search is a "variably-mapped" [10, 11] task; that is, targets and distractors switch roles between trials (the same file might be the target on one occasion and a distractor on another). Thus, icon search should be a relatively slow, controlled process that improves only minimally with practice [10, 11].

Target knowledge. Not only is there variable mapping between the target and the distractors, but knowledge about the target may be incomplete. Often, people do not know either the exact name of the file or what icon it will have before they begin to search for it.

Multiple matches and semantic necessity. Often in a search, the target document will share its icon picture with other documents present on the display. This makes it necessary to read the label attached to the document most or all of the time in order to compensate for the fact that the icon picture alone is ambiguous in that it may or may not represent the target document.

Motor task. In realistic measurement of icon search, there will also be a motor component—that of mouse movement to the target. This should be affected primarily by physical properties of the system, such as icon size (often a function of screen resolution), the distance to be moved, and the characteristics of the input device.

Visual factors

Size. Generally, all of the icons on a display are the same size. So while size may be a useful way to discriminate between icons, it is typically not employed in current displays.

Color. This is an interesting dimension, because many GUI displays do not support color. Thus, even if using color as a discriminating feature is effective, reliance on color differences may not be appropriate for icon designers.

Form. This is the primary dimension on which icons vary and includes a number of sub-dimensions, such as the level of detail in the form and the meaningfulness of the form.

Spatial organization. Some interfaces impose a grid-like organization on icons, others a "staggered grid" organization, and still others no organization at all—icons can be anywhere. This may or may not have an effect on search.

Number of objects. Both the size of the search set and the number of icons with pictures matching the target can vary widely from trial to trial. Probably the best-studied factor in visual search is that of set size, while the effect of multiple visual matches has been relatively ignored.

Semantic factors

Words/non-words/near-words. Since the labels on icons can typically be assigned almost any string up to a certain size, the labels may be made up of actual words, things that are not words at all (numbers, dates, etc.), and near-words (such as abbreviations). A given display may contain a mixture of all of these things.

Sorting. Often, the user has the option of sorting the display based on some features of the labels, the most common example being alphabetical sorting. Other forms of sorting are possible and available on some systems, such as sorting by time of modification or document type, which are also semantic variables.

While this description may not be entirely complete, it gives an indication of the number of issues that must be explored before one can fully understand what appears to be a fairly simple operation of pointing and clicking. Designing the "best" icons and icon-based GUI file system interface will require knowledge about all of these factors, and likely many of their interactions.

MECHANISMS

Research on Icons

By and large, the preceding factors have not been systematically investigated. Jones and Dumais [8] have done research demonstrating that it is difficult for people to find documents based on "spatial filing." In the spatial filing task studied, subjects were given documents to file according to whatever spatial scheme they wanted, and then later asked to retrieve the documents. Performance was generally poor, and clearly inferior to simple alphabetic filing. This begins to suggest that semantic information (names) are more important than visual information, but the support for this conclusion is not strong.

Byrne [2] suggests that pictures on icons do have an effect on people's ability to find documents: they slow people down. In this study, it was found that for a set size of twelve, simple empty rectangles work as well or better than icons with pictures. This work, however, was not very comprehensive and its claims are based on a relatively small sample with few trials. More critically, only one set size was examined, making it difficult to get a clear understanding of the processes involved.

Other work on icons (e.g. [1, 4, 6]) has addressed issues relating to the form of icons such as subjective user preference [6] and meaningfulness [1, 4], but generally says little about how that form will affect performance in a task such as icon search.

A Model of Mixed Search

Since the literature has largely not covered the icon search question, one approach is to develop a model of the process. There is some evidence (the verbal reports of subjects in [2], see also [4, 5]) that mixed search is actually a two-stage process: the user uses a visual search to narrow down the semantic search. For example, the user first finds all the icons on the display with pictures that match the application he or she was using and then reads the names only on those icons. Then, search time should be $f_m(n) = f_v(n) + f_s(t)$, where $f_m(n)$ is the total time for the mixed search. This is the simple sum of $f_v(n)$, the visual search time as a function of set size, and $f_s(t)$, the semantic search time as a function of the average number of icons with pictures matching the target icon (t).

A reasonable assumption to make about the semantic search is that it is a standard self-terminating serial search. Since there is no visual guidance for the search, one can reasonably expect that the user will read each file name and decide if it matches the target. If that is the case, search time should be a function of the average number of items necessary to search, which is n/2. That is, $f_s(n)$ for set size n, should be a linear function of the form $f_s(n) = m_s*(n/2) + b_s$. Additionally, if the user knows the exact file name, the decision process should be faster, and therefore the slope m_s shallower.

Times for $f_m(n)$ can be empirically estimated, and then it should be possible to estimate $f_v(n)$ by simply subtracting $f_s(t)$. Since t will be at most n (and usually less), the effectiveness of using the visual search is dependent on the nature of $f_v(n)$. If $f_v(n)$ is a linear function ($m_v n + b_v$) and its slope is large, then $f_m(n)$ can be expected to be a linear function such that $f_m(n) > f_s(n)$; that is, mixed search would always be slower than semantic search for any n. However, if $f_v(n)$ has a shallow or even zero slope, $f_m(n)$ should be more efficient than $f_s(n)$.

The visual characteristics of the icon pictures should determine $f_v(n)$ [3, 6, 12, 13]. If it is possible to discriminate icons based on one simple feature, it should be possible to conduct the visual search in parallel [12], and thus m_v should be close to zero. However, restriction to a single feature would severely limit the number of different icons possible, and most icons currently found in GUIs contain numerous features.

If it is possible to distinguish icons on the basis of a few features or a simple conjunction of features, m_v should be relatively shallow [6], perhaps even near zero [3, 13]. In this case, icons would act as a legitimate aid in searching for files. However, if icons largely contain all the same features or the same combinations of features, m_v should be steep [6, 3/13] and probably no advantage over m_s.

Empirical investigation

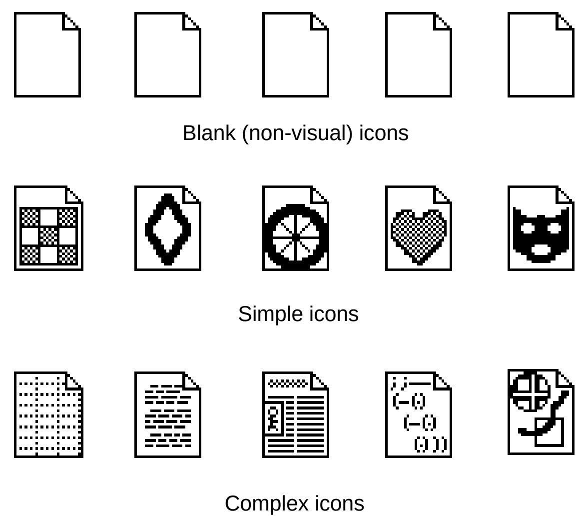
The experiment described in the remainder of this paper was intended to do two things: estimate the effects of several of the factors described in the task analysis and evaluate the mixed search model. Since the model predicts that set size should have an effect on the speed of icon search regardless of the type of search conducted, set size was an important variable in the study. Since the mixed search model includes a semantic search over the set of icons with pictures that match the picture on the target icon (the "match set," which includes the target icon) was also an important variable, primarily used in the estimation of the visual search function.

Since the nature of the visual search function should itself be a function of the visual characteristics of the icon pictures, this was also varied. Color and size were not varied, since color is not possible on all displays and size is typically constant. What was varied was the complexity of the visual form of the icons, using three types of icons (see Figure 1).

Another factor that is likely to vary widely in a naturalistic setting is the knowledge that the user has about the target before the search. There are two important classes of knowledge: file name knowledge and picture knowledge. Users may know the exact name of the file and they may not. Picture knowledge varies similarly; users may not know the picture they are searching for, they may have a guess as to the picture, or they may know exactly what picture they seek.

Since it is not realistically possible to vary the task mapping, semantic necessity, or the motor task, these factors were not investigated. Other factors were generally held constant: the same spatial organization (grid-aligned) was used, all file names were common English words of two or more syllables from 6-13 characters in length, and there was no systematic sorting of any kind.

Figure 1. Icon types (actual size)



METHOD

Subjects

The subjects were 45 undergraduates from the Georgia Institute of Technology who were participating for extra credit in a psychology course. Subjects in this population are typically familiar and comfortable with computers, though actual amount of experience with GUIs varied from "almost none" to "very extensive."

Apparatus/Materials

All instructions and stimuli were presented to subjects on an Apple Macintosh IIfx personal computer running Claris HyperCard version 2.1.

Design

The design of this experiment is relatively complex. There were five factors evaluated, one of which was between subjects and the rest within. The between-subjects factor was icon type, and subjects were randomly assigned to one of three levels. The three levels correspond to the levels used in the 1991 Byrne study [2]: blank (non-visual), simple, and complex. (See Figure 1.)

The second factor, a within-subjects factor, was set size. Four set sizes were used: 6, 12, 18 , and 24. The size of the match set was also varied, which was a factor termed match level. The match level (1, 2, or 3) had a different meaning in each set size. With a set size of 6, the size of the match set was 1, 2, or 3, corresponding to the match levels of 1, 2, and 3. With a set size of 12, the match levels 1, 2, and 3 corresponded to 1, 4, and 6 matching pictures. Set sizes of 18 and 24 yielded match sets of 1, 6, 9 and 1, 6, 12, respectively. Note that in each set size, a match level of 1 corresponded to a unique (to that display) target icon.

The remaining factors, also within-subjects, were the amount of knowledge the subject had about the target before the search, both picture knowledge and file name knowledge. There were three levels of icon knowledge provided. The first level, "none," meant that no special information was provided to help the subject infer the target icon before the trial. The second level, "some," meant the subject was provided with the name of the application used to create the target document, which subjects could then use to infer the correct icon. The third level, "icon," meant that the subject was explicitly given the icon to be found.

The file name knowledge factor was varied with two levels: "yes," meaning the subject was explicitly given the name of the target, and "no," meaning the name of the target could have been any word found in the document description.

Thus, a fully-crossed design yielded 72 cells (4 x 3 x 3 x 2) for each subject, and a total of 216 cells overall. Each subject performed one trial in each cell. The factors were not counterbalanced but were randomly selected from a list prior to each trial to control for systematic interaction with practice and/or subjects' anticipation of what kind of trial would come next.

There were 15 subjects used per icon type (45 subjects total) for a grand total of 3240 observations.

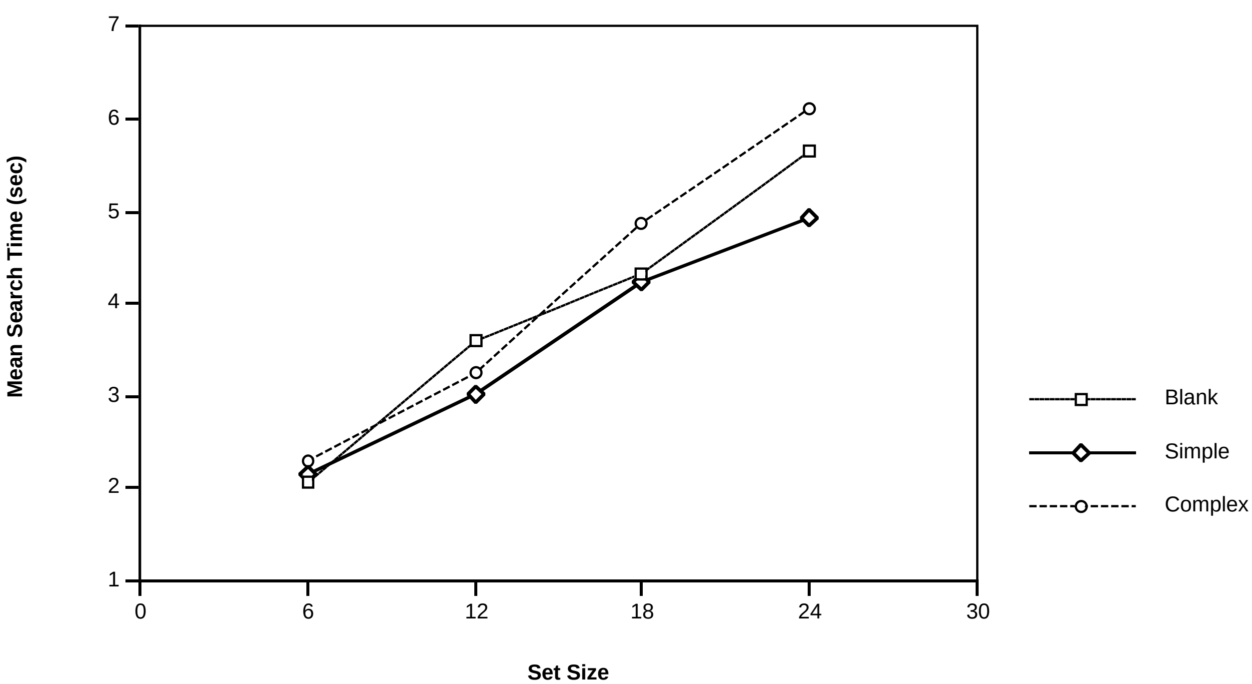
Procedures and Stimuli

Subjects first received initial instructions about how to perform the task, and were given a preview of the icons that would be used in the experiment. Subjects were given three practice trials and then 72 experimental trials.

Each trial consisted of three stages: an encoding stage, a decay stage, and a search stage. In the first stage (the encoding stage), subjects were shown a screen containing a portion of the target document, a brief description of that document (which always contained the target file name), and possibly some "hints," depending on the trial. "Hints" consisted of the target icon or explicit indication of the file name, according to the manipulations described above.

The next stage (the decay stage) consisted of a blank screen with a dialog box into which the subject was instructed to type a string of 15 random lowercase characters. If the string was typed incorrectly, the subject was given a new string to type. This continued until the subject typed in a full string correctly. This was included to prevent rehearsal and avoid floor effects, which were observed in a pilot study conducted as part of [2].

Figure 2. Time vs. Set Size by Icon Type



The final stage (the search stage) consisted of a screen resembling the Macintosh Finder. However, the main window in which the subject searched was initially blank. When he subject clicked on a button labeled "Ready," the icon display appeared and the subject attempted to click on the correct icon as quickly as possible.

The location of the target icon was randomly selected for each trial, as were the names on all the distractors. Distractor names were randomly selected without replacement from a list of 525 names until the list was exhausted, and then the list was recycled for the second half of the experiment.

RESULTS

Overall, the grand mean of the data was 4.36 seconds with a standard deviation of 6.61 seconds. Trials outside of three standard deviations from the grand mean were considered outliers and were discarded for the computation of means and regressions, but replaced by the grand mean for the ANOVA analysis (this is necessary to maintain a balanced design). 36 of the 3240 trials were outliers, which is 1.1% of the data.

Effects involving Icon Type

While there was no overall effect for icon type, it was involved in several interesting interactions. The icon type by set size interaction was significant, $F(6, 126) = 3.13$, $p = .007$; Figure 2 presents a graph of the means. The divergence of the means between groups at large set sizes suggests that different processes are operating according to what type of icon pictures are used.

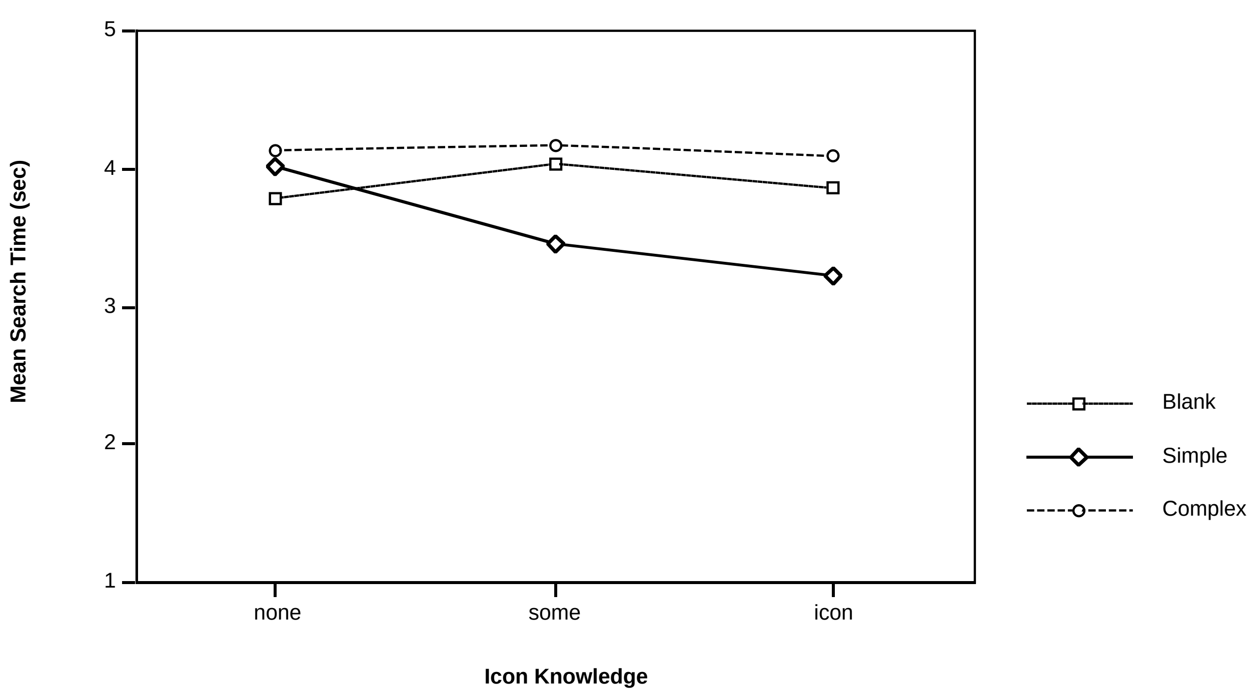
Icon type also interacted with picture knowledge, $F(4,84) = 3.51$, $p = .011$; Figure 3 shows the means. Icon knowledge affected the three types of icons differently, with simple icons apparently benefiting the most. Apparently, subjects were only able to make use of icon knowledge when the icons were simple.

Icon type interacted with file name knowledge, $F(2,42) = 4.34$, $p = .019$ (see Table 1 for means). In this case, knowledge of the file name had an effect in all three conditions, and clearly the greatest effect in the blank icon condition. However, the detrimental effect of not receiving the file name was less in the simple icon condition than it was in the complex icon condition, again suggesting that a different process is operating when searching through simple icons.

Table 1. Icon Type by File Name Knowledge means (in seconds)

	Received Filename	
Icon Type	yes	no
Blank	3.36	4.44
Simple	3.40	3.72
Complex	3.78	4.49

Figure 3. Time vs. Picture Knowledge by Icon Type



Finally, icon type interacted with match level, F(4,84) = 8.21, *p* < .001. Since match level only makes sense for the simple and complex conditions (all icons appear identical in the blank condition), the overall mean for the blank icon condition is presented for comparison in Table 2. This is one of the more interesting effects. If the target icon is unique to the display in the simple icon condition, the advantage of that condition over the others is substantial. However, if the display consists of half targets, this advantage is substantially reduced or even eliminated.

Table 2. Icon Type by Match Level means (in seconds)

	Match Level		
Icon Type	1	2	3
Blank	3.89	3.89	3.89
Simple	3.06	3.46	4.20
Complex	3.87	4.08	4.45

Other results

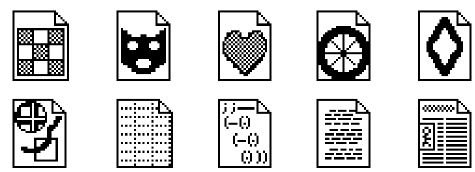
Besides results involving icon type, there were several other effects of interest. First, a practice effect was observed F(1,3202) = 5.02, *p* = .025—subjects got slightly faster over time. However, while this effect is statistically reliable, it is of little practical significance. This regression accounts for less than 0.2% of the variance (R^2 = .0016), and has a very shallow slope (a .0057 second improvement per trial).

There was also an effect for the particular icon being searched for. Obviously, with the blank icons, there was no such effect—the means are included merely for comparison. However, there was an effect for the particular icons within the simple icons, F(4,1062) = 2.96, *p* = .019, and within the complex icons, F(4,1062) = 3.51, *p* = .007. Table 3 presents the means. Not surprisingly, the complex icon that was the most successful was the one that could likely be found on the basis of a single feature (curvature, for example). The "best" icon in the simple group is the only icon with repeating horizontal lines, another unique feature.

Finally, regression analyses were conducted to test the visual search model, $f_m(n) = f_v(n) + f_s(t)$. Due to the fact that there were interactions involving icon type, set size, icon knowledge and file name knowledge, only a portion of the data was used to ensure consistency of process for all comparisons. In the real-world task, there are typically a fixed number of icons available, so a user is most likely to know what the icon is. File names are more variable, so it is assumed that the file name is not exactly known. That is, data in the "icon" condition of icon knowledge and the "no" condition of file name knowledge was selected for regression analysis.

For the blank icons, the linear regression of time on set size was reliable, F(1,527) = 163.0, p < .0001, and accounted for nearly a quarter of the variability (R^2 = .24), which is quite good for non-aggregated data. (If the data is collapsed for each subject, the R^2 rises to near .70.) The regression yields the equation $f_s(n)$ = .930sec + .235sec*n. The number of matching icons in each trial (t) was then plugged in to the preceding equation and subtracted from the total time, which is $f_m(n)$, yielding an estimate of $f_v(n)$. Mean $f_v(n)$'s for each set size are presented in Figure 4. Clearly, the $f_v(n)$ for complex icons increases much more with increases in set size than does the $f_v(n)$ for simple icons. In fact, $f_v(n)$ for simple icons seems to follow the predictions made by the model of Cave and Wolfe [3, 13]: an initial increase at small set sizes which flattens out as set size grows.

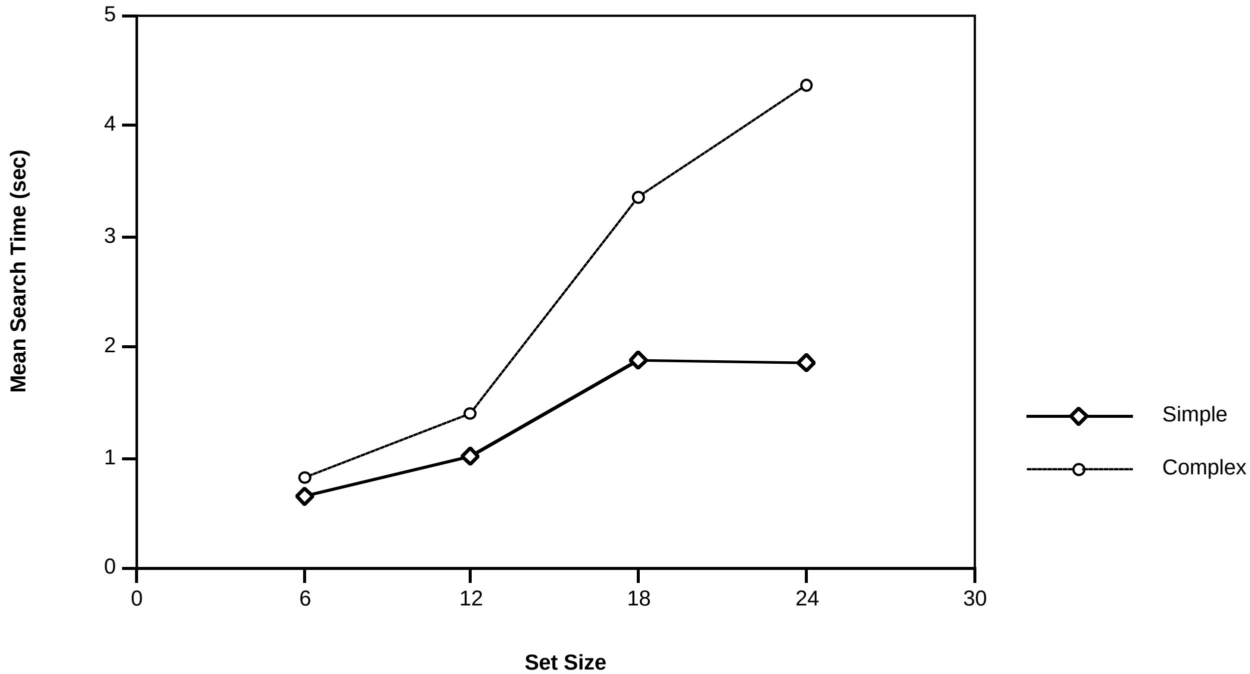
Table 3. Means for specific icons (in seconds)



	Icon				
Icon Type	1	2	3	4	5
Blank	3.89	3.89	3.89	3.89	3.89
Simple	3.08	3.34	3.65	3.88	3.88
Complex	3.52	4.09	4.14	4.36	4.59

The estimate of visual search time was regressed on set size for each type of icon. Due to the large number of data points and excellent power, accounting for even small amounts of variance yields statistical reliability. The best comparison, then, is in the amount of variability accounted for by the linear effect. For the simple icons, the regression was, of course, reliable, F(1,176) = 10.0, *p* = .0018. However, this accounted for only about 5% of the variability (R^2 = .05). For complex icons, the regression is again reliable, F(1,176) = 52.5, *p* < .0001. In this case, the proportion of variance accounted for is much greater—almost five times as great, in fact (R^2 = .23). This suggests that a linearly increasing model is much more appropriate for the visual search with complex icons than for simple icons, which means that simple icons should hold an advantage, particularly at larger set sizes.

Figure 4. Visual Search Time vs. Set Size by Icon Type



DISCUSSION

Clearly, one of the most important factors in icon search is the type of icon used. Simple icons clearly outperform blank and complex icons, especially with larger set sizes. There is support for the two-pass model, and the slope for the visual search component with simple icons is much flatter than the slope for the purely semantic search or for the visual search with complex icons. The lesson: for icons to be effective aids to visual search, they must be simple and easily discriminable. Simple icons are more effective with larger set sizes, allow effective use of icon knowledge, are less affected by a lack of file name knowledge, and are especially effective when they are unique to the display. With simple icons, there is reason to accept the design assumption that icon pictures make finding files easier.

Complex icons are consistently worse than even blank icons. They seem to clutter the display with information that users are unable to employ to their advantage. Thus, if the icons are complex and difficult to discriminate quickly, the assumption that icon pictures are useful in this task clearly does not hold.

Though the experiment was done in monochrome, one could easily extrapolate the basic guideline to color as well: use only a few, easily discriminable colors. Simply because many computer displays now support some 16 million colors should not be taken as a mandate to use more than a few.

Perhaps counter-intuitively, practice does not seem to help users find documents more quickly. This is one area where it seems that the novice/expert distinction is not meaningful. It does not seem possible to compensate for poor icon design simply with practice.

This research does not address several important issues, such as semantic sorting (e.g. alphabetically by name) or spatial organization (aligned to a grid vs. scattered placement). These factors may play a significant role in finding documents with icons. It is interesting to note that many experienced Macintosh users, for instance, choose to view their files not by icon, but as a list of names sorted by date or alphabetically by file name.

This research also does not address the use of icons in tasks other than visual search. It may be the case that complex icons can usefully serve other functions in which the speed of visual search is not particularly important or in which there are few icons on the display.

ACKNOWLEDGEMENTS

I would like to thank David Cater for his help in running subjects and generating experimental materials. I would also like to thank Susan Bovair and Neff Walker for their comments on earlier drafts of this paper. This research was supported primarily by a graduate fellowship to the author from the National Science Foundation.

REFERENCES

1. Baecker, R., Small, I., & Mander, R. (1991). Bringing icons to life. *Proceedings of CHI'91*, 1-6.

2. Byrne, M. D. (1991). The misunderstood picture: A study of icon recognition. Poster presented at CHI'91.

3. Cave, K. R., & Wolfe, J. M. (1990). Modeling the role of parallel processing in visual search. *Cognitive Psychology, 22*, 225-271.

4. Fisher, D., Coury, B. G., Tengs, T. O., & Duffy, S. A. (1989). Minimizing the time to search visual displays: The role of highlighting. *Human Factors, 31(2)*, 167-182.

5. Fisher, D., & Tan, K. C. (1989). Visual displays: The highlighting paradox. *Human Factors, 31(1)*, 17-30.

6. Fisher, D. L., & Tanner, N. S. (1992). Optimal symbol set selection: A semiautomated procedure. *Human Factors, 34(1)*, 79-95.

7. Hemmenway, K. (1981). Psychological issues in the use of icons in command menus. In *Proceedings of CHI'81*, 20-23.

8. Jones, W. P., and Dumais, S. T. (1986). The spatial metaphor for user interfaces: Experimental tests of reference by location versus name. *ACM Transactions on Office Information Systems, Vol. 4, No. 1*, 42-63.

9. Nolan, P. R. (1989). Designing screen icons: Ranking and matching studies. In *Proceeding of the Human Factors Society 33rd Annual Meeting* (pp. 380-384). Santa Monica, CA: The Human Factors Society.

10. Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search and attention. *Psychological Review, 84*, 1-66.

11. Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Detection, search, and attention. *Psychological Review, 84*, 127-190.

12 Triesman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology, 12*, 107-141.

13. Wolfe, J. M., Cave, K. R., & Franzel, S. L. (1989). Guided search: An alternative to the feature integration model for visual search. *Journal of Experimental Psychology: Human Perception and Performance, 15*, 419-433.