The Spatial Metaphor for User Interfaces: Experimental Tests of Reference by Location versus Name

WILLIAM P. JONES and SUSAN T. DUMAIS
Bell Communications Research

The enduring dichotomy between spatial and symbolic modes of representation and retrieval acquires an added pragmatic dimension through recent developments in computer-based information retrieval. The standard name-based approach to object reference is now supplemented on some systems by a spatial alternative—often driven by an office or desktop metaphor. Little rigorous evidence is available, however, to support the supposition that spatial memory in itself is more effective than symbolic memory.

The accuracy of spatial versus symbolic reference was assessed in three experiments. In Experiment 1 accuracy of location reference in a location-only filing condition was initially comparable to that in a name-only condition, but deteriorated much more rapidly with increases in the number of objects filed. In Experiment 2 subjects placed objects in a two-dimensional space containing landmarks (drawings of a desk, table, filing cabinets, etc.) designed to evoke an office metaphor, and in Experiment 3 subjects placed objects in an actual, three-dimensional mock office. Neither of these enhancements served to improve significantly the accuracy of location reference, and performance remained below that of a name-only condition in Experiment 1. The results raise questions about the utility of spatial metaphor over symbolic filing and highlight the need for continuing research in which considerations of technological and economic feasibility are balanced by considerations of psychological utility.

Categories and Subject Descriptors: H.1.2 [**Models and Principles**]: User/Machine Systems—*human factors*; *human information processing*; H.3.2 [**Information Storage and Retrieval**]: Information Storage—*file organization*; H.4.1 [**Information Systems Applications**]: Office Automation

General Terms: Experimentation, Human Factors, Management, Performance

Additional Key Words and Phrases: Computer-human interaction, personal filing systems, spatial representation, user interface

1. SPATIAL VERSUS SYMBOLIC FILING: EVIDENCE FROM COGNITIVE PSYCHOLOGY

The metaphorical use of space as a means of organizing thoughts or ideas dates at least back to the formalization of the Method of Loci during the fifth century B.C. (see Yates [18] for a discussion). Recently, there has been an attempt to give the abstract objects of computer filing systems (e.g., files, commands,

Some of the material in this article was presented at the 25th Annual Meeting of the Psychonomics Society, Nov. 1984, San Antonio, Tex., and at the CHI'85 Human Factors in Computing Systems Meeting, April 1985, San Francisco, Calif.

Authors' present addresses: W. P. Jones, Microelectronics and Computer Technology Corporation (MCC), Echelon #1, 9430 Research Blvd., Austin, TX 78759; mccabe@MCC. S. T. Dumais, Bell Communications Research, 435 South St., Morristown, NJ 07960; std@bellcore.

procedures) locations as well. For example, many new systems employ desktop metaphors in which an object can be given a location instead of, or in addition to, a name through its placement in a two-dimensional space on the terminal screen. This approach is pushed close to a logical extreme in the Spatial Database Management System (SDMS) [2], in which picture-quality representations of objects are placed in a simulated three-dimensional space representing a setting familiar to the average user (e.g., an office or a library).

These systems attempt to engage a spatial facility that is of clear importance to us in interactions with the tangible objects of our world. It is not enough to know what we are looking for; we must also know where to look for it. Although the idea of using spatial metaphors in user interfaces sounds like a reasonable one, many recent studies of computer–human interaction have shown that the intuitions of systems designers are not always in accord with the needs or abilities of users. It is worth examining the basis of the argument before embracing it wholeheartedly. We begin by considering existing evidence from psychological studies of memory for the location of objects.

Mandler et al. [13] report experiments in which subjects were asked to recall the locations of 16 toys, placed in a 6 × 6 array of 36 locations. The recall of both objects and their locations was measured following instructions that either both would be tested (intentional learning of location) or only object recall would be tested (incidental learning of location). This instructional manipulation had only small, marginally significant effects on location recall. Furthermore, the intention to recall the location of objects did not appear to hurt the subject's ability to recall the names of these objects. These results have been interpreted to suggest both that the processing of location information is incidental (in the sense that it is stored in the absence of any conscious intention to do so) and that it is effortless (in the sense that it does not appear to utilize processing resources needed to remember other object information); see Hasher and Zacks [7] for a review. A similar pattern of results has been found for the recall of locations and labels of pictures [17] and for the recall of passage content and location in a text [20]. Incidental recall of passage location has been demonstrated in other experiments as well [4, 14, 19].

The apparently incidental and cost-free nature of location memory for even the abstract material of a text [4, 14, 19, 20] suggests that such spatial facilities might be profitably engaged in a computer database situation to increase the ease with which objects are organized and retrieved. However, this conjecture must be qualified in at least two respects. First, research on the incidental and automatic aspects of spatial memory is comparatively sparse, and nonconfirmatory results have been found in at least two studies [1, 15]. Second, the applied use of spatiality in a computing environment raises a set of questions that have not been addressed previously in the psychological literature.

The traditional and still dominant mode of object reference on computing systems is symbolic; data files, commands, programs, etc., are initially labeled and subsequently referred to by name. To make a valid comparison with spatial representations, we need to know more about the comparative utility of symbolic and spatial modes of object reference. One important measure of performance is the accuracy with which a user can specify or find an object. The accuracy of object reference clearly depends upon a user's ability to connect an object's

> Astronaut Sally Ride and her four male colleagues are cramming in a few final hours of training before Saturday's launch of the seventh space shuttle mission.
>
> Ms. Ride, commander Robert Crippen, pilot Rick Hauck, and mission specialists John Fabian and Dr. Norman Thagard were to arrive at Cape Canaveral for the training late today, piloting three T-38 jets from the astronaut training base in Houston.
>
> Ms. Ride is the first woman named to a U.S. space flight.
>
> At the launch pad, crews were preparing to start the countdown at 3:00 a.m. EDT Thursday, aiming for lift-off of the shuttle Challenger at 7:33 a.m. Saturday for the six-day orbital flight. The count will begin when launch director Al O'Hara issues a formal "call to stations," which will summon more than 50 engineers and technicians to launch control center consoles.
>
> O'Hara said the countdown for the mission has been cut in half from nearly five days to slightly more than two because of experience gained on the first six shuttle flights.

Fig. 1. A sample news article.

internal contents and its external representation. How will users' memories for an object's location compare with their memories for that object's name? It is especially important to determine how these memories compare as a function of crowding, that is, increases in the size of the database. The spatial memory resolution required to specify an object successfully in even a moderately sized database, say 20 or 30 objects, has not yet been demonstrated in empirical studies.[1] Finally, it is important to determine the extent to which symbolic and spatial aspects of an object's representation can be combined usefully in a computerized information retrieval system. A database object could have both a name and a location. But, to what extent can a user integrate these diverse sources of information during retrieval? Is the time needed to select such a dual representation better spent in the elaboration of a single-mode representation (i.e., name only or location only)?

The three experiments described in this paper were an initial attempt to address these issues concerning the applied use of spatiality as a means of representing, organizing, and retrieving information. Specifically, these experiments are intended to provide a basis on which to compare the accuracy of location-based, name-based, and location+name-based object reference under conditions of an expanding database.

2. EXPERIMENT 1: FILING BY NAME VERSUS LOCATION

2.1 Method

2.1.1 *Database Objects: News Articles.* In Experiment 1 subjects read and "filed" news articles in four filing conditions. The news articles covered a wide range of topics of general interest to the average person. A sample article is printed in Figure 1. The articles were truncated to the first five paragraphs

[1] In studies demonstrating incidental recall of passage location [4, 14, 19, 20], for example, subjects are typically required to specify only the eighth or quarter of the page in which a passage occurs.

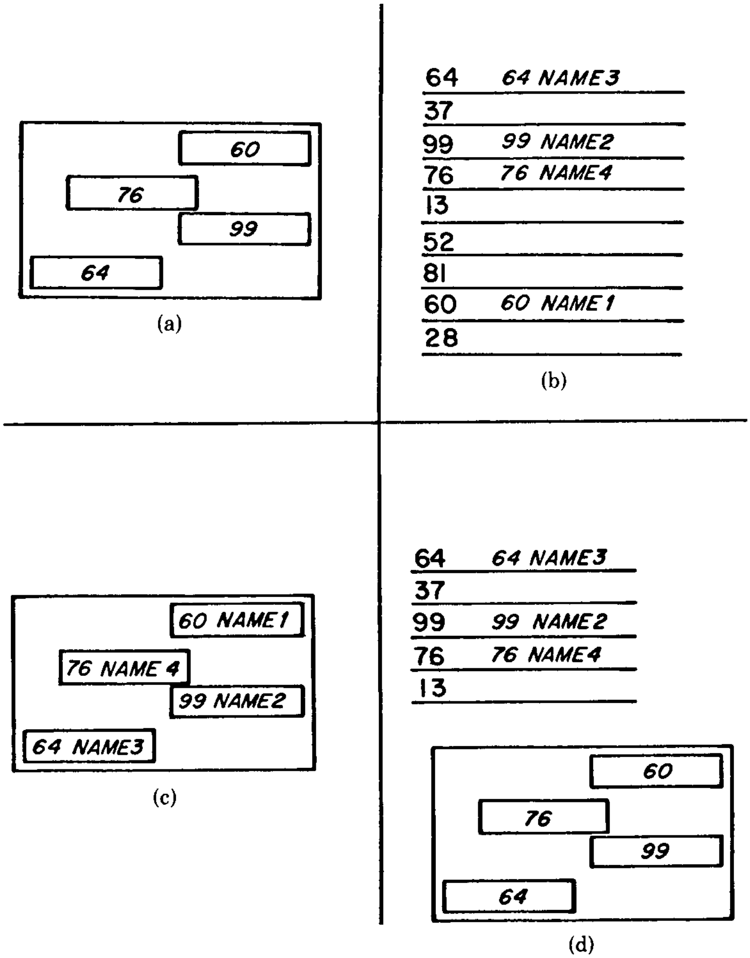

Fig. 2. Schematization of the four filing conditions of Experiment 1: (a) location only; (b) name only; (c) name+location combined; (d) name+location separate. (Note that the actual size of labels relative to the filing sheets was much smaller than it appears in the figure.)

(about 12–14 lines). Each article was printed on a separate sheet of paper in a notebook that subjects worked through, one sheet at a time. A unique two-digit identifying number was printed in the upper left-hand corner of each article sheet.

2.1.2 *Conditions.* Subjects filed an article immediately after reading it by sticking a small gummed label containing the article's identifying number on a separate sheet. The way in which this was done depended on the filing condition to which a subject was assigned (see Figure 2).

The *location-only* condition (see Figure 2a) represents a filing scheme in which objects are represented and retrieved by their spatial locations. Subjects in this condition filed an article by writing the article's identifying number on a gummed label, and then placing this label at a location of their choosing on a blank 14 × 11-inch filing sheet. The labels could overlap as long as all the identifying numbers remained visible. In the instructions for this condition, the filing sheet was referred to as the "desk sheet." The subjects were encouraged to arrange the labels on their desk sheet in a manner similar to that in which they might arrange the paper objects on an actual desktop.

The *name-only* condition (see Figure 2b) is analogous to a typical computer filing system in which a name is assigned to each object and serves as the primary route to its subsequent retrieval. In this condition, subjects selected a name for each article after reading it. They wrote this name, along with the article's identifying number, on a gummed label, and then placed the label on an 8.5 × 11-inch filing sheet (referred to as the "name sheet" in the instructions). The subjects were not free to select an article's location. Instead, labels were placed so that their article numbers matched corresponding numbers on the name sheet. The numbers on the name sheet were randomly arranged in two columns. The subjects were also severely constrained in the selection of names—a name could be only two letters in length.[2]

In the *name+location-combined* condition (see Figure 2c) subjects selected both a name and a location for each article. After reading an article, subjects first wrote the article's identifying number and a two-letter name of their choice on a gummed label. They then placed this label at a position of their choice on a blank 14 × 11-inch filing sheet. This filing system represents symbolic and spatial information in an integral fashion; that is, the objects on the desktop are labeled.

The subjects also chose both the name and location of articles in the *name+location-separate* condition (see Figure 2d). In this condition, however, the two kinds of information were separately represented. The subjects first placed one label for an article on a desk sheet (in the same way as in the location-only condition). They then placed a second label containing their name choice on a separate sheet (in the same way as in the name-only condition). Although both spatial and symbolic information were available, they were not integrated into a single filing system.[3]

2.1.3 *Measuring Recall.* In all four conditions, the news articles that a subject encountered were organized into one practice block and three experimental blocks of ten articles each.[4] The subjects were asked to read and file each article according to the scheme outlined for their condition. After each block of ten

[2] In pilot studies, subjects who were allowed to use longer names had near perfect success in retrieval. This left no room for us to observe any added benefits of location cues (see Figure 2c and d).

[3] This condition is not as anomalous as it might first appear. In actual computing environments, users can usually select longer names; most of today's systems permit names of at least eight characters. If these names are placed in a metaphorical desk or office space, however, that space will quickly fill up. It might thus be better to give an object in such a space a representation that takes up less room (perhaps an icon) and to store a more complete name separately. In addition, it is sometimes useful to have different views of the same information. For example, files can be listed alphabetically or by time of creation. The spatial and symbolic sheets in this condition provide two different views of the same objects.

[4] The 40 articles used for a given subject were randomly selected from a pool of 60 articles. Twenty different subsets were constructed, and no two subjects in a given condition used the same subset. Subsets were reused in different conditions.

Table I. A Breakdown of the Experimental Task Illustrating the Composition of Successive Retrieval Tests[a]

	Articles	Retrieval statements	Block to which statements refer
Practice	10	3	
Experimental			
Block 1	10	3	Block 1 (3)
Block 2	10	6	Block 1 (3) Block 2 (3)
Block 3	10	9	Block 1 (3) Block 2 (3) Block 3 (3)

[a] After filing a block of ten articles, subjects completed a retrieval test consisting of from three to nine statements.

articles, the subjects completed a retrieval test consisting of a series of statements. Each statement uniquely referred to one of the articles that the subjects had previously read and filed. For instance, the statement referring to the article in Figure 1 was: "*She and four others crammed in a few final hours of training before the seventh of its kind got underway.*" The subjects had three chances to guess which article each statement referred to. They indicated their guesses by writing the articles' identifying numbers, which could be obtained by referring to the filing sheet(s), in the appropriate blanks on the statement sheets.[5] In all four filing conditions, therefore, subjects responded in exactly the same way to test statements (three two-digit identifying numbers), although their responses were based on very different object representations. Subjects received no feedback concerning the accuracy of their guesses.

In the selection of statements, care was taken to avoid the inclusion of topic words, that is, words that directly referred to the topic in the corresponding articles. This was done to ensure that guesses in all conditions were mediated by a subject's memory for the originally selected article representation (i.e., names and/or locations). Specifically, we wanted to minimize the possibility that a statement in the name conditions could be accurately answered through a direct match between words in the statement and names on the filing sheet. (This was considered to be unlikely in any case because the names were constrained to be only two letters in length.)

Table I illustrates the statement composition for the different retrieval tests. The retrieval test after the practice block consisted of three statements, each of which referred to a different article that the subjects had just filed. After completing this test, the subjects turned in their practice sheets and were given new filing sheets to be used in the remainder of the experiment. The retrieval

[5] In postexperimental interviews, subjects in all four filing conditions generally indicated that they had no direct memory for an article's identifying number. Performance on the retrieval tests, therefore, critically depended on the use of the filing sheets. Those subjects who did volunteer that they were occasionally able to directly recall an article's identifying number appeared to be evenly distributed among the filing conditions.

test after the first experimental block consisted of three statements, each of which referred to an article in the block of ten articles that had just been filed. The retrieval test after the second block consisted of six statements. Three of the statements referred to articles that had been filed in the first block, and three referred to articles filed in the second (most recent) block. Finally, the retrieval test after the third block consisted of nine statements. Three referred to articles filed in the first block, three to articles filed in the second block, and three to articles in the third (most recent) block. Over the course of the three retrieval tests, no article was tested more than once. (The subjects were not told about this constraint, nor were they told about the composition of statements in the retrieval tests.)

Note that during the experimental phase, more and more articles were added to the filing sheet during successive blocks. This provides a way to assess the effects of crowding on a person's ability to retrieve a desired object in the different filing conditions. The nature of the statement composition in successive retrieval tests also permits us to separately examine the effects of *retroactive* and *proactive* inhibition [5, pp. 217–263]. To do this, we partition the test statements at the point at which the articles that they reference were filed. A *first-block* statement refers to one of the ten articles filed in the first block (i.e., the block immediately preceding the first retrieval test), and a *most-recent-block* statement refers to one of the ten articles filed in the block immediately preceding the current retrieval test. (Note that the three statements of the first retrieval test are necessarily both first-block and most-recent-block statements.) With respect to first-block statements, retroactive inhibition increases in succeeding test blocks, whereas the effects of proactive inhibition remain constant. With respect to most-recent-block statements, proactive inhibition increases in successive blocks (i.e., the total number of articles filed increases), whereas retroactive inhibition remains constant.

2.1.4 *Subjects.* Seventy-one Murray Hill, New Jersey, area homemakers ranging from 23 to 72 years of age participated in Experiment 1. Many of these subjects had previously participated in psychological experiments, but none had experience with computerized filing systems. Twenty subjects were assigned to the name-only condition, and 17 were assigned to each of the other three conditions. The subjects were tested in groups of 7 to 10, but they were allowed to work through the articles and statements at their own pace. All the subjects in a group worked in the same filing condition.

Fifty-two subjects (14 location-only, 12 name-only, 13 name+location-combined, 13 name+location-separate) were able to return 13–17 days later to complete a *30-statement* retrieval test. Eighteen of these statements were identical to those that subjects had previously answered during the three retrieval tests of the first session; the remaining 12 statements referred to the articles not tested during the first session. The subjects used their original filing sheets to answer the 30 statements, but they were not allowed to review the actual articles.

2.2 Results

2.2.1 *Comparing Conditions.* The performance measure of primary interest in Experiment 1 is the accuracy with which subjects generated the correct response

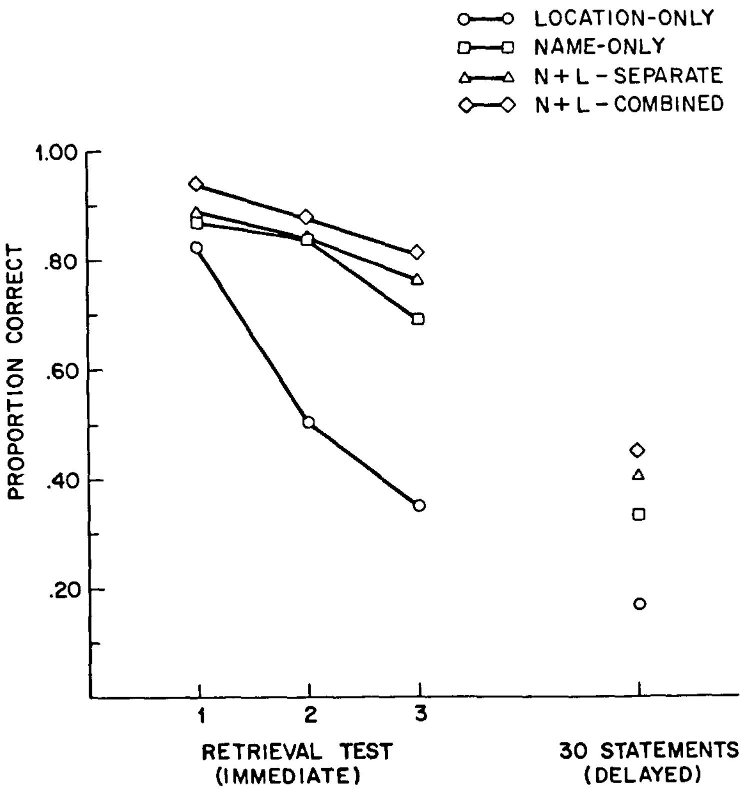

Fig. 3. First-guess accuracy (proportion correct) in the three retrieval tests of the first experimental session and the 30-statement delayed test of the second session for each of the four filing conditions of Experiment 1.

to a test statement on the first guess. Figure 3 presents mean first-guess accuracy for the three retrieval tests of the first session and for the delayed 30-statement test for each of the four filing conditions. An analysis of variance on the data for the first experimental session revealed significant effects for the between-subjects factor of filing condition, $F(3, 67) = 10.01, p < .001$, the within-subjects factor of retrieval test, $F(2, 134) = 30.01, p < .001$, and the interaction between these two factors, $F(6, 134) = 4.94, p < .001$.

Three preplanned pairwise contrasts between filing conditions were also completed. In a comparison of the name-only condition with the name+location-separate condition, the factor of filing condition is not significant, $F(1, 35) < 1$, nor is the interaction between condition and retrieval test significant, $F(2, 70) < 1$. In a pairwise comparison of the name-only condition with the name+location-combined condition, the factor of filing condition is again not significant, $F(1, 35) = 2.56, p > .110$, nor is the interaction between condition and retrieval test significant, $F(2, 70) < 1$. In contrast, a pairwise comparison between the name-only and location-only conditions reveals a significant main effect for filing condition, $F(1, 35) = 11.94, p < .001$. In addition, the factor of filing condition interacts with the factor of retrieval test, $F(2, 70) = 8.27, p < .001$.

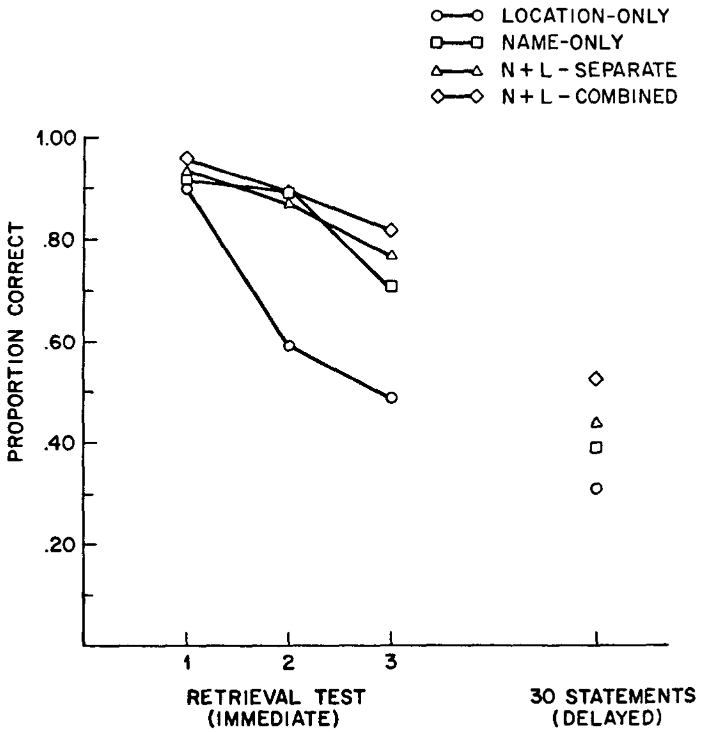

Fig. 4. Three-guess accuracy (proportion correct when all three guesses to a statement are considered) in the three retrieval tests of the first experimental session and the 30-statement delayed test of the second session for each of the four filing conditions of Experiment 1.

2.2.2 *Delayed Tests.* An analysis of variance was also conducted on the first-guess accuracy data for those subjects who completed the 30-statement task of the second experimental session. This analysis revealed a significant main effect of filing condition, $F(3, 47) = 6.72$, $p < .001$. In planned pairwise comparisons, the difference between the name-only and location-only conditions was significant, $F(1, 24) = 7.27$, $p < .013$, and the difference between the name-only and name+location-combined conditions was marginally significant, $F(1, 22) = 3.84$, $p < .063$. A comparison of the name-only and name+location-separate conditions was not significant, $F(1, 23) = 2.01$, $p > .170$.

The improvement in performance that results from a consideration of all three guesses to a statement (a statement is judged to have been correctly answered if at least one of the three guesses is correct) is above chance in all conditions. However, as Figure 4 demonstrates, the pattern of results is essentially the same as that for first-guess accuracy. Analyses of variance reveal the same significant effects as those reported for the first-guess data, and these results are not described in detail here.

Analyses of variance were performed separately on the first-guess accuracy for the first-block and most-recent-block statements (as these were described in the methods section). Again the pattern of results is essentially the same as that described earlier. That is, for both first-block and most-recent-block statements, the effects of filing condition, retrieval test, and their interaction are significant. In addition, all pairwise contrasts between the location-only and the name-only conditions are significant. The only discrepancy from previous analyses occurs in the pairwise contrast on most-recent-block statements between the name-only and the name+location-combined conditions. In this contrast, the main effect of filing condition is now marginally significant, $F(1, 35) = 2.86$, $p < .100$, and its interaction with retrieval test is significant, $F(2, 70) = 3.41$, $p < .039$.

First-guess performance on first-block and most-recent-block statements are plotted separately in Figure 5 as a function of retrieval test for each of the four filing conditions. Across conditions, accuracy on the most-recent-block statements is marginally superior to that for first-block statements, $F(1, 67) = 3.36$, $p < .071$. The factor of statement type (first-block versus most-recent-block) interacts with filing condition, $F(3, 67) = 2.84$, $p < .044$.[6] The data in Figure 5 suggest that the difference between first-block and most-recent-block statements is negligible for all but the location-only condition.

2.3 Summary and Discussion of Experiment 1

Overall, performance in the location-only condition was considerably worse than in any condition involving a name. More important, the decline in performance over successive retrieval tests (i.e., as the space became more crowded) was much more precipitous in the location-only condition than in the name conditions. (Note that performance in the first retrieval test was roughly comparable in all conditions.) The data indicate, therefore, that a strictly spatially based information retrieval system will be particularly vulnerable to increases in the size of the database. The data in Figure 5 further suggest that location memory for less recently filed items is especially poor.

Why is performance in the location-only condition so poor? It is possible that the initial representation of an object's location is not sufficiently precise to distinguish it from objects subsequently filed. When only a few objects are filed, for example, the user may only need to remember an object's approximate location (e.g., "upper left-hand corner"). Such a representation, although initially adequate, would present problems as the space became more crowded. It is also reasonable to suppose that memory for a location is not absolute and is represented instead as part of a configuration involving other objects in the filing space. Retrieval problems are then to be expected when there is a mismatch between the user's representation and the current configuration of the filing space.

For first-block statements, however, the delay between the selection of an object's representation and its subsequent use also increased in succeeding retrieval tests. It is possible, therefore, that location memory is simply more vulnerable to general factors of decay associated with the passage of time. The results of the delayed test provide some support for this notion. When first-guess performance on the 18 statements encountered in both the first session and the delayed tests is compared in subjects who appeared in both sessions,

[6] The data from the first retrieval test were excluded from these analyses since first-block and most-recent-block statements are necessarily identical.

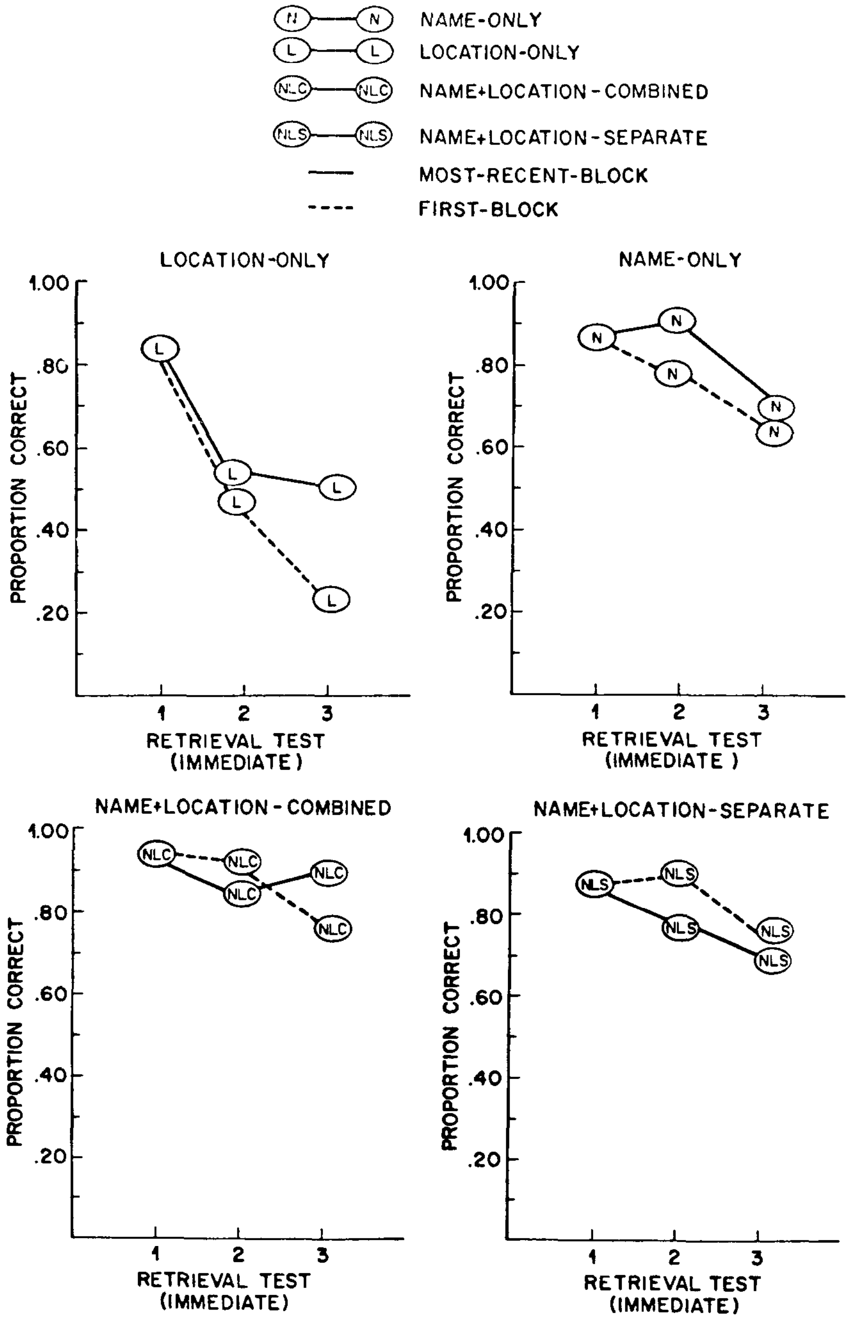

Fig. 5. First-guess accuracy in each of the filing conditions of Experiment 1, separately plotted for first-block and most-recent-block statements.

location-only performance in the delayed condition is only 33.2 percent of that observed in the first session (a weighted average of performance on the three retrieval tests). Comparable numbers for the other conditions are 50.7 percent (name-only), 56.1 percent (name+location-separate), and 59.1 percent (name+location-combined).

In all four filing conditions, the improved performance that results from a consideration of all three guesses is well above chance. If an imperfect memory for an object's representation is insufficient to guarantee success on the first guess, it may nevertheless serve to constrain the set of reasonable alternatives so that the likelihood of success on subsequent guesses is good. However, despite the fact that the absolute levels of improvement are largest in the location-only condition (when the three-guess data are considered), there is no indication that partial recall is more useful in the location-only condition than in the name-only condition. The probability of generating a correct answer on the second or third guess, given that the first guess was incorrect, is 21.6 percent in the location-only condition and 19.1 percent in the name-only condition.[7]

Leaving aside the poor performance in the single-mode, location-only condition, the data of Experiment 1 do little to support the utility of a dual-mode, name+location representation. Performance in the two conditions in which subjects selected both an object's name and its location was only slightly better than that in the name-only condition.[8] In contrast, data from a pilot study indicate that greater gains can be realized in a name-only system that simply permits longer names (i.e., more than two letters). In a condition in which names of up to ten letters were allowed, the mean first-guess performance (for a group of ten homemakers from the same subject pool as that used in Experiment 1) was 96.7, 91.3, and 88.1 percent in successive retrieval tests of the first session. The corresponding figures for the best condition of Experiment 1 (name+location-combined) were 94.1, 88.2, and 83.3 percent. This is not to suggest that performance in a name-only condition will be superior to performance in

[7] There is reason to believe that even this small difference might disappear if feedback were given about the correctness of a guess (a situation more representative of an actual computerized filing system). In Experiment 1, the second and third guesses were often the same as the first guess. We believe that this reflects a subject's confidence, often misplaced, that the first guess was correct. Subjects were more likely to repeat their guesses in the name-only condition than the location-only condition. Given an incorrect first guess, subjects generated the same response as a second guess 61 percent of the time in the name-only condition compared to only 44 percent in the location-only condition.

[8] Interestingly, one can predict performance in the name+location-combined condition from performance in the name-only and location-only conditions. Assume that one fails to retrieve an object in the name+location-combined condition if one can retrieve neither its name nor its location. That is,

$$(1 - \text{name+location-combined}) = (1 - \text{name-only}) * (1 - \text{location-only}).$$

In the present experiment, estimates obtained in this manner are reasonably close to observed performance (.91 versus .88 for the first three retrieval tests and .47 versus .51 for the delayed test), especially considering that different subjects served in the different conditions. These results suggest that the benefit derived by combining name and location cues is simply a probabilistic summation. A similar benefit is not derived in the name+location-separate condition. Perhaps this is due to subjects relying primarily on the name information.

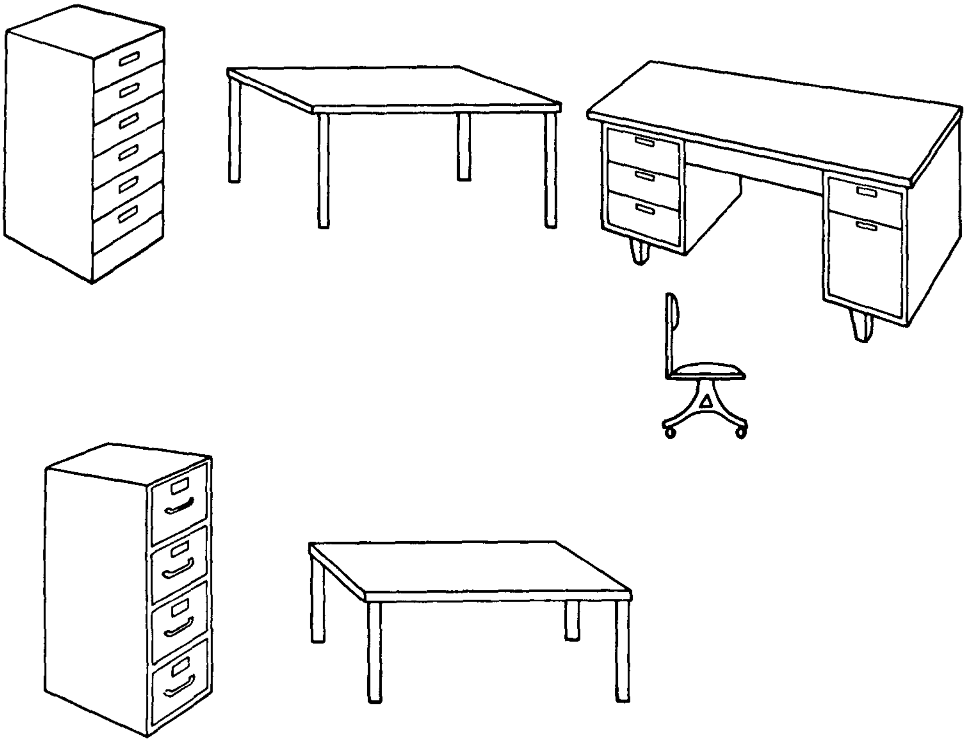

Fig. 6. The "enhanced" filing sheet used in Experiment 2.

a name+location-combined condition with names of comparable length. This comparison does suggest, however, that the time spent selecting an object's location might be better spent in the elaboration of the object's name.

The results of Experiment 1 suggest that the usefulness of spatial information, either alone or in combination with symbolic information, may be limited in information retrieval settings. One caveat, however, is immediately obvious. The space offered by the blank "desk sheet" used in Experiment 1 was very barren (although perhaps not unlike that offered by a blank CRT screen). Performance might improve if the space were enriched through the introduction of landmarks or the use of graphics providing a more credible simulation of a three-dimensional space. We explored these possibilities in Experiments 2 and 3.

3. EXPERIMENTS 2 AND 3: DO SPATIAL ENHANCEMENTS HELP?

Experiment 2 replaced the blank filing sheet of Experiment 1 with an identically dimensioned, but "enhanced" sheet containing representations of various pieces of office furniture (see Figure 6). In all other respects, the filing system and general procedure were identical to the location-only condition of Experiment 1. The enhanced space was intended to increase the number of landmarks by which subjects could anchor the position of an object, and to reinforce a spatially oriented office metaphor for filing. Twenty-nine new subjects from the same pool as that used in Experiment 1 completed the first session of Experiment 2.

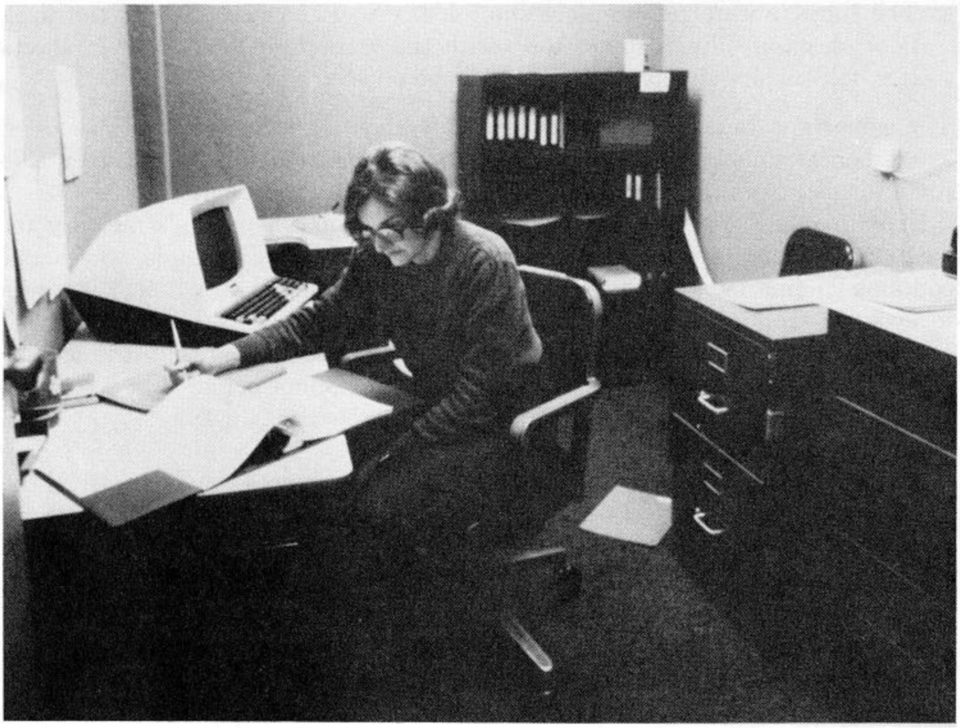

Fig. 7. A photograph of the mock office used in Experiment 3.

Twenty-five of these subjects were able to come back about two weeks later to complete a 30-statement retrieval test (identical to that used in Experiment 1).

A location-based office metaphor for information retrieval was carried considerably further in Experiment 3 by placing subjects in an actual three-dimensional office setting (see Figure 7). Twenty new subjects from the same pool as Experiments 1 and 2 completed a single session of Experiment 3. The second session was omitted because of problems associated with reconstructing the office setting so that its appearance at the outset of the second session would be identical to its appearance at the end of the first session for a given subject. (The same office space was used for all subjects in Experiment 3.)

The general procedure of Experiment 3 was identical to that of Experiments 1 and 2. The use of an actual three-dimensional office, however, necessitated the following changes in the filing procedure:

1. The subjects were tested individually, rather than in groups, as in the two previous experiments.
2. After reading an article, the subjects tore the sheet that it was printed on out of their notebook and placed it in a standard manila folder.
3. The label numbered by subjects was placed directly on this folder. The folder was then stapled shut. This was done to discourage subjects from peeking inside the folder during the retrieval tests.

4. The subjects were free to place the folder on any surface in the room (e.g., floor, desktop, filing cabinet top), with the restriction that existing objects (i.e., folders already placed) were not to be moved.

The subjects were permitted to move freely about the room at all times. The experimenter unobtrusively observed the subject's activities from an outside room to ensure proper compliance with the prescribed procedure.

Experiments 2 and 3 explored the extent to which a person's memory for location depends on the nature of the space involved, for example, whether it contains landmarks, whether it is two or three dimensional, or whether it serves to evoke a familiar (to some subjects, at least) situation of object reference. Special interest can be attached to Experiment 3 to the extent that it is seen as a test of the utility of a spatially oriented office metaphor per se.

3.1 Results and Discussion of Experiments 2 and 3

Figure 8 presents mean first-guess accuracy[9] in Experiments 2 and 3 for the three retrieval tests of the first session and also for the 30-statement test of the second session in Experiment 2. (Recall that subjects did not complete the 30-statement test in Experiment 3.) Comparable values from the location-only and name-only conditions of Experiment 1 are included for comparison.

There is no indication that the inclusion of landmarks (drawings of office furniture) on the two-dimensional desk sheet of Experiment 2 did anything to improve the accuracy of spatial reference. Performance in Experiment 2 was nearly identical to that in the location-only condition of Experiment 1 ($F < 1$ for the main effect of filing condition). In fact, performance in the first retrieval test was somewhat worse in Experiment 2 than in the location-only condition of Experiment 1. This raises the possibility that the introduction of landmarks into a filing space may actually have an adverse effect on performance. This could occur, for example, if these landmarks effectively increased the crowding or clutter of the space so that retrieving the location of an object became more difficult.

It is perhaps not surprising that the two-dimensional desk sheet of Experiment 2, with its abstract, colorless representations of office furniture, is not much of an improvement over the blank desk sheet of Experiment 1. One could still argue that there is nothing wrong with location-based filing or the introduction of landmarks as a means of improving performance, but that the problem lies, instead, with the specifics of the implementations of this approach in Experiments 1 and 2.

For this reason, special interest is attached to the results of Experiment 3. In Experiment 3, a schematic two-dimensional space was replaced by an actual three-dimensional space. Although first-guess performance in Experiment 3 was better than in the location-only condition of Experiment 1, this difference is not significant, $F(1, 35) = 2.17$, $p > 0.150$. However, the interaction between the filing condition (Experiment 3 versus the location-only condition of Experiment 1) and the retrieval test is reliable, $F(2, 70) = 4.90$, $p < 0.010$. The decline of

[9] Since the three-guess data are identical in all important respects to the first-guess data, only the latter are presented for Experiments 2 and 3.

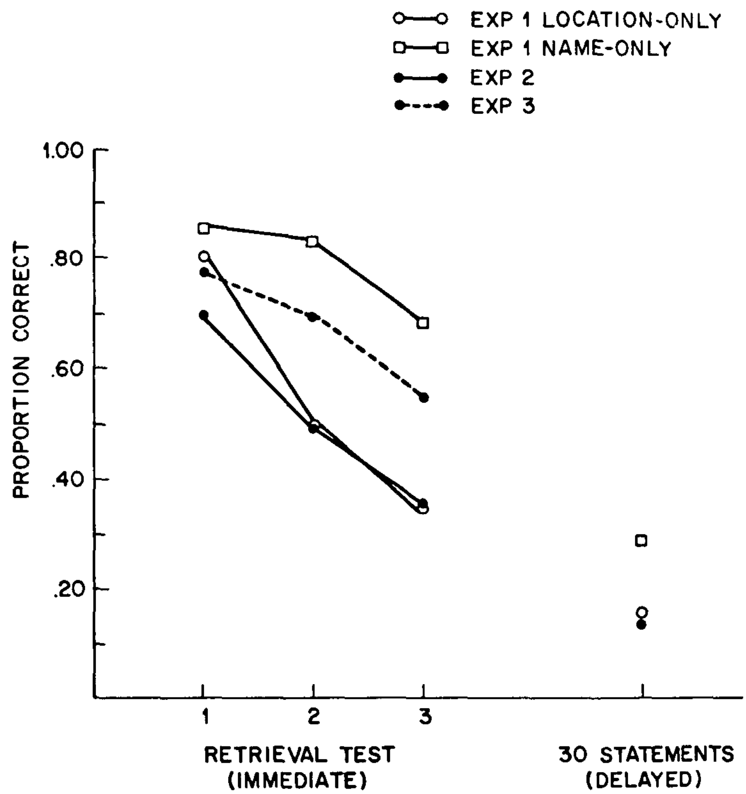

Fig. 8. First-guess accuracy (proportion correct) in Experiments 2 and 3. (Name-only and location-only performance from Experiment 1 is included for the sake of comparison.)

performance in the successive retrieval tests was less precipitous in Experiment 3 than in the location-only condition of Experiment 1, suggesting that an actual three-dimensional space may be more resistant to the effects of crowding. Nevertheless, first-guess performance in Experiment 3 was still not quite so good as in the highly impoverished name-only condition of Experiment 1. This difference is marginally significant, $F(1, 38) = 3.84, p < .057$. (The interaction between the retrieval test and the filing condition does not approach significance, $F < 1$.)

First-guess performance on first-block and most-recent-block statements is presented in Figure 9 as a function of retrieval tests in Experiments 2 (Figure 9a) and 3 (Figure 9b). In agreement with results from the location-only condition of Experiment 1, performance is worse for first-block statements than most-recent-block statements in both Experiment 2, $F(1, 28) = 17.02, p < .001$, and Experiment 3, $F(1, 19) = 5.28, p < .033$. Although the data in Figure 9 also suggest that performance on first-block statements is more adversely affected by the crowding manipulation, this interaction is not significant in either Experiment 2, $F(1, 28) = 1.18, p > .286$, or Experiment 3, $F(1, 19) = 1.63, p > .217$.

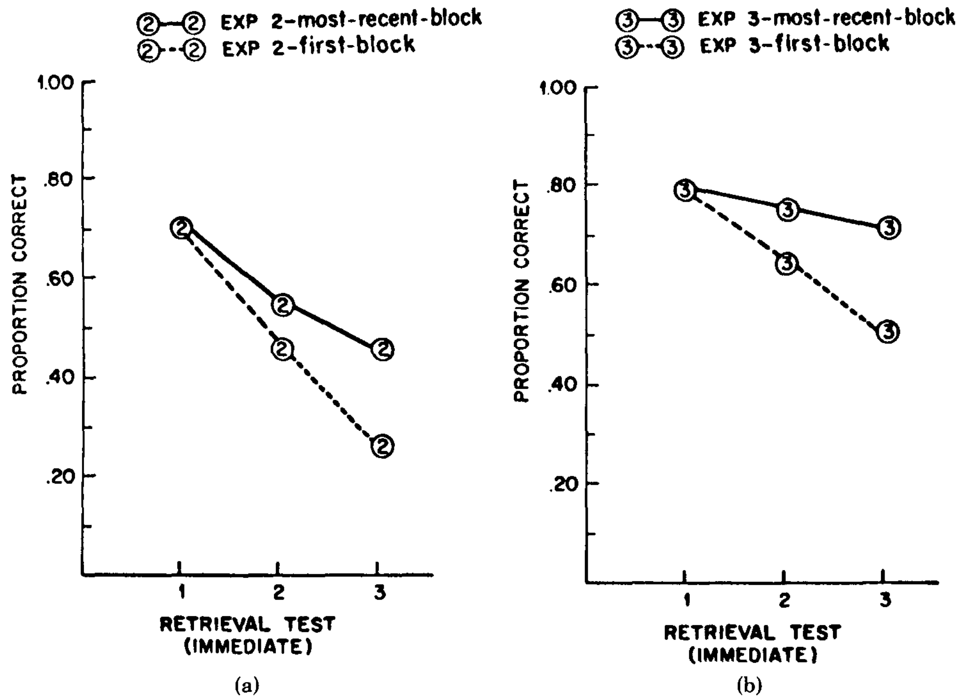

Fig. 9. First-guess accuracy in Experiments 2 and 3, separately plotted for first-block and most-recent-block statements: (a) Experiment 2; (b) Experiment 3.

4. CONCLUSION

The results of these experiments suggest that the efficacy of a spatially oriented approach to object reference in a computing system, whether this approach is used alone or in conjunction with symbolic reference, may be severely limited. Performance in the location-only condition of Experiment 1 was considerably worse than that in the name-only condition. Improvements obtained by combining symbolic and spatial information (compared with just symbolic information) were very modest.

In Experiment 2 subjects worked with desk sheets containing schematic drawings of office furniture. These were intended to serve as landmarks by which to anchor an object's location and also to heighten the office metaphor for storing information. This enhancement failed to produce any appreciable benefits. On the contrary, performance in the first retrieval test was somewhat worse than that in the location-only condition in Experiment 1. Although this difference is not significant, it raises the possibility that enrichments in a space, until familiarity with them has been attained, may actually crowd and interfere with memory for object location.

Any attempt, no matter how elaborate, to simulate a three-dimensional space such as an office setting in the fundamentally two-dimensional medium of paper or a CRT terminal screen is subject to the same criticism. Much more can and should be done, so the criticism goes, to increase psychological similarity between "virtual" and actual space; for example, the colors could be richer, the resolution greater, and tactile and kinesthetic sensation could be introduced. If actual space is held as an ideal to be more and more closely approximated with each new advance in technology, then actual space should be tested to assess the probable limits on performance that might be realized in the most realistic of computer-based, spatially oriented information systems.

This was done in Experiment 3 by putting subjects in an actual office setting within which articles were placed and retrieved. The adverse effects on performance of the crowding manipulation are less severe in Experiment 3 than in the location-only condition of Experiment 1. Nevertheless, the overall performance difference between these two conditions was not reliable. More important, performance in Experiment 3 remained below that observed in the severely constrained name-only condition of Experiment 1. The results of Experiment 3 cast serious doubt on the wisdom of any approach that wholeheartedly embraces a spatially oriented, "real-world" ideal.

In all filing conditions of the experiments, performance declined in successive retrieval tests. However, this decline was especially precipitous in the location-only conditions, and was more noticeable for retrieval statements dealing with less recently filed articles. This pattern suggests that a spatial filing system may be most useful as a temporary holding area or work space in which a small number of recently encountered objects can be stored. Such a system might possess some of the same advantages that are evidently realized in an actual office setting in which unlabeled "piles" are used [3, 11]. In particular, such a spatial system might permit users to defer the selection of an object's more permanent symbolic representation until more is known about the object and about the context(s) within which the object will eventually be used. It has been shown that such contextual knowledge can be an important determinant of the goodness (e.g., recallability, recognizability) of a user-selected representation [9]. The experiments described in this article also confirm an intuition, expressed by several subjects in Malone's investigation [11], that the retrievability of unlabeled objects in such a work space will rapidly decline with increases in the total number of unlabeled objects.

In a name+location-combined condition, it is also possible that spatiality might have some use as a means of more quickly "homing in" on a desired object. Such a use is better indicated by measurements of speed rather than by the accuracy measurements used in the experiments discussed in this article. This possibility notwithstanding, many of the promises of spatiality are given a miragelike quality by the experimental results discussed here; that is, when these promises are subject to closer scrutiny, they seem to fade away. It should be emphasized, however, that the results examined in this article generally complement rather than contradict those of previous psychological investigations of spatial memory. Previous findings indicate that incidental memory for location is better than chance, although the differences are usually modest [10]. The current experimental results indicate that intentional memory for location is considerably above chance, although modest when compared with memory for names. It also should be noted that incidental memory is by no means a unique property of spatial information. It has also been observed for symbolic information (e.g., words), where its magnitude can equal that observed under intentional conditions depending, in part, on the way in which the information is processed [8, 12].

It may be true that the formation of location memory in many situations is essentially effortless or, perhaps more accurately, that it is a gratuitous by-product of actions that must be performed in any case (e.g., head and eye movements that are needed to focus on an object). However, this is not necessarily the case in a computer filing situation. It takes time and effort to select an object's representation and to communicate this representation to the computer—whether this representation be a name or a location.

In a computer filing situation it is important to consider the trade-offs between the costs of specifying an object's representation (measured in terms of time, errors, inconvenience, etc.) and the benefits of this representation (measured in terms of memorability, descriptiveness, etc.). If users are forced to choose, our results indicate they will be better off with names than locations. (We assume that the cost of specifying a two-letter name roughly equals the cost of specifying a location.) Even in the absence of such a forced choice, the evidence suggests that an elaboration of a symbolic representation is better than a dual code consisting of spatial and symbolic information. As noted earlier, substantial gains were obtained in a pilot study in which subjects were simply allowed to select longer names. Moreover, experiments by Furnas et al. [6] suggest that even greater gains can be achieved when objects are referred to by more than one name.

The experiments of this article must be regarded as a first attempt to assess the role of spatiality in an information retrieval system. As such, the results do as much to suggest directions for further research as they do to provide any firm conclusions. In one direction, it might be worthwhile to pursue a finer grained analysis in which the *recallability* and *recognizability* of representations are separately considered. Any computer-based retrieval attempt is essentially an act of communication between a user and a computer system in which a link must be established between the user's representation of an object and the computer's corresponding representation of this object. When users simply *recall* the computer's representation of an object (e.g., when they recall the name originally given to that object), traversal of the link is primarily in one direction—going from the user's representation to the computer's. Near the other extreme are situations in which the user does not know or remember anything about how the object is represented in the computer. Users may then elect to scan all of the computer's object representations in the hopes that one will elicit their own internal representation of the desired object. When this occurs, traversal of the link is in the opposite direction and we can say that the computer's representation of an object has been *recognized* rather than recalled. From this perspective, the recall and the recognition of the computer's object representation differ primarily in the nature of the cues involved and in the direction of associative traversal. (See Tulving and Thompson [16] for a discussion of the essential cue-dependent nature of both recall and recognition.) Of course, in many situations the connection between user and computer representations of an object is formed through an interplay between the activities of recall and recognition. The user, for example, may recall that the computer's representation begins with an "r" or that it was originally placed in the upper right-hand corner of a filing space. The user can then use this partial recall of the computer's representation to restrict the number of alternatives that must be looked at during a recognition attempt.

In the context of the experiments discussed here, recall can be measured by a subject's ability to generate an article's representation (either its name or spatial location) in the absence of a filing sheet; recognition can be measured by the extent to which an article's representation (as this appears on the filing sheet) elicits an appropriate representation of the article in the user's own memory. Preliminary research (S. T. Dumais and W. P. Jones, unpublished research, 1984) confirms an intuition that symbolic representations are more easily recognized than spatial representations. An object's name, even if it is only two letters in length, seems to carry a great deal of information concerning the object's content or its purpose; an object's location appears to carry considerably less of the corresponding information. This research, however, also indicates that names are more easily recalled than are locations. That is, when filing sheets are taken away, names are easier to generate than locations.

In another direction of further research, it is important to learn more about the range and nature of computing situations to which the results given here might apply. On-line replications involving tasks that a user is likely to perform using computerized databases are clearly desirable. In an on-line experiment, it becomes possible to give subjects feedback concerning the contents of an incorrect choice. Such feedback is clearly important to the users of an actual filing system. However, there is no compelling reason to suppose that performance under different modes of object reference will be differentially affected by such feedback.

In replications and extensions of this manuscript's experiments, special attention should be given to two factors: (1) the nature of the objects to be stored and retrieved and (2) the nature of the cues that serve to signal the need for a given object in a given task situation. Neither of these factors was manipulated in the experiments reported here, and the choices that were necessarily made with respect to each factor are a source of caveats concerning the generalizability of the results.

Perhaps the objects (news articles) that were filed in the experiments do not especially lend themselves to a spatial organization. In this respect, they are perhaps not much different from a variety of other information objects (programs, datasets, memoranda, letters, etc.) that a user might want to store electronically. The possibility remains, however, that for certain materials (e.g., travelogue descriptions) a spatial representation may be preferable to a symbolic one.

In our experiments, objects were cued by statements that were themselves expressed symbolically. It is possible that, in the name conditions, some of these statements were "giveaways" in the sense that subjects did not have to rely upon their memory for an object's representation in order to generate the correct answer. For example, a word in a statement may sometimes have borne a considerable resemblance to the name of the appropriate object. Precautions were taken (names were only two letters in length, and statements contained no topic words from the articles they referenced) that probably reduced, but surely did not eliminate, the likelihood of such occurrences. In another followup study

(W. P. Jones and S. T. Dumais, unpublished research, 1984) we have found that subjects are indeed better than chance when they use someone else's two-letter names to guess the correct answer to a statement, even if they have never read the article. Although this accounts for only a small portion of the name to location advantage in Experiments 1–3, the problem of target specification clearly deserves further consideration. In future experiments we would like to provide subjects with a range of cues more closely approximating those that users encounter in the course of their interactions with a computer.

As an aside, we note that an interesting analogy can be drawn between giveaway statements (i.e., statements that can be answered in the absence of any memory for an object's name), and an ideal of computer–human interaction. When an object is adequately specified by the cues of a task environment, users need not rely on their own memories for the object's representation. The achievement of such an ideal may be especially important in the use of large public databases (where users cannot be expected to know the specific name or location of a desired object). The achievement of such an ideal will also aid new users who are unfamiliar with a system's idiosyncratic terminology.

In computer filing systems, we have the freedom to use whichever of many possible retrieval cues (shape, size, location, name, etc.) are most likely to lead to success. We need not (and should not) be bound by the same constraints that we face in retrieving the tangible objects in our world. The fact that we must remember *where* "the Smith account" is located in an office or where the milk is located in the grocery store in order to retrieve it does not mean that similar constraints should be imposed on the electronic offices of tomorrow. We therefore conclude by emphasizing the essential role that psychological experimentation must play in the development of computing systems. The increased sophistication and economic feasibility of computer graphics and of analog input devices create wonderful new opportunities to improve the interface between users and computers. However, to forego experimentation on the application of these newfound capabilities is to run the risk of indulging in a costly kind of technological trendiness in which new systems are implemented "because we can do it" and not as a consequence of their proved psychological utility.

ACKNOWLEDGMENTS

We would like to thank Susan Cook, Sharon Greene, Tom Landauer, and Steve Poltrock for helpful comments on earlier versions of the manuscript. Linda Blinn, Shari Fabri, and Julie Brady helped with various aspects of data collection and entry.

REFERENCES

1. ACREDELO, L. P., PICK, H. L., JR., AND OLSEN, M. G. Environmental differentiation and familiarity as determinants of children's memory for spatial location. *Dev. Psychol. 11*, 1975, 495–501.
2. BOLT, R. A. Filing and retrieving in the future—spatial data management. Paper presented at Infotech "User Friendly Systems" State of the Art Conference (London, 1979).
3. COLE, I. Human aspects of office filing: Implications for the electronic office. In *Proceedings of the Human Factors Society—26th Annual Meeting.* Human Factors Society, Santa Monica, Calif., 1982, pp. 59–63.

4. CRISTIE, J., AND JUST, M. A. Remembering the location and content of sentences in a prose passage. *J. Educ. Psychol. 68*, 1976, 702–710.
5. CROWDER, R. G. *Principles of Learning and Memory.* Erlbaum, Hillsdale, N.J., 1976.
6. FURNAS, G. W., LANDAUER, T. K., GOMEZ, L. M., AND DUMAIS, S. T. Statistical semantics: Analysis of the potential performance of key-word information systems. *Bell Syst. Tech. J. 262*, 1983, 1753–1806.
7. HASHER, L., AND ZACKS, R. T. Automatic and effortful processes in memory. *J. Exp. Psychol. Gen. 108*, 1979, 356–388.
8. HYDE, T. S., AND JENKINS, J. J. Differential effects of incidental tasks on the organization of recall of a list of highly associated words. *J. Exp. Psychol. 82*, 1975, 472–481.
9. JONES, W. P., AND LANDAUER, T. K. Context and self-selection effects in name learning. *Behav. Inf. Technol. 4*, 1985, 3–17.
10. LOVELACE, E. A., AND SOUTHALL, S. D. Memory for words in prose and their location on the page. *Mem. Cognition 11*, 1983, 429–434.
11. MALONE, T. W. How do people organize their desks? Implications for the design of office information systems. *ACM Trans. Off. Inf. Syst. 1*, 1983, 99–112.
12. MANDLER, G., AND PEARLSTONE, Z. Free and constrained concept learning and subsequent recall. *J. Verbal Learn. Verbal Behav. 5*, 1966, 126–131.
13. MANDLER, J. M., SEEGMILLER, D., AND DAY, J. On the encoding of spatial information. *Mem. Cognition 5*, 1977, 10–16.
14. ROTHKOPF, E. Z. Incidental memory for location of information in text. *J. Verbal Learn. Verbal Behav. 10*, 1971, 608–613.
15. SCHULMAN, A. I. Recognition memory and the recall of spatial location. *Mem. Cognition 1*, 1973, 256–260.
16. TULVING, E., AND THOMPSON, D. M. Encoding specificity and retrieval processes in episodic memory. *Psychol. Rev. 80*, 1973, 352–373.
17. VON WRIGHT, J. M., GEBHARD, P., AND KARTTUNEN, M. A developmental study of the recall of spatial location. *J. Exp. Child Psychol. 20*, 1975, 181–190.
18. YATES, F. A. *The Art of Memory.* University of Chicago Press, Chicago, Ill., 1966.
19. ZECHMEISTER, E. B., AND MCKILLIP, J. Recall of place on a page. *J. Educ. Psychol. 63*, 1972, 446–453.
20. ZECHMEISTER, E. B., MCKILLIP, J., PASKO, S., AND BESPALEC, D. Visual memory for place on the page. *J. Gen. Psychol. 92*, 1975, 43–52.