Order Number 1345286

Information displays: The effects of organization and category distinctiveness on user performance

Halgren, Shannon Lee, M.A.

Rice University, 1991

U·M·I
300 N. Zeeb Rd.
Ann Arbor, MI 48106

RICE UNIVERSITY

INFORMATION DISPLAYS: THE EFFECTS OF ORGANIZATION AND CATEGORY DISTINCTIVENESS ON USER PERFORMANCE

by

SHANNON LEE HALGREN

A THESIS SUBMITTED
IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE
MASTER OF ARTS

APPROVED, THESIS COMMITTEE

Nancy J. Cooke, Ph.D., Director
Assistant Professor, Department of Psychology

Ken Laughery, Ph.D.
Professor, Department of Psychology

Henry Roediger, Ph.D.
Professor, Department of Psychology

Houston, Texas

December, 1990

INFORMATION DISPLAYS: THE EFFECTS OF ORGANIZATION AND CATEGORY DISTINCTIVENESS ON USER PERFORMANCE

by
Shannon L. Halgren

Abstract

The goal of this research was to test the effect of display organization on user performance under a situation representative of non-experts' interactions with an online display. Alphabetical, categorical, and random organizations were tested for response time and accuracy on a visual search task (Experiment 1) and on a problem solving task (Experiment 2). Term or definition targets were searched for in displays consisting of items from distinct or overlapping categories. Performance with alphabetical and categorical organizations was similar when targets were terms and categories were distinct, however, these conditions are atypical of non-experts' interactions. Categorical organizations were superior when task difficulty increased. Surprisingly, overlapping categories resulted in decreased accuracy with alphabetical organizations relative to the distinct category conditions, whereas, performance with categorical organizations remained unaffected. This result and evidence suggesting that the individual display items influence how these factors affect performance have implications for interpreting past display organization research.

Acknowledgements

I would like to thank several people who helped make this thesis possible. First and foremost thanks must go to my advisor, Nancy Cooke. Nancy was always available to solve problems, provide emotional support, and read the seemingly endless series of drafts that accompanied this thesis. I would also like to thank my committee members, Henry Roediger and Ken Laughery for their valuable comments which strengthened this thesis. Finally, I could not have completed this work without the encouragement and support provided by my family and best friend, Eric.

Table of Contents

List of Tables

List of Figures

Figures | Page

Displays of information are found in an unlimited number of settings. Common interactions with information displays include using the Yellow Pages to look up a number or address, using an online card catalog to locate a book or journal, and using a dictionary to look up the spelling of a word. Information displays can also contain a smaller number of items such as a grocery list or the menu of options on an automatic-teller machine. Regardless of display size, the way that people use displays to access information has intrigued cognitive psychologists and recently those who study human-computer interaction (e.g., Card, 1982; McDonald, Stone & Liebelt, 1983).

Within the last decade menu-driven computer systems have become increasingly popular among the growing number of novice and casual users (Card, 1982; Dumais & Landauer, 1984; Giroux & Belleau, 1986). Menu systems appeal to this population because they only require the user to recognize commands, rather than recall them from memory as is required by command-based systems. However, menu-driven systems are only as successful as the information displays or menus that they contain.

Which elements of a menu or information display affect how quickly and accurately an item can be found within it? Despite the fact that there have been numerous laboratory investigations designed to answer this question, it appears that there is still some uncertainty about its answer (Paap & Roske-Hofstrand, 1986). Human-computer interaction psychologists were not the first scientists to discover search performance is complex and dependent on several variables. Cognitive psychologists have also sought to explain and predict search performance (DeRosa & Tkacz, 1966; Naus, Glucksberg, & Ornstein, 1972; Neisser, 1963, 1964; Sternberg, 1966).

Search performance of both memorized and visually presented lists of items have been studied in the domain of cognitive psychology. Results from both types of investigations are useful to the human-computer interaction psychologist because they provide clues about the factors that may influence search of computer displays.

Memory Search

Numerous studies investigating search of memorized lists have found a number of factors that influence performance (DeRosa & Tkacz, 1976; Lachman, Lachman & Butterfield, 1979; Sternberg, 1966). One of the pioneers in memory search was Saul Sternberg (1966) who found that length of a memorized list affects the amount of time subjects need to report whether a target item is present in the list. Subjects who memorized a list of six digits took longer to search their list for a target than subjects who memorized shorter lists of digits.

Search of a memorized list is also influenced by list organization (DeRosa & Tkacz, 1976). Subjects studied sets of pictures that told a short story when arranged sequentially. Each set was arranged either randomly or sequentially. Subjects were then presented with a probe item and asked if it was in the studied set. There was no relation between set size and reaction time in the sequential organization conditions, whereas a linear function was present between these two variables in the random organization condition. This suggests that subjects who memorized the set arranged randomly searched their mental set serially; subjects who memorized the set arranged sequentially searched their mental set in parallel. Other investigators have concluded that when searching memorized word lists that comprise a single category for a probe item, search is serial and exhaustive. When words comprise two or more categories however, the search is serial, but will end after the probe's category has been exhaustively searched (Naus et. al, 1972).

Visual Search

Interest in visual search stemmed from the work of Neisser (1963) who investigated various factors that affect visual search. Some of Neisser's work sought to explore how perceptual features influence search time. In one of his studies (1964) subjects were asked to search for a target letter (Z) in a list of letters which were either perceptually similar (e.g., W, X, E) or dissimilar (e.g., O, Q, D) to the target. The list containing letters with similar features to the target took more time to search than the list containing letters with dissimilar features.

Semantic features affecting a visual search task have also been of interest to cognitive psychologists. Several studies were conducted to investigate the influence of categorization on visual search. A typical visual search paradigm involves asking subjects to find a digit or letter target located in an array of digits or letters. Results have consistently indicated that a between category search (looking for a digit among letters or a letter among digits) is faster than a within category search (looking for a digit among digits or a letter among letters (Jonides, & Gleitman, 1976)). The arrangement of the numbers and digits in the array influences visual search as well (Krueger and Hettinger, 1984). Subjects asked to find a target in a list of numbers or digits arranged regularly (in sequential order) searched faster than when items were arranged randomly.

Investigators have found several other factors which influence visual search as well. Some of them are: number of elements in an array (Shiffrin & Schneider, 1977), the consistency of mapping between targets and distractors (Shiffrin & Schneider, 1977), target location within a display (Hammond &

Green, 1982; Neisser, 1964), and the type of display (letter vs. shape) being searched (Hammond & Green, 1982). Although the results from these studies have implications for the human-computer interface, they do not directly address some of the factors of particular importance to this application.

The goal of most display research in human-computer interaction is to determine the array or display design which results in optimal search; whereas, the primary goal of search studies in cognitive psychology (both memory and visual) is to understand the human information processing system. The latter goal results in interesting implications for human-computer interaction, however, there are many questions left unanswered. For example, how search performance is affected by larger displays. Search studies in cognitive psychology often incorporate displays of only a few items (i.e., Shiffrin & Schneider, 1977), whereas human-computer interactionists are often interested in much larger displays. Therefore, it is necessary to conduct research which will explicitly test various factors specific to computer displays. Some of the factors human computer-interaction psychologists are interested in manipulating include: novice and skilled user performance, explicitness of the item for which the user is searching, category distinctiveness, display size, and display organization. More importantly, the complex interactions of these factors are of particular importance in this domain. In addition, whereas visual search is interesting and certainly involved in computer display interactions, the use of displays is not limited to the visual search task.

Numerous studies investigating computer displays have already been conducted. Results of these studies have indicated several variables which influence search of computer displays and thus, should be attended to in display design.

Computer Display Search

Organization. Experiments manipulating computer display organization have consistently found significant effects of organization on search time and accuracy (Card, 1982, 1984; McDonald, Stone & Liebelt, 1983; Perlman, 1985; Vandierendonck, Van Hoe, & De Soete, 1988). Vandierendonck, et al. (1988) found that categorical organizations afford faster search than random organizations. They concluded that subjects search computer menus based on knowledge. Eye-movement analysis and reaction times suggested that search is not random, rather "knowledge facilitating factors" such as listing items in a categorical arrangement allowed the subjects to begin their search in close proximity to the target, thereby reducing search time. This explains why displays arranged alphabetically have also been found to facilitate visual search (Card, 1982, 1984; Perlman, 1985). Most users are knowledgeable about alphabetical organization and are able to use this arrangement to narrow their search.

Most psychologists and software designers agree that the items in a menu or display, should be arranged in some sort of logical order, especially when the display contains a large number of items. The organization should be obvious and consistent in order to allow users to anticipate where items are located (Dumais & Landauer, 1984; Sisson, Parkinson & Snowberry, 1986; Teitelbaum, & Granda, 1983). In addition, the display should be congruent with the type of task the user will be performing (Bødker, 1989; Boehm-Davis, Holt, Koll, Yastrop, & Peters, 1989; Miller, 1984) . These general guidelines were created in response to research which suggested that display organization influences speed and accuracy of a user's interactions with the display.

Organization of items within a computer display is not the only factor that influences visual search, but has been found to interact with several other factors such as target type. These other factors must also be taken into account in display design.

Target Type. Depending on the users' task and expertise, they will search a display for one of several types of targets. These targets can be broken down into two types: explicit targets and implicit targets. Users will be searching for an explicit target when they know exactly what word they are looking for in the display. For example, when a person is searching for the word *cut* and the word *cut* exists on the display somewhere as opposed to a synonym such as *delete* or *kill*, the target is explicit. Search for explicit targets will usually occur when the user is experienced and has a complete and accurate understanding of the system.

When users don't know the exact name of the menu item for which they are searching, the target is considered implicit. This often occurs when the user is new to a system and has a goal in mind, but doesn't know the exact name of the command needed to achieve the goal (McDonald et. al., 1983).

Card (1982) has studied how search time is affected by menu organization when the subject is searching for an explicit target. Subjects were presented with a text editing command and were asked to find it in a list of 18 vertically arranged commands, then to select it with a mouse. The commands were arranged either alphabetically, functionally or randomly. The functional arrangement was similar to a categorical arrangement -items related functionally were located in close proximity to each other. Mean search times favored alphabetical arrangement, followed by functional and random arrangements, respectively. Because subjects were given the exact word to locate they merely had to scan the menu to find the target. The alphabetical arrangement allowed the subjects to quickly narrow their search. Subjects in the functional condition found the target faster than the subjects in the random condition indicating that subjects were beginning to learn the functional organization which allowed them to save some time. Despite the fact that the functional group performed poorly in comparison to the alphabetical group, studies of implicit targets have found that in situations in which users must make a decision about the most suitable menu item, a categorical organization is superior (Hollands & Merikle, 1987; McDonald et al., 1983).

Studying implicit targets is important not only because the type of target appears to interact with display organization, but also because these types of targets are thought to provide more "ecological validity." It is more often the case that menu-based interfaces are used when recall is imperfect, as opposed to command-based interfaces which require more exact knowledge of the system. In addition, for the novice or occasional user it is most likely that the exact name of an item is either not known or not remembered. This may also happen to an experienced multiple system user or a casual user (McDonald et al., 1983). Thus, because a significant portion of the population of menu users are not experienced system users, it is important to create conditions in the lab that reflect the situations that they typically encounter.

A study was conducted by Hollands and Merikle (1987) which tested menu organizations using both explicit and implicit targets. Subjects were given either a psychology term or a definition of a psychology term as the target. For novices in psychology, an explicit term was found faster in a list of terms arranged alphabetically than in a list of terms arranged categorically. For psychology experts in the same target condition, categorical and alphabetical arrangements produced similar search times. When subjects were provided with an implicit definition target, the categorical arrangement produced superior performance for both experts and novices.

The above study suggests that the type of target (term or definition) as well as the user's experience with the items, has a direct effect on which display organization optimizes search. This is further supported by McDonald, Stone and Liebelt (1983). In this study, five menu organizations were compared using menus of 64 items belonging to one of four natural categories (animals, food, cities, and minerals). Alphabetical, random and three categorical organizations were tested. The categorical organizations consisted of items arranged categorically with members of the same category placed in the same column. Items in the categorical columns were further arranged either alphabetically, randomly or categorically. Subjects either searched for an explicit target in an identity matching task or searched for an item corresponding to a single line definition in an equivalence matching task.

The explicit matching task resulted in faster searches than the definition matching task across menu organization. The menu organization effect was significant (categorical-categorical resulting in fastest judgements and random resulting in the slowest reaction judgements) and was found to interact with target type. When searching for definition targets, the categorical-categorical organization produced the fastest search times; however, this condition was not significantly faster than the other two categorical conditions. All three categorical organizations were significantly faster than the alphabetical arrangement. In the explicit target conditions all three categorical conditions were equivalent in search time to the alphabetical organization.

The above findings suggest that when the user does not know the name of the target, categorical organizations will result in the fastest selection times. However, when the user knows the name of the target, alphabetical organizations are equally or more effective than categorical organizations. The explicit condition was similar to Card's study in which the alphabetical organization was found to be superior to the categorical arrangement, and to Holland and Merikle's explicit condition in which alphabetical arrangements were superior to categorical arrangements for novices, but equal for experts. In McDonald et al.'s explicit condition, alphabetical and categorical arrangements produced similar reaction times and in the implicit condition categorical arrangements were superior. These inconsistent results might be due to the distinctiveness of the categories used. When categories are distinct as in McDonald's study (food, animals, cities, minerals), or well-learned as they were for the experts in Holland and Merikle's study, then it is likely that subjects immediately identify the categorical arrangement and have little trouble selecting the appropriate category under which to begin search. As a result, the advantage of an alphabetical over a categorical organization would not be as great for distinct or well-learned categories. In Card's study subjects may have had more problems identifying the functional similarity between the text editing commands which limited the usefulness of the categorical arrangement. Thus, the benefits of categorical organizations seem to be tied to the meaningfulness or distinctiveness of the categories to the subjects.

Category Distinctiveness. In the studies discussed above alphabetical organizations facilitated search over categorical organizations in some cases (Card, 1982; Hollands & Merikle's novice condition,1987), but not others (Hollands & Merikle's expert condition,1987; McDonald et. al., 1983). These inconsistencies can be explained by category distinctiveness. It is possible that when distinct, semantically clear, categories are used, categorical arrangements may be utilized to reduce search time when the target is a definition. However, when there is some conceptual overlap between categories, it is likely that the subject's ability to use the categorical arrangement to limit their search will be impaired (Card, 1984; Giroux & Belleau, 1986; Hollands & Merikle, 1987; Mehlenbacher, Duffy & Palmer, 1989; Paap & Roske-Hofstrand, 1986; Schwartz & Norman, 1986). That is, categorical organizations should be successful to the extent that they effectively group the items and describe the categories so that users are able to anticipate their contents (Dumais & Landauer, 1984). A categorical organization should be ineffective when items can be initially placed in more than one category and as a result, the organization will not be useful until the categories and instances are learned. The more distinct the categories used (McDonald et. al., 1983) or the more experience the user has with the categories (Hollands & Merikle, 1987), the more a categorical organization should facilitate search in comparison to an alphabetical organization.

Of course, this argument seems rather obvious at one extreme. With extremely poor or overlapping categories search should be comparable to a random organization. However, at intermediate levels of categorical distinctiveness the benefit over alphabetical organization is less clear and requires empirical support. The effect of category distinctiveness and the type of target is also unclear. One might expect that categorical organizations are particularly helpful when the target is implicit, although less so as categorical distinctiveness decreases.

It is important for research in areas such as human-computer interaction which directly address practical questions, to recreate situations in the lab that are closely matched to everyday experiences. In order to achieve a realistic recreation of the "everyday" it is necessary to make the distinction between distinct and overlapping categories. Natural categories that are encountered in normal conversation and problem solving are not distinct but overlap a great deal. Membership in natural categories is not a dichotomous all-or-none question, but rather one of degree (McCloskey & Glucksberg, 1979; Rosch, 1978). In addition, natural categories can vary over time, people, and tasks making them even less distinct (Dumais & Landauer, 1984). Finally, distinct categories are not the norm for interaction with computer systems, unless the user is an expert and a regular user of the system.

On several occasions of menu use it is likely that the categories will not be as semantically clear to the user as the categories were in McDonald et al.'s study. For example, the novice or occasional user might not know which categories to search when looking for a *save* or *quit* command. This was suggested by Card's (1982) study in which categorical arrangements were slower than alphabetical arrangements when searching for explicit text editing commands. It is also suggested by the categories used in actual systems in use today. Microsoft applications, such as Word, list commands under pull-down menus. Organization of the commands under the various menus is determined by category. Commands of similar function are placed under the same menu. However, the categories used (i.e., file, edit , format, utilities) are not usually intuitive to novices and can even be quite fuzzy to experienced users. As a result, a large amount of practice is required to use the categorical arrangement to effectively facilitate search. Because menu items describing actual databases are often unfamiliar to novices, any categorical arrangement should take some time to learn. This learning phase is suggested by the results of Hollands and Merikle (1987) in which novices took longer to find an explicit target in a categorical arrangement than in an alphabetical arrangement, whereas experts took the same amount of time for both arrangements.

Distinct categories should be learned faster than overlapping categories. Mandler and Pearlstone (1966) have suggested that subjects impose a categorization on the display before all the instances of the array have been inspected. If this hasty categorization is different from the categorization definition in the system design, it will have to be overcome before the organization can be utilized to narrow search. Subjects initial categorization will most likely be closer to the system's organization when categories are distinct because there is less room for error or variability. It is unlikely, on the other hand, that subjects using a categorized display of overlapping categories will identify the rules for category membership accurately on the first or even second attempt.

In summary, research has indicated that both organization and target type are important to visual search. Inconsistent results of these studies might be explained by distinctiveness of categories used in the organization conditions. The goal of Experiment 1 was to investigate how category distinctiveness interacts with organization and target type to influence search time and accuracy on a computer display. Subjects were asked to search for either explicit or implicit targets in a display arranged either alphabetically, categorically or randomly. The items within each display comprised either distinct or overlapping categories.

Generation of categories

Before the first experiment was conducted a short norming study was run in order to generate the categories used in the distinct and overlapping category conditions. A category was considered distinct if all or nearly all of its items were judged to belong only to that category. A category was considered overlapping, on the other hand, if several items were judged to belong to that category and also to one or more other categories. The goal of this study was to generate four distinct categories of 16 instances and four overlapping categories of 16 instances. In addition, two of the categories were to serve in both the distinct and overlapping conditions, thereby controlling for item differences. This is possible because category distinctiveness is affected by the alternative categories in each set.

A card sorting technique was used to measure semantic distance between the categories and their instances. This technique for determining how lexical information might be stored and organized in memory was validated by Miller (1969).

Twelve Rice undergraduate students participated in this study to satisfy course requirements. Data from two subjects were discarded because English was the second language for these subjects and they therefore had difficulty completing the task. Subjects were given two piles of 100 3 x 5 inch index cards. The name of each item from hypothesized distinct categories was printed on the cards from one pile and the name of each item from hypothesized overlapping categories was printed on the cards from the other pile. The distinct categories pile contained 20 items from each of the following categories: tools, furniture, musical instruments, clothing, and vehicles. The overlapping categories pile contained 20 items from the following categories: clothing, vehicles, games, sports equipment, and camping equipment. (Note: Although only four categories of 16 instances were needed for each category conditions, additional items and categories were included so that the most overlapping or distinct items could be selected from a larger pool). Several of the categories and their instances were taken from Battig and Montague (1969) and Hunt and Hodge (1971), two studies presenting normative data describing common categories and their most typical instances. The categories and instances of games, sporting equipment , camping equipment, and clothing were developed by the experimenter.

The two piles were shuffled separately and labeled either pile one or pile two. Subjects were instructed to sort each pile of cards into smaller piles representing categories. They were free to choose the number of piles to create and the definition of each pile. Several blank index cards were given to subjects and they were encouraged to make duplicate cards in order to add an item to more than one category if necessary. When subjects were satisfied that their cards were sorted, the experimenter asked them to double check their piles to confirm that duplicate cards were made for items that belonged to more than one category. Subjects were then asked to label each of their piles. The order in which the two category piles were sorted was counter balanced across subjects.

A correlation matrix of the frequencies that two items were sorted together across subjects was created for each condition. The matrices were submitted separately to an overlapping cluster analysis using the OVERCLUS procedure in SAS. This procedure forces the items into a set of clusters representing items that were most frequently grouped in the same category. Items could belong to more than one cluster.

The procedure created twelve clusters for each of the conditions. Five of the clusters from the distinct categories condition were easily identified as the five hypothesized categories (furniture, vehicles, clothes, musical instruments, or tools). They each contained only the 20 items from the category, with the exception of the vehicles cluster which contained the word "plane" from the tools category (this item was thrown out). The remaining seven clusters were subsets of the five larger clusters.

The clusters from the overlapping categories condition were not as easy to identify. There seemed to be five clusters representing the five hypothesized categories (clothing, vehicles, games, sporting equipment, camping equipment), however they all included one or several additional items with the exception of the games category which was ultimately thrown out because it was too distinct from the other categories. Of the remaining seven clusters, four could be described as subsets of the five categories mentioned above. The other three would best be described as new categories developed by the subjects. It was not easy to label these new categories or to determine the rule for cluster (category) membership.

To determine which category to eliminate from each group of five categories, the frequency of category overlap was calculated. Ideally, the most overlapping category would be eliminated in the distinct category condition and the most distinct category would be eliminated in the overlapping category condition. In the distinct category condition the only two categories to ever overlap within the same cluster were the clothing and vehicle categories (two of the twelve clusters which could best be described as summer items (e.g., shorts, swimsuit, bicycle) and winter items (e.g., scarf, ice skates, skis), contained instances from both categories). These two categories were kept, however, because they overlapped with both the sporting equipment and camping equipment categories in the overlapping category condition and thus could serve as controls for item differences. None of the other three distinct categories overlapped with another category. Therefore, musical instruments was arbitrarily deleted. In the overlapping condition, the games category was deleted because only two of its instances were placed in another cluster.

Four instances from each category were also deleted. Instances were systematically removed until the combination was found that made the categories maximally distinct or overlapping. The four categories and their instances in each condition are listed in Appendix A.

After the appropriate categories and instances were deleted, several calculations were performed to verify that distinctiveness was indeed different for the two conditions. These calculations are summarized in Table 1.

The number of instances across all the clusters in the distinct category condition was 105 compared to 188 in the overlapping category clusters. This suggests that more duplicate cards were made while sorting the overlapping categories pile than when sorting the distinct categories pile (125 vs. 41). Several of these duplicates were actually the same item occurring in several piles. When repetitions are not counted, the actual number of items placed in more than one cluster is reduced to 29 in the distinct categories condition and 52 in the overlapping category condition.

Although some items occurred in more than one cluster, some categories were labeled similarly by subjects or could not be easily identified as one of the four categories. Thus some overlap may be overestimated, and in order to take this into account, each cluster was labeled with one of the four categories

Table 1

Summary of Results of Category Distinctiveness Analysis

Analysis Description	Number of Occurrences in Condition	
	Distinct Categories	Overlapping
Categories		
Total number of instances across all clusters[a]	105	188
Number of times an item occurred in >1 cluster	41	125
Number of items in >1 cluster	29	52
Number of times a category instance occurred in a cluster of a different category name	4	66

[a]The number of instances is greater than 64 (the number of separate items) because items could be placed in more than one cluster.

names. The category name selected was the category to which the majority of instances within the cluster belonged. The number of times an item occurred in a different cluster according to these more stringent requirements was calculated. This type of overlap occurred 4 times in the distinct categories condition compared to 66 times in the overlapping categories condition.

Finally, the overlap between each pair of categories was examined separately. The number of times an item from a cluster representing one category occurred in a cluster representing another category was calculated. The results of this analysis are summarized in Table 2. As mentioned earlier, only the clothing and vehicles categories overlapped in the distinct categories condition. However, amount of overlap between these two categories in the overlapping category condition is much greater (114 times vs. 32). Thus the context of the alternative categories affected judgements of distinctiveness between categories.

The results of the above analyses suggest that the distinct categories overlap with each other much less than the overlapping categories. The clusters that emerged in the cluster analysis verified the original assignment of items to categories and the hypothesis that the two sets of categories differed in terms of distinctiveness.

Table 2

Overlap Between Each Category

Category Condition	Categories A	Categories B	Overlap[a]
Distinct	Clothes	Vehicles	32
	Clothes	Furniture	0
	Clothes	Tools	0
	Vehicles	Furniture	0
	Vehicles	Tools	0
	Furniture	Tools	0
Overlapping	Clothes	Vehicles	114
	Clothes	Camping Equipment	227
	Clothes	Sporting Equipment	159
	Vehicles	Camping Equipment	207
	Vehicles	Sporting Equipment	340
	Camping Equipment	Sporting Equipment	253

[a]Overlap equals the number of times an instance from category A was occurred in the same cluster as an instance from category B.

Experiment 1

The goal of this experiment was to determine what effect, if any, category distinctiveness has on visual search of information displays. Although the items and categories used in this display were not taken from any system currently in use, subjects in this study were exposed to more realistic display conditions compared to subjects from previous studies. It was assumed that overlapping rather than distinct categories and definitions rather than explicit targets, are features that are more representative of a novice or casual user's interactions with an online information display.

Display organization (alphabetical, categorical, and random), target type (term and definition), and category distinctiveness (distinct and overlapping) were examined. Two predictions were made based on the above literature review. First, in the distinct categories condition, results from the studies discussed above should be replicated. When targets are explicit terms, no difference should be found between search times of alphabetical and categorical organizations (McDonald et. al, 1983; Holland & Merikle, 1987, expert condition). When targets are specified implicitly by definitions, a categorical organization should facilitate search (McDonald et. al, 1983; Holland & Merikle, 1987, both expertise conditions).

According to Paap and Roske-Hofstrand (1986), when categories are overlapping the categorical arrangement should not be as beneficial as when categories are distinct. The conceptual overlap of the categories used in the categorical organization should make it more difficult to use the arrangement to limit the search as several of the items might theoretically fit under more than one category. Consequently, the second prediction is that when categories are overlapping and targets are terms, alphabetical organizations should produce faster initial search times. This result would concur with previous studies (Card, 1982; Holland & Merikle, 1987, novice condition). When targets are definitions, the advantage of the categorical organization found in the distinct category condition should be lessened in the overlapping category condition. In this condition, performance with categorical arrangements should approach, if not equal, performance with alphabetical arrangements.

Method

Subjects. A total of 62 Rice University undergraduates participated in this study. Two subjects' data were unusable due to the occurrence of system errors during their session. The students all received course credit for their participation.

Design. The design was a 2 x 2 x 3 x 2 mixed factorial design: category distinctiveness (distinct vs. overlapping), target type (definition vs. term), organization (alphabetical, categorical and random), and blocks (one and two). Category distinctiveness and organization were between subject variables whereas target type and blocks were within subject variables.

Materials and Apparatus. As discussed earlier, the items in the distinct condition comprised the four categories clothing, vehicles, tools, and furniture and the items in the overlapping condition comprised the categories clothing, vehicles, sporting equipment, camping equipment. Each category consisted of 16 items. The items were arranged in four vertical columns of 16 items each. Menu organizations were created by listing the 64 items alphabetically, categorically without headings, or randomly. For the categorical organization, the words were arranged randomly within each category.

Definition targets were derived form Webster's New World Dictionary and Webster's Dictionary (a pocket dictionary). If items were not found in either dictionary (e.g., running shoes and sweat pants) or if the words had definitions of over one line, new definitions were created by the experimenter (See Appendix B for a listing of the term and definition targets).

Target presentation and response collection was done with four Macintosh Plus computers. The experimental program was written in HyperCard.

Procedure. Subjects were randomly assigned to category type (distinct vs. overlapping) and organization (alphabetical, categorical, or random). On each trial subjects were either shown a term as the target, or were given a one-line definition of the target. Half of the subjects received terms as the first 32 targets and definitions as the second 32 targets and half received definitions first and then terms. This sequence of 64 trials was repeated, in a different order, in the second block.

Instructions explaining the task were written on the computer screen and each subject was given 10 practice trials (5 term targets and 5 definition targets). Subjects were encouraged to ask any questions that they had about the task before beginning the experimental trials.

Subjects began the program by selecting an OK button located in the center of the screen with the mouse. The target was then displayed below the OK button and remained on the screen for approximately one second (term target) or four seconds (definition target). According to pilot subjects, presentation of the definition target needed to be at least four seconds in order to read the entire definition. If the subjects had seen the term target for four seconds as well, they may have time to plan their move, consequently giving an unfair advantage to this condition.

After the target disappeared, the screen displayed the menu of 64 items arranged either alphabetically, categorically or randomly depending on the subject's condition. Subjects were then required to find the target in the menu. When the item was found, subjects were instructed to move the cursor to the target word and select it using the mouse. If a non-target word was selected, the word ERROR was displayed on the screen for approximately one second, after which the next trial was performed. At the end of each block of 64 trials the subjects were told the number of correct selections they had made. Subjects searched for each of the 64 targets to complete a block of trials. Each subject performed two blocks. The order of the targets was randomized across subjects and blocks. Prior to debriefing, subjects were given a questionnaire to determine if there were any special difficulties encountered and to identify what display organization, if any, the subject detected.

The performance variables measured were search time and accuracy. Search time consisted of the time from target presentation to the time the subject selected a word in the 64-item display.

Results

Several analyses were performed on the error-free search time and accuracy data from Experiment 1. The total search time and accuracy data for all trials were first submitted to an analysis of variance procedure. Because most effects in the first block decreased or disappeared completely in the second block, the trials from the first block were then analyzed separately. Finally, the trials from the two categories used in both category distinctiveness conditions were analyzed alone to determine what effects, if any, were present after controlling for item variance. This analysis of identical item trials is the focus of the Discussion section.

Practice Effects. The search time and accuracy means for each condition are listed in Table 3. Error-free search time and error data were each submitted separately to a 2 x 2 x 3 x 2 analysis of variance. Each of the four variables yielded significant main effects for search time. Term targets were found faster than definition targets (F (1, 54) =86.91, p<.0001), and targets were found faster when items were from distinct rather than overlapping categories (F (1, 54) =6.20, p<.05). Alphabetical and categorical organizations facilitated search over random organizations, but were not significantly different from each other (F (2, 54) =39.98, p<.0001; Tukey: st (54) = 3.4, ∂ = .05). Finally, search times were faster in the second block than in the first (F (1, 52) =133.12, p<.0001).

A significant interaction was found between target type and category distinctiveness (F (1, 54) =4.71, p<.05). Category distinctiveness only affected search time when targets were definitions (st (54) = 3.74, ∂ = .05). In addition, the effects of target type, organization and category distinctiveness each interacted with block- effects were much larger in early trials than in later trials (although for organization and target type they were still significant in the second block (F (1, 51) =48.63, p<.0001, Tukey: st (57) = 3.74; F (2, 52) =3.25, p<.05, Tukey: st (59) = 4.17; F (1, 52) =4.52, p<.05, Tukey: st (59) = 3.74, ∂ = .05, respectively)).

The three-way interaction between target type, category distinctiveness and block was also significant (F (1, 51) = 5.08, p<.05). A post hoc Tukey procedure (st (57) = 4.45, ∂ = .05) indicated that the significant interaction of target type and category distinctiveness found in the first block disappeared in the second block of trials. All other interactions were not significant.

The error data for these same conditions were in the same directions as the search time data described above, although several of the effects failed to

Table 3

Error-Free Search Time (seconds) and Percentage of Errors from all Block 1 Trials

BLOCK 1

Category Distinctiveness:	Distinct			Overlapping			
Organization:	Alphabetical	Categorical	Random	Alphabetical	Categorical	Random	Means
Target Type:							
Term	3.38 (0.31)	4.04 (0.00)	6.97 (0.00)	3.56 (0.31)	5.10 (0.00)	6.90 (0.63)	5.02 (0.21)
Definition	5.61 (9.06)	5.85 (6.56)	9.23 (14.96)	7.98 (10.31)	7.79 (5.31)	10.03 (7.50)	7.75 (8.95)
Means:	4.56 (4.69)	4.95 (3.28)	8.10 (7.48)	5.77 (5.31)	6.45 (2.66)	8.47 (4.07)	
Means:		5.89 (5.24)			6.90 (4.01)		

Figure 3, Continued

BLOCK 2

Category Distinctiveness:	Distinct			Overlapping			
Organization:	Alphabetical	Categorical	Random	Alphabetical	Categorical	Random	Means
Target Type:							
Term	3.03 (0.63)	3.37 (0.31)	5.28 (0.35)	3.07 (0.00)	3.49 (0.31)	5.13 (0.00)	3.89 (0.27)
Definition	3.86 (2.19)	4.15 (2.50)	5.55 (4.51)	4.63 (4.51)	3.40 (3.13)	5.74 (3.13)	4.64 (3.33)
Means	3.45 (1.41)	3.76 (1.41)	5.41 (2.43)	3.85 (2.26)	3.73 (1.72)	5.43 (1.57)	
Means		4.17 (1.75)			4.36 (1.85)		

Note. Error percentages are listed in parentheses

reach significance. As with the search time data, the main effects of target type and block were significant (F (1, 54) = 122.66, p<.0001; F (1, 52) = 38.49, p<.0001, respectively), as well as the target type by block interaction (F (1, 52) = 35.86, p<.0001). In addition, the target type by organization by category distinctiveness interaction (F (2, 54) = 3.27, p<.05) was significant. Because the accuracy effects are all in the same direction as the search time effects, the possibility of a speed-accuracy tradeoff is not supported.

The interactions in both the search time and accuracy data indicate that effects of the independent variables on performance decrease with practice. This pattern of results has been found by several others as well (e.g., Card, 1982; McDonald et. al., 1983). Since the goal of this study was to recreate a scenario representative of a novice's interactions and since the effects are greatly reduced by practice in the second block, the data from the first block were analyzed separately.

Block 1. Analysis of these data yielded the effects of target type (F (1, 53) = 90.92, p<.0001), organization (F (2, 55) =26.36, p<.0001), and category distinctiveness (F (1, 55) =7.20, p<.01). These effects were all in the same direction as those for all trials reported above. The interaction between target type by category distinctiveness was also significant (F (1, 53) = 6.29, p<.08). Post hoc Tukey procedure (st (53) = 3.75, ∂ = .05) revealed that when categories are distinct, the time it takes to locate a term or definition target is the same. When categories were overlapping, term targets were found faster than definition targets. In both target type conditions, targets were found faster when categories were distinct. All other interactions were not significant.

Across all conditions and trials, the average error rate was 4.57%. Analyses of these data resulted in the main effects of target type (F (1, 54)

=104.63, p<.0001) and organization (F (2, 54) =3.72, p<.05). Subjects searching for term targets made fewer errors than when searching for definition targets. Alphabetical and categorical organizations produced fewer errors than random organizations (st (54) = 4.15, ∂ = .05). There was not a significant difference between the number of errors in the alphabetical and categorical organizations. Thus, the error effects do not indicate that a speed-accuracy tradeoff occurred.

Identical Items-Block 1. Since the clothing and vehicles categories were included in both the distinct and overlapping conditions, the trials from these two categories were analyzed separately. These results should be free from any effects caused by item differences which might be present in the data described above. Mean search time and accuracy rates are listed in Table 4. Organization and type of target both produced significant main effects (F (2, 54) =26.62, p<.0001; F (1, 54) =43.64, p<.0001, respectively). Term targets and non-random organizations facilitated search time over definitions and random organizations. As in the previous analyses, there was no significant difference between alphabetical and categorical organizations (st (54) = 3.41, ∂ = .05). The lack of a significant main effect for category distinctiveness previously found suggests that effects of category distinctiveness found in the previous two analyses are due to item distinctiveness.

No interactions were significant in this analysis. One explanation for this may be that some effects occur in very early trials but are washed out with minimal practice. An examination of learning trends during the first block of trials suggests that some effects might be present during the earliest trials. See Appendix C for analysis of learning trends.

Table 4

Error-Free Search Time (Seconds) and Percentage of Errors from Trials of Identical Items

Category Distinctiveness:

	Distinct		Overlapping		
Target Type:	Term	Definition	Term	Definition	
Organization: Means					
Alphabetical	3.36 (0.63)	5.73 (0.94)	3.49 (0.00)	6.96 (7.81)	4.89 (2.35)
Categorical	4.18 (0.00)	5.83 (2.19)	5.20 (0.00)	6.85 (3.75)	5.52 (1.49)
Random	6.77 (0.00)	8.69 (5.63)	6.99 (0.31)	8.93 (5.31)	7.89 (2.81)
Means	4.77 (.21)	6.84 (2.92)	5.20 (.10)	7.58 (5.63)	
Means	5.80 (1.57)		6.39 (2.87)		

Note. Error percentages are listed in parentheses

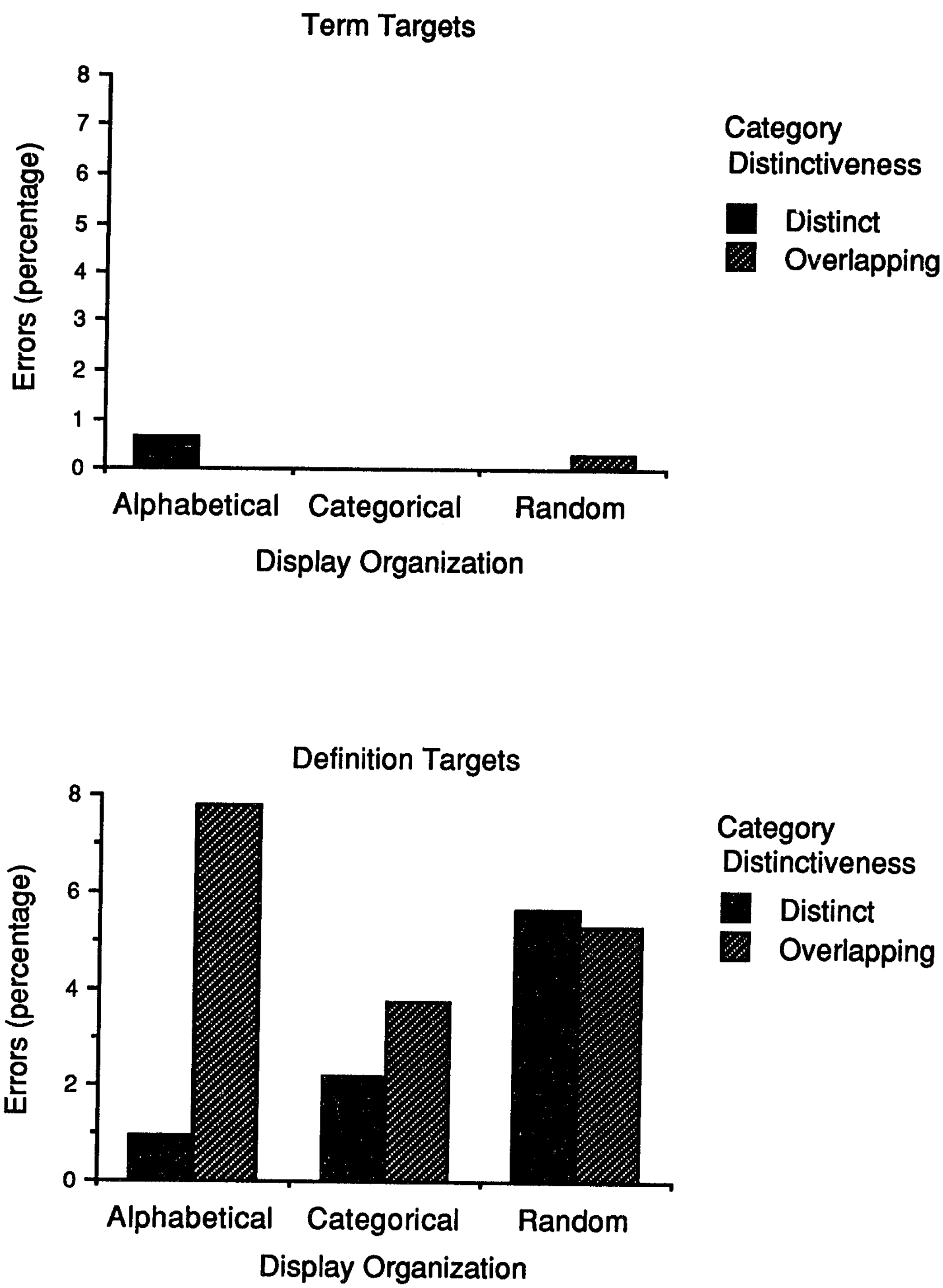
Across all conditions, subjects made an error on the average of 2.22% of the identical item trials. When the errors were analyzed from the this set of data, a different pattern of results emerged. Contrary to search time results, category distinctiveness appears to influence accuracy a great deal. Category distinctiveness not only produced a significant main effect (F (1, 54) = 6.40, p<.05), but also interacted with both organization (F (2, 54) = 3.33, p<.05) and target type (F (1, 54) =7.05, p<.01). In addition, category distinctiveness, organization, and target type produced a significant three way interaction (F (2, 54) =5.25, p<.001). This interaction is shown in Figure 1. Searching for items in distinct categories resulted in fewer errors than searching for items in overlapping categories, whereas searching for term targets was more accurate than searching for definition targets. Category distinctiveness effects are greatest in alphabetical organizations (categorical distinctiveness by organization interaction) and when targets are definitions (categorical organization by target type interaction). In fact, the main effect of category distinctiveness can probably be attributed to the condition in which the organization is alphabetical and targets are definitions (st (54) = 4.83, ∂ = .05). The finding that category distinctiveness effects alphabetical organizations but not categorical organizations is counter-intuitive and thus, quite interesting. Possible reasons for this pattern of results will be presented in the discussion section.

Finally, the main effect for target type was again found (F (1, 54) =60.33, p<.0001). Searching for term targets resulted in fewer errors than searching for definition targets.

Post Questionnaire. Each subject was asked to complete a post-test questionnaire after the experiment was complete. The questionnaires

Figure 1

Accuracy as a function of organization, target type and category distinctiveness



confirmed that nearly all subjects identified the type of display organization for their condition. Only four subjects made incorrect identifications. Two subjects from the distinct categories/categorical organization condition guessed the organization was random and one subject from each of the distinct categories/random organization and the overlapping categories/random organization conditions attempted to describe a non-random organization.

Discussion

In this experiment, the effects of display organization, target type, category distinctiveness, and block on search time and accuracy were studied. As previous research has found, effects of organization and target type on search time are reduced by a very small amount of practice (64 trials (Card, 1982; McDonald et. al., 1983)). Similarly, category distinctiveness effects decrease with practice. However, the fact that these effects decrease with experience does not warrant their dismissal as factors to be considered in interface design. In fact, these factors are likely to affect performance of new or infrequent users of a system. In addition, the amount of continuous practice received by subjects in this experiment is atypical of users' interaction with real systems (Mehlenbacher, Duffy & Palmer, 1989). Subjects in this study spent roughly an hour continuously locating targets on the display. In a normal interaction with a system, it is unlikely that a user would search for 128 targets in one interaction, rather the same number of searches would be drawn out over numerous interactions. In this type of situation, it would probably take even longer to reach the practiced level of the subjects in this experiment. Additionally, users would most likely forget at least some of what they had learned about the display organization on previous interactions and would need to relearn the organization or categories implicit within the display items.

Although most effects of the independent variables on search time seemed to disappear by the end of the first block of trials, only the effect of target type on accuracy interacted with bock. This is another reason to account for these factors in display design. It can be argued that effects of accuracy are even more important than effects of search time. When search time is significantly affected by a factor such as organization, it rarely causes delays of more than a few seconds. However, when accuracy is reduced, the amount of time lost is much greater. Consider a hierarchical menu system in which a series of levels must be traversed before the user can find the bottom level information desired. If the wrong menu item is selected in one of the top-level menus, it could be a number of minutes before users realize their mistake and then are able to determine where the mistake was made and correct it. For this reason, factors which affect accuracy of search should be closely attended to in display design.

As found by McDonald et. al. (1983), term targets are found faster and more accurately than definition targets. This was a robust finding in this experiment as in McDonald et. al.'s. Definition targets clearly make the search task more difficult and are more often the rule than the exception in every day interactions with computer systems.

As predicted from McDonald et. al.'s (1983) study and Holland and Merikle's expert-explicit target condition (1987), there was no difference in search time or errors between alphabetical and categorical organizations when the targets were explicit terms and the categories were distinct. On the other hand, based on studies with seemingly less distinct categories (Card, 1982; Holland & Merikle's novices, 1987), it was expected that the benefit of an alphabetical organization over a categorical arrangements should increase when targets were explicit terms, but categories were overlapping. This is based on the reasoning that the categorical organization is not as effective at narrowing search when the categories are ambiguous. However, search time and errors were again the same for both types of organization. There are several plausible explanations for this result.

Perhaps subjects quickly learned the categorical organization even though it was not as clean as the distinct categories organization. In fact, the trial-by-trial analysis (Appendix C) shows that for both distinct and overlapping conditions the alphabetical organization produced faster search times in the first few trials.

Another explanation is that the alphabetical advantage found in other studies was caused by factors other than category distinctiveness. For instance, Card (1982) may have found an advantage of an alphabetical organization because the groups in the functional organization condition were too small to benefit from a categorical organization. The entire display of 18 items was broken down into five categories with as few as two items in some categories. It is possible that a relationship was not seen between the few items in these categories, thus reducing the advantage of a functional organization. It is also possible that there were so few items in the list, a categorical organization was not needed to narrow search. Nonetheless, when targets were explicit terms (which occurs only rarely) neither alphabetical nor categorical organizations resulted in faster or more accurate search, especially after some learning had taken place.

Results from the implicit conditions tell a similar, but not identical, story. McDonald et. al. and Hollands & Merikle found that a categorical organization facilitated search for implicit targets. Contrary to predictions, results from this study indicate that there is no significant difference between alphabetical and categorical organizations when categories are distinct. In addition, as predicted there was no difference between the two organizations when categories were overlapping. However, an interesting pattern emerged from the organization by category distinctiveness by target type interaction for the error data.

When targets are definitions it appears that the accuracy of search through alphabetical organizations is reduced as category overlap increases, but accuracy of categorical organization search is unaffected. This result is surprising and counter intuitive. It was predicted that a decrease in categorical distinctiveness would not affect performance with alphabetical organizations, but instead, categorical organizations would lose their effectiveness in narrowing search.

Perhaps when subjects don't know the exact target they are searching for, they search through the list in a serial fashion and select an item that closely matches the definition target. In an list containing items from overlapping categories, there may be more items that closely match the target than in a list containing distinct categories. In this case, subjects select the first close match they come to. This in not a problem in a categorical organization, since search is narrowed, thereby reducing the probability of selecting the wrong item. However, in alphabetical organizations search is not narrowed and the probability of coming to a plausible item before the correct target is found increases. Furthermore, subjects in categorical organizations may have acted as did the subjects in Naus et al.'s (1972) study and exhaustively searched the entire target category, thereby further reducing the probability of selecting an incorrect item.

An example will clarify this point. A subject given the definition target "a wagon or van designed to be pulled by an automobile or truck" might be tempted to select *camper-trailer* if this item is seen before the correct target, *trailer*. In a categorical organization, the subject would likely look under the vehicles category for a target to fit this description. Under this category only the item *trailer* is located, thereby leading to a correct selection.

Aside from the differential effect category distinctiveness has on the search accuracy of alphabetical and categorical organizations, these two organizations were not significantly different from each other under any other circumstances. The fact that an advantage of categorical organization has been found by others (Hollands & Merikle's expert condition, 1987; McDonald et. al., 1983), but did not extend to this study can be attributed to some methodological differences between this study and previous ones. One important difference is that the items used in previous studies formed categories that were most likely very familiar to the subjects. McDonald et. al. (1983) used the categories of food, animals, minerals and cities whereas Hollands and Merikle (1987) used biopsychology, cognition, personality, perception, social psychology. These categories and the items that comprise them should have been very familiar for experts in psychology such as college professors. The categories and items used in this study seem less familiar than these categories and items. Thus subjects in the distinct categories condition may not have benefited from the categorical organization simply because the categories and items were less familiar and therefore, less distinct than those used in previous studies. The importance of specific items and categories is supported by the fact that many of the category effects were eliminated once the effects of individual items was controlled for by analyzing identical items. In addition, contrary to this experiment, the categories in Hollands and Merikle's (1987) displays were labeled at the top of the display. It is likely that the labels increased the subject's ability to utilize the categorical organization.

Category distinctiveness did not have a significant effect on search time as was predicted. However, category distinctiveness did appear to affect accuracy when identical items were analyzed separately. In general, fewer errors were made when categories were distinct. This effect of distinctiveness can largely be attributed to the condition in which the organization is alphabetical and targets are definitions. Finding the categorical distinctiveness effect when measuring accuracy is just as important as its presence in search time results, if not more so. As discussed previously, making an error can be more detrimental than taking longer to find the correct target. Therefore, the distinctiveness between categories is important in display design as well.

In summary, initial interactions with a display can be made more accurate and be performed more quickly when explicit targets are provided; however, it is typically not the case that users know exactly what they are searching for as it is often difficult to provide non-expert users with an explicit target. Performance is also enhanced when non-random organizations are used (i.e., either alphabetical or categorical organizations). Finally, distinct categories reduce errors.

All of the above conclusions are based on data collected from a simple visual search task and may only be relevant to this type of task. Subjects in past studies have used database or menu to locate a specific target. Once the target was located, the subject was required to either select it with a mouse or enter the item's numerical code and the process was repeated. In actuality, this is not the way information displays are typically used, and so it should not be the only way displays are tested.

Often a user of a database system will be looking for multiple targets as opposed to a single target. Multiple target tasks are affected by menu organization differently than are single-target selection tasks. McDonald, Dayton, & McDonald (1986) gave subjects orders of 1 to 4 fast food products and had them select the products on a touch pad keyboard. When subjects searched for multiple items, menu organization effects did not disappear with practice as they did with a single-item selection task (Card, 1982; McDonald et. al., 1983). If multiple item search tasks are affected differently by menu organization, it is possible that the results of Experiment 1 would not generalize to more complex tasks, such as a problem solving task (Bødker, 1989). In addition, even though categorical organizations did not result in faster search times than alphabetical organizations in Experiment 1 (maybe because of category unfamiliarity), it is possible that categorical organizations would facilitate search over alphabetical organizations if the task is more complex.

Some display organizations might encourage a semantic understanding of the system structure or an accurate mental model (Norman, 1983). This, in turn, should enhance problem solving performance. The semantic nature of a categorical organization should force users to think about the system design in a broad, conceptual way. This type of knowledge should prove more helpful in a complex problem solving task than surface knowledge encouraged by alphabetical organizations.

Information displays are frequently used to solve problems. Database information can be located using a menu and and then used to make decisions or solve problems. These types of tasks have not been studied in relation to display organization (Paap & Roske-Hofstrand, 1986), target type or category distinctiveness. Experiment 2 builds on McDonald et al's (1986) study by using a task which requires users to locate multiple targets. Subjects are then required to integrate target information to solve a simple or complex problem.

Experiment 2

Experiment 2 is similar to the first experiment; however, subjects performed two types of problem solving tasks (a simple information retrieval task and a complex integration task), rather than a simple visual search task. Display organization, category distinctiveness, and target type were each examined again. It is expected that problem solving performance should be affected by the difficulty of the problem and may interact with other factors (organization, target type, category distinctiveness) as well.

The simple retrieval problem used in Experiment 2 required the subjects to search the display in a manner similar to search tasks employed in previous studies. Thus, results from Experiment 1 are expected to be replicated in this condition. That is, alphabetical organizations should not differ in terms of search time or accuracy from categorical organizations. The second type of problem, the complex integration problem, required subjects to search for more than one target. It is hypothesized that a categorical organization should result in equal or faster problem solving times than should an alphabetical organization for this type of problem, at least when the target is implicit. This prediction is based on the premise that the complex problems require more thought and integration within and between categories and should, therefore, benefit from an organization arranged semantically.

In addition, the information needed to solve the problems was not be directly accessible on the screen, but was be accessed by selecting an item.

Thus, memory may play a greater role in Experiment 2 than in Experiment 1. As subjects interact with the display, they should be forming a representation of it in memory. A categorical organization should emphasize semantic relations of the items over the alphabetical or random organizations, thereby encouraging the formation of a deeper, more meaningful representation than the other organizations might. When presented with a complex problem, subjects might need to consult their memorized representation to assist in decision making. A meaningful organization such as a categorical one should prove beneficial in this situation. In addition, for similar reasons, as problem difficulty increases, it is predicted that factors such as target type and category distinctiveness may have an even greater effect.

Method

Subjects. Seventy Rice undergraduate students participated as subjects in this study. Data from four of the subjects were eliminated due to the following problems: A system error occurred during the experimental session (2 subjects), one subject was a non-native English speaker and had difficulty completing the task (1), and one subject was given the wrong stimuli (1). A total of sixty-six subjects' data were used in the analyses. Subjects participated in the study to partially fulfill course requirements.

Materials and Apparatus. Three Macintosh Plus computers were used to record search time and responses. A HyperCard program was used to simulate an electronic mail-order catalog. The list of items on the screen (identical to the screen seen by subjects in the first experiment) represented the items sold by the store. At the bottom of the screen the subject was able to choose between a browsing mode and a purchasing mode by selecting radio buttons with a mouse. The buttons were located next to the words BROWSE or PURCHASE.

In the browsing mode, information about the product's price, color, weight and shipping cost was displayed when a product was selected using the mouse. In the purchasing mode, selecting an item with the mouse resulted in highlighting the item. The item could be "unhighlighted" by selecting it a second time. Highlighted items were purchased by selecting the SUBMIT ORDER button located at the top of the screen. See Appendix D for example screens from the catalog simulation.

An interactive online tutorial was created to instruct subjects in the use of the catalog. The tutorial explained how to select radio buttons and recognize if they have been selected, and how to highlight the products and recognize that they have been highlighted. The purchasing process described above was also explained in detail. As part of the tutorial, subjects were given five practice problems to solve. The problems required subjects to buy products from a reduced list of 6 products. None of the items in the practice list were also in the experimental list.

A list of 24 problems describing a purchase was given to each subject. The problem lists contained three types of problems. The simple information retrieval problem asked the subject to gather a specific piece of information about a single item and then use the information to determine whether or not to buy the product (i.e., Buy a turtleneck if it comes in red. If red isn't available, buy a raincoat). Complex integration problems required the subject to integrate information from more than one product to determine which products should be purchased (i.e., Buy a sleeping bag, backpack and a tent for your trip to the mountains. If the combined price of all three items is over $1,000, then buy the two cheapest items.). Complex integration problems were further divided into two types of problems. Integration problems either asked subjects about products that belonged to the same category, or about products that belonged to two separate categories. Each of the three types of problems was worded in such a way that the subject would be searching for either term or definition targets (See Appendix E for a complete list of all the problems used in this study).

Each subject's problem list contained each of the three problem types. Half of the problems required subjects to search for term targets, whereas the other half required subjects to search for definition targets. An effort was made to equate the difficulty between term and definition problems. Unlike the last experiment, each problem or target did not have both an term and an definition form. Rather, each problem was associated with only one target type.

Lists either presented problems with term targets in the first half and problems with definition targets in the second half, or vice versa. The order of this presentation was counterbalanced across subjects. Problems within the first and second halves were randomized to eliminate any effect of problem order.

Design and Procedure. The factorial design was nearly identical to the design used in Experiment 1. Display organization (alphabetical, categorical and random) and category distinctiveness (distinct and overlapping) were between subject factors, whereas target type (definition and term) and problem type (simple information retrieval, complex integration within the same category, complex integration between two different categories) were within subject factors.

Subjects began the study by working through the online tutorial which explained how to use the mail-order catalog. Because the tutorial was self-paced, the experimenter left the testing cubicle. The experimenter re-entered the cubicle to observe the subjects perform at least one of the practice problems to ensure that they understood how to use the catalog. Subjects were also given an opportunity to ask any questions they had about the catalog before starting the experiment.

The session began by asking the subject to read the first problem on their problem lists. When they were finished, they were instructed to select an OK button located in the center of the screen. The catalog which was arranged either alphabetically, categorically or randomly was then displayed. Subjects used the browse and purchase modes to read about the various products and select products to buy. When the desired products were highlighted for purchasing, the SUBMIT ORDER button was selected and the screen displayed a message instructing the subjects to read the second problem on their problem list. This process was repeated until the subjects had completed all 24 problems.

Both completion time and accuracy were measured. For each problem, the program recorded problem solving time and the items purchased. Problem solving time began after the subject had read the problem and selected an OK button which caused the catalog to be displayed. The time ended when the subject selected the SUBMIT ORDER button. A problem was considered incorrect if the products purchased did not meet the criteria set in the problem. Some of the problems allowed for more than one correct answer. A questionnaire was given to the subjects after they had completed the experimental trials. The questionnaire required subjects to identify any difficulties encountered and to identify under what type of organization, if any, their catalog was arranged.

Results

As in the first experiment, analyses were performed on the error-free response time and error data from all problems as well as on data from problems requiring subjects to search for identical items from the same categories in both distinctiveness conditions. The results from the identical items data will be the focus in the discussion.

Before any inferential statistics were performed, all trials with problem solving times over three standard deviations from the response time cell mean were removed from the data set. On a few occasions the timer was left running while a subject emerged from her cubical to ask a question of the experimenter resulting in a few outliers. Nine trials were removed from the data set.

Error-free problem solving times and error data for the remaining trials were each submitted separately to a 3 x 2 x 2 x 3 ANOVA. Organization (alphabetical, categorical and random) and category distinctiveness (distinct vs. overlapping) were between subject variables, whereas target type (definition vs. term) and problem type (simple, complex within one category (complex1) and complex between categories (complex2)) were within subject variables.

Analysis on All Items. Mean problem solving times are listed in Table 5. The main effects of organization, target type and problem type were each significant (F (2, 60) =22.50, p<.0001; F (1, 59) = 32.57, p<.0001; F (2, 120) =403.23, p<.0001, respectively). Post hoc Tukey analyses (st (60) = 3.40, ∂ = .05) indicated that alphabetical and categorical organizations were responded to faster than were random organizations, but were not significantly different from each other. Term targets resulted in faster response times than definition targets and simple information retrieval were solved faster than complex1

Table 5

Error-Free Search Time (seconds) and Percentage of Errors from Experiment 2

Simple Problems

Category Distinctiveness:	Distinct				Overlapping				
Target Type:	Term		Definition		Term		Definition		Means
Organization:									
Alphabetical:	20.53	(3.64)	22.41	(0.00)	20.88	(5.36)	34.75	(10.91)	24.54 (5.23)
Categorical:	21.93	(4.00)	22.37	(6.12)	22.73	(3.37)	26.50	(9.09)	23.42 (5.74)
Random:	31.03	(10.91)	39.53	(2.27)	24.22	(10.91)	29.44	(20.00)	30.82 (11.48)
Means:	24.45	(6.25)	27.94	(2.92)	22.57	(6.63)	30.23	(13.33)	
Means:	26.08 (4.71)				26.25 (9.97)				

Complex 1 Problems

Category Distinctiveness:	Distinct				Overlapping				
Target Type:	Term		Definition		Term		Definition		Means
Organization:									
Alphabetical:	43.70	(11.34)	56.15	(4.54)	45.58	(2.27)	65.84	(20.45)	52.37 (9.66)
Categorical:	44.78	(10.00)	47.87	(2.08)	39.87	(4.55)	58.97	(0.00)	48.11 (3.98)
Random:	55.75	(6.98)	75.16	(2.27)	55.60	(0.00)	85.88	(11.36)	67.68 (5.14)
Means:	48.03	(9.45)	59.39	(2.94)	47.14	(2.27)	69.90	(10.61)	
Means:	54.10 (6.08)				58.01 (6.44)				

Table 5, Continued.

Complex 2 Problems

Category Distinctiveness:	Distinct				Overlapping				
Target Type:	Term		Definition		Term		Definition		Means
Organization:									
Alphabetical:	47.07	(6.06)	64.45	(15.15)	55.03	(9.38)	94.36	(22.73)	61.56 (12.5)
Categorical:	51.92	(6.67)	49.34	(16.67)	54.65	(0.00)	83.00	(9.09)	57.63 (8.26)
Random:	65.95	(11.43)	81.78	(9.38)	72.94	(6.06)	96.74	(36.36)	76.59 (13.93)
Means:	55.08	(8.16)	65.01	(13.86)	60.86	(5.10)	90.55	(22.73)	
Means:	59.97 (11.06)				71.39 (12.20)				

Note. Error percentages are listed in parentheses

problems which were solved faster than complex2 problems (st (120) = 3.36, ∂ = .05).

Several interactions were also found to be significant. Most included the problem type variable. Problem type interacted with each of the other three factors significantly (organization: F (4, 120) =4.47 , p=.005; category distinctiveness: F (2, 120) = 9.88, p<.0001; and target type: F (2, 115) = 19.19, p<.0001). Post hoc Tukey procedures (st (120) = 4.47, ∂ = .05) indicated that across all problem types, alphabetical organizations were similar to categorical organizations, but both were significantly faster than random organizations and especially so for the more complex problems. Thus, problem complexity increased the effect of organization on reaction time.

Similarly, the problem type by category distinctiveness interaction indicates that problem complexity increased the effect of distinctiveness. Distinct categories resulted in faster response times over overlapping categories only for the complex2 problems (st (120) = 4.10, ∂ = .05). The other two types of problems were solved equally fast with distinct and overlapping conditions. Again, in the target type by problem type interaction the effect of problem complexity tended to increase the main effect of target type. As problems became more complex, the advantage of an explicit target increased.

In addition to the above two-way interactions, two three-way interactions reached significance: organization by target type by problem type (F (4,115) =3.60, p<.01) and category distinctiveness by target type by problem type (F (2, 115) =16.31, p<.0001). These interactions are shown in Appendix F. The interaction discussed previously between problem type and organization was even more pronounced for definition targets (st (115) = 5.05, ∂ = .05). Similarly, the interaction between category distinctiveness and problem type depends on target type. The interaction was enhanced when targets were definitions (st (115) = 4.72, ∂ = .05).

The overall error rate for these data was 8.01%. Main effects of target type (F (1,59) = 4.94, p<.05) and problem type (F (2, 120) = 4.08, p<.05) on errors were each significant. Post hoc Tukey tests confirmed that definition targets produced more errors compared to term targets and complex2 problems caused more errors than either the simple or complex 1 problems. Simple and complex1 problems were not significantly different from each other (st (120) = 3.36, ∂ = .05).

The category distinctiveness by target type and problem type by target type interactions were also significant (F (2, 59) =13.10, p=.001; F (2, 118) = 4.92, p<.01). Overlapping categories resulted in more errors than distinct categories, but only if targets were definitions (st (59) = 3.74, ∂ = .05). The problem type by target type interaction confirms earlier conclusions. Performance effects were amplified as complexity is introduced.

<u>Analysis on Identical Items Only.</u> As in the first experiment, the problems from both the clothing and vehicle categories were analyzed separately to eliminate any effects due to differences among individual problems. Error-free problem solving times in seconds and percentage of incorrect trials are listed in Table 6. Main effects for problem solving times were found for organization (F (2,60) = 37.49, p<.0001), target type (F (1, 59) = 122.93, p<.0001) and problem type (F (2,119) = 131.20, p<.0001). Tukey tests (st (63) = 3.40, ∂ = .05) on problem solving times indicate that alphabetical and categorical organizations facilitated response time over random organizations, but were not significantly different from each other. In addition, term targets resulted in faster times than definition targets and simple problems were solved faster than complex1

problems which, in turn, were solved faster than complex2 problems (st (125) = 3.35, $\partial = .05$). As in the analysis of the full data, there was no significant effect of category distinctiveness.

Again problem type interacted with organization (F (4, 116) = 6.49, $p<.0001$) and target type (F (2, 71) = 41.53, $p<.0001$). The effect of organization was increased by problem complexity. As problems became more difficult, the advantage of a categorical organization became increasingly larger, although it was never significantly faster than an alphabetical organization. In addition, the advantage of an term target was enhanced as problems increased in complexity. Contrary to the last experiment, an organization by target type interaction was significant for these data (F (2, 60) = 17.10, $p<.0001$). Post hoc Tukey procedures (st (63) = 4.16, $\partial = .05$) indicate that the effect of organization was dependent on target type. A categorical organization is fastest when targets are definitions, whereas either an alphabetical or categorical organization facilitates problem solving time when the target is a term (see Figure 2).

The three-way problem type by target type by organization interaction was also significant (F (4, 71) = 9.58, $p<.0001$). This interaction is shown in Figure 3. This interaction indicates that the pattern seen in Figure 2 is dependent on problem type. That is, the term target pattern holds across all problem types whereas the definition target pattern holds for complex2 problems only. These two interactions suggest an advantage of categorical organizations over alphabetical organizations. This difference becomes significant when the task increases in difficulty, that is, targets are definitions and problems are complex2 (st (108) = 5.05, $\partial = .05$). These two interactions are important because they suggest that conditions needed to be quite difficult

Table 6

Error-free Problem Solving Time (seconds) and Percentage of Errors from Identical Items Data

Simple Problems

Category Distinctiveness:	Distinct				Overlapping				
Target Type:	Term		Definition		Term		Definition		Means
Organization:									
Alphabetical:	14.19	(0.00)	18.18	(0.00)	17.63	(4.35)	21.00	(10.00)	17.37 (3.49)
Categorical:	15.37	(4.76)	22.17	(0.00)	15.38	(0.00)	19.87	(0.00)	18.26 (1.15)
Random:	20.38	(9.09)	35.02	(0.00)	16.16	(0.00)	28.91	(4.55)	26.07 (3.45)
Means:	16.57	(4.62)	25.08	(0.00)	16.39	(1.49)	23.02	(4.69)	
Means:	20.89 (2.33)				19.57 (3.05)				

Complex 1 Problems

Category Distinctiveness:	Distinct				Overlapping				
Target Type:	Term		Definition		Term		Definition		Means
Organization:									
Alphabetical:	41.26	(18.18)	59.58	(9.09)	49.00	(4.54)	57.90	(23.81)	51.64 (13.79)
Categorical:	40.85	(13.64)	44.72	(4.76)	39.12	(4.55)	46.39	(0.00)	42.75 (5.88)
Random:	56.24	(4.76)	78.07	(4.55)	55.03	(0.00)	78.81	(0.00)	67.16 (2.30)
Means:	46.38	(12.31)	61.07	(6.15)	47.57	(3.03)	61.86	(7.94)	
Means:	53.97 (9.23)				54.36 (5.43)				

Tabel 6, Continued

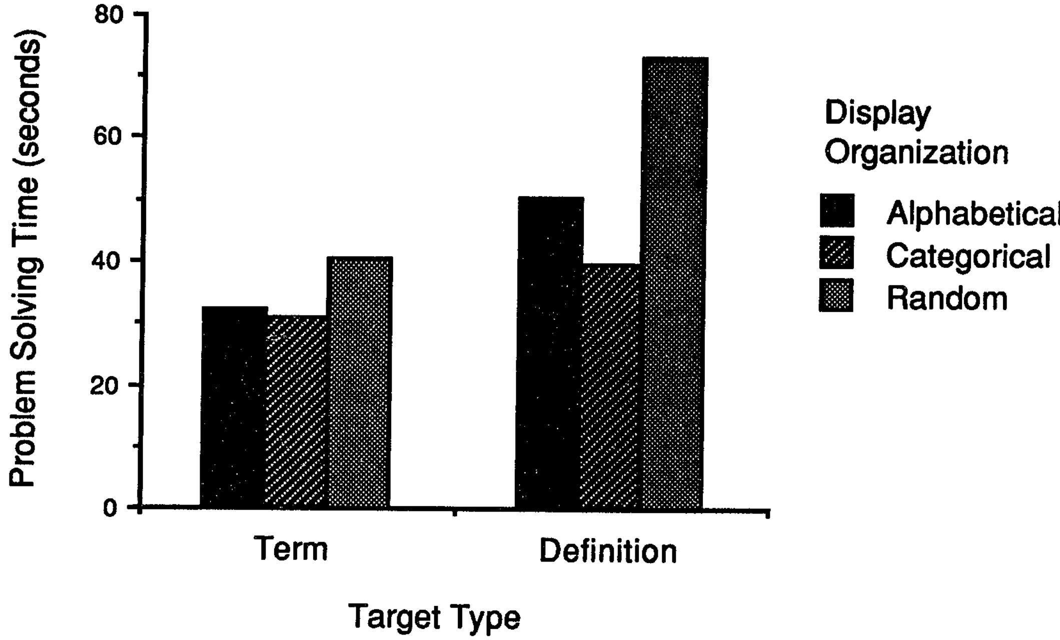
Complex 2 Problems

Category Distinctiveness:	Distinct				Overlapping				
Target Type:	Term		Definition		Term		Definition		Means
Organization:									
Alphabetical:	47.66	(9.09)	99.68	(27.27)	39.75	(9.09)	78.64	(27.27)	66.58 (20.00)
Categorical:	38.56	(0.00)	54.33	(45.45)	48.67	(0.00)	63.21	(33.33)	51.68 (22.22)
Random:	56.94	(9.09)	145.90	(9.09)	53.76	(18.18)	146.57	(59.09)	100.82 (30.91)
Means:	47.42	(6.06)	107.60	(27.27)	47.22	(9.09)	88.78	(40.00)	
Means:	73.68 (16.67)				70.71 (29.59)				

Note. Error percentages are listed in parentheses

Figure 2

Mean problem solving times across organization and target type



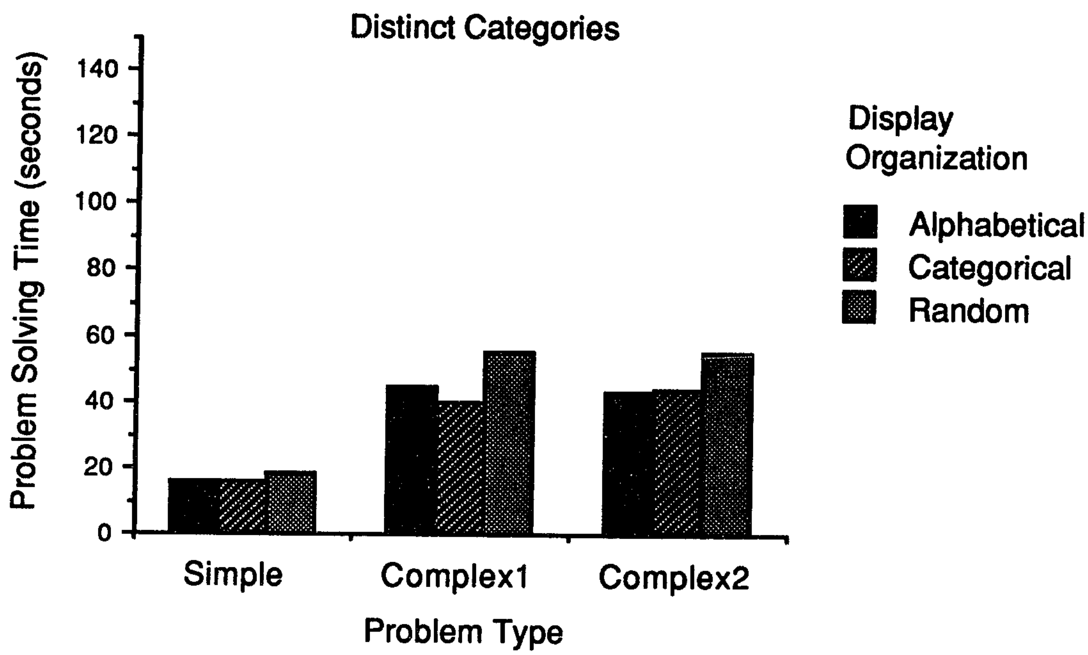
in this experiment in order to produce the significant advantage of categorical organizations found in previous studies (Holland & Merikle, 1987; McDonald et. al., 1983). Possible reasons for this result are presented in the Discussion section.

As in the full data analyses the problem type by target type by category distinctiveness interaction was significant for these data (F (2, 101) = 6.08, p<.005). As seen in Figure 4, problem complexity enhanced the effect of target type which was mediated by category distinctiveness. The category distinctiveness effect was significant only when problems were complex2 (st (108) = 4.73, ∂ = .05). In this situation, however, the overlapping definition condition produced faster problem solving times than the distinct definition condition. As will be discussed shortly, this effect can probably be attributed to a speed-accuracy tradeoff.

The mean problem-solving error rate across all conditions was 9.53%. The percentage of errors individual subjects made ranged from 0 to 30%. Analysis of the error data for identical items indicated that accuracy was greatly affected by independent item differences. Several new effects and interactions were found in the identical item data, whereas only a few remained consistent with the full data. Overall, term targets and distinct categories resulted in fewer errors than definition targets and overlapping categories (F (1, 59) = 11.71, p<.001; F (1, 60) = 3.57, p=.06). These two factors also interacted with each other (F (1, 59) = 9.87, p<.005). Category distinctiveness only affected accuracy when targets were definitions.

As in the response time data, problem type affected accuracy performance (F (2, 120) = 34.56, p<.0001) and interacted with organization, target type, and category distinctiveness (F (4, 120) = 3.01, p<.05; F (2, 117) =

Figure 3
Mean problem solving time across organization, target type and problem type



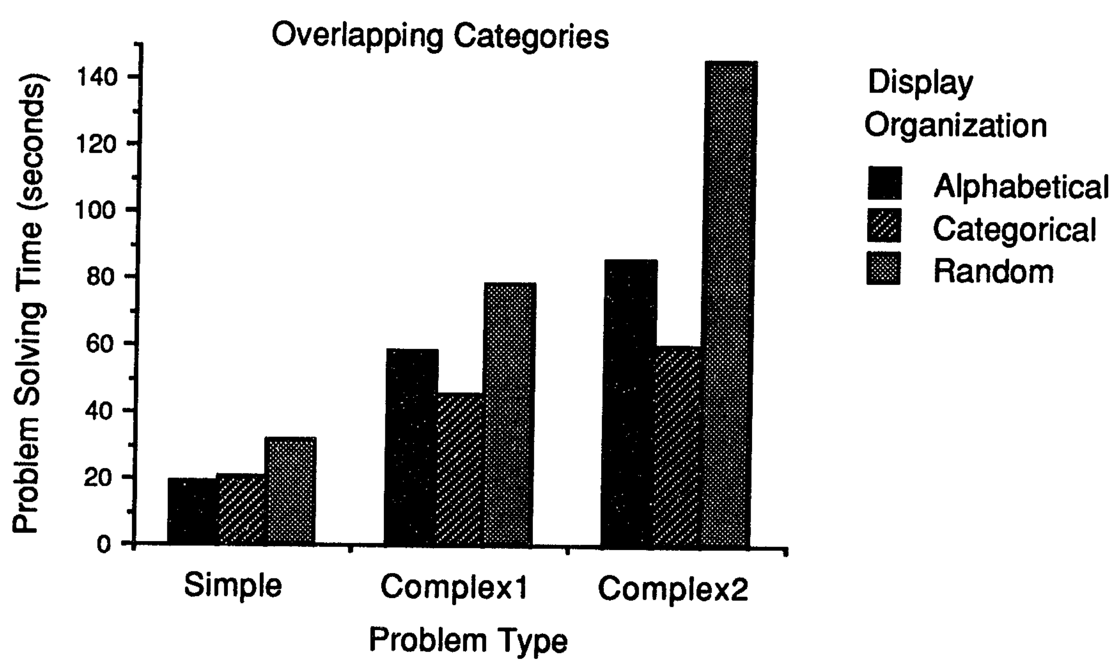
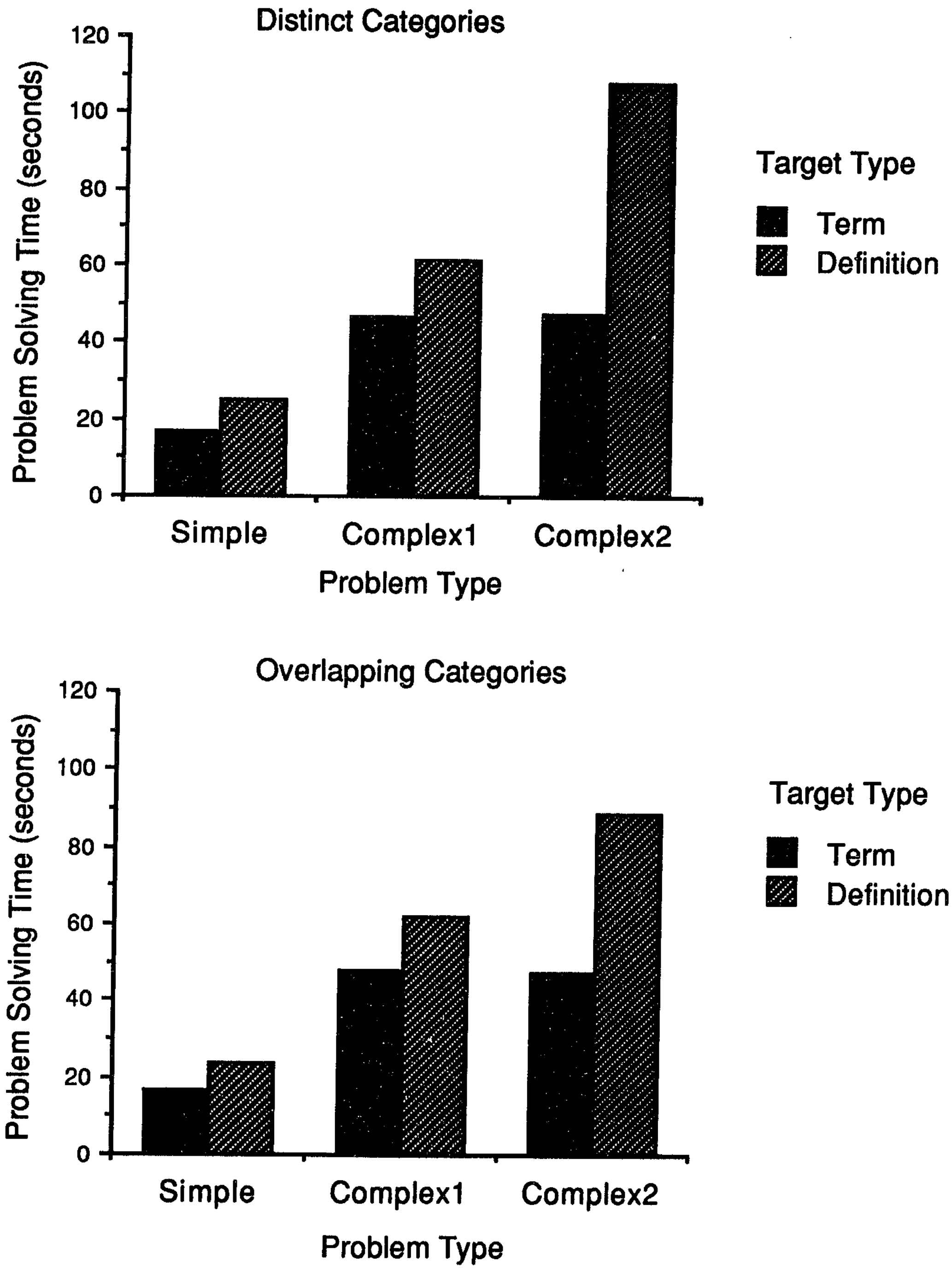


Figure 4
Mean problem solving times across problem type, target type and category distinctiveness



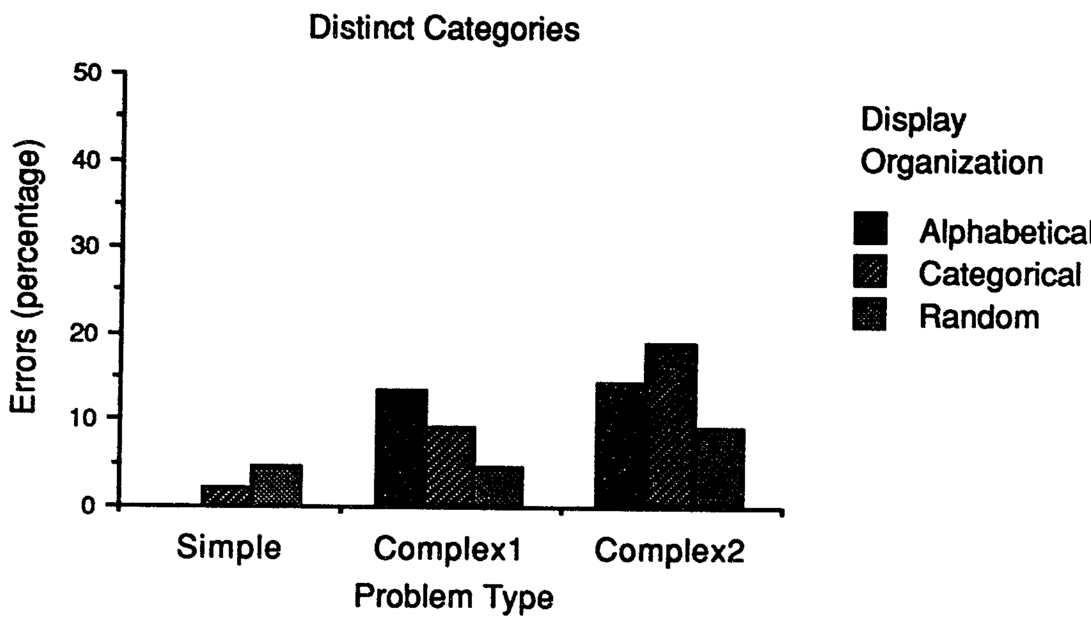
15.26, $p<.0001$; F (2, 120) = 4.07, $p<.05$, respectively). These effects increase as the problem becomes more difficult, that is, definition targets are searched for in random organizations and items are from overlapping categories. The three-way interaction between problem type, organization and category distinctiveness also supports this conclusion (F (4, 120) = 3.41, $p<.05$). See Figure 5.

The interaction between organization and category distinctiveness reached significance in these data (F (2, 60) = 3.15, $p=.05$). As shown in Figure 6, when categories are overlapping and the organization is alphabetical or random accuracy is decreased. Introducing overlapping categories to a categorical organization, however, has no effect on accuracy.

Although the problem type by category distinctiveness by target type interaction was not significant in the these data, it is felt that there is enough support to suggest that the counter-intuitive response time advantage that overlapping categories had in the definition complex2 condition can be described as a speed-accuracy tradeoff. The means from these two points in the accuracy data were in an opposite direction from the reaction time data. When targets were definitions and problems were complex2, displays with overlapping categories produced more errors than displays with distinct categories (40.00% vs. 27.27%). These two means are significantly different (t (60)= 4.46, $p<.0001$). Furthermore, as discussed above, the main effect of category distinctiveness and the interactions between category distinctiveness and problem type and category distinctiveness, organization and problem type were each significant and consistently indicated that fewer errors were made when categories were distinct.

Figure 5

Mean accuracy across problem type, organization and category distinctiveness



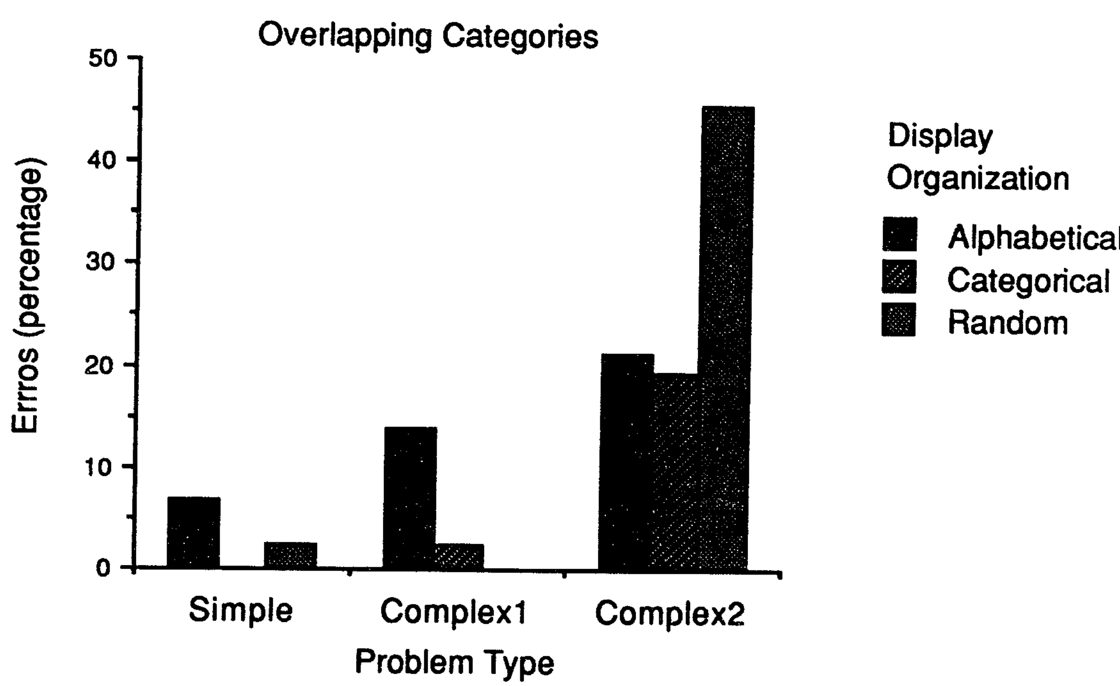
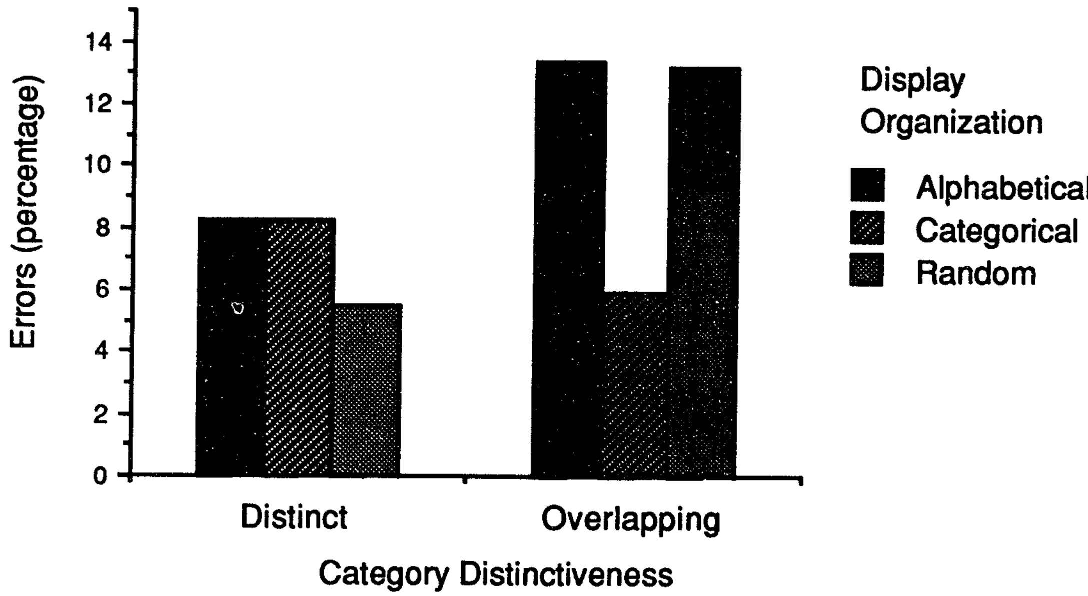


Figure 6

Mean accuracy across organization and category distinctiveness



Again with this study, it was felt that plotting the learning trends for the organization by category distinctiveness conditions would provide an interesting look at learning rates for the various conditions. However, the variability between the individual problems was so large that it was difficult to draw conclusions about learning trends. These graphs are presented in Appendix G.

Post Questionnaire. Every subject in each condition correctly identified their catalog organization. No serious problems were encountered with the experimental program.

Discussion

The goal of Experiment 2 was to determine if the factors studied in Experiment 1 (organization, target type and category distinctiveness) affected problem solving time and accuracy on three types of problems (simple, integration within a single category, integration between two categories). As in the first experiment in which a simple search task was used, all three factors affected problem solving performance.

One of the most robust findings in Experiment 1 was that explicit targets were found faster and located more accurately than definition targets. This also seems to be the case with a problem solving task. Problems that required subjects to locate items stated explicitly in the problem were solved faster and more accurately than when items were stated implicitly as definitions in the problem. This was especially true when categories were overlapping (error only) and problems were complex. When searching for a definition, the subject is forced to use meaning to help narrow search. When categories are overlapping, there are more items with similar meanings and accuracy decreases. Search is slowed when the subject has to keep in mind more than one definition as required by complex problems. This is especially difficult when the items come from more than one category so that the meanings are more semantically distant. When targets are explicit, these problems either do not exist or are not as severe.

At this point it should be noted that the effect of target type cannot be isolated from any independent item differences due to the use of different problems in each condition. It is possible that the problems used in the definition condition were simply more difficult than those in the term condition. However, because the effect of target type was present in the first experiment and in several previous studies (Hollands & Merikle, 1987; McDonald et. al., 1983) in which the term and definition targets were identical, it is unlikely that this effect in Experiment 2 can be completely accounted for by item differences.

Category distinctiveness had little effect on problem solving time. There existed only one situation in which the category distinctiveness effect was significant. When targets were definitions and problems were complex2, displays with overlapping categories were responded to faster than displays with distinct categories. However, this counter intuitive result can be explained by a speed-accuracy tradeoff as discussed previously. In general, as in Experiment 1, the accuracy data were influenced to a greater extent by category distinctiveness. In each case, distinct categories were responded to more accurately than overlapping categories supporting the speed-accuracy tradeoff explanation. The accuracy advantage of distinct categories is especially large when targets are definitions and problem are complex, as suggested by the interactions of category distinctiveness by target type and by problem type.

Predictions about organization effects were supported in this experiment. Overall, non-random organizations facilitated problem solving times and reduced errors. When problems included term targets to be searched for there was no difference between alphabetical and categorical organizations. In addition, there was no difference between alphabetical and categorical organizations when problems included definitions and problems were simple. These results all replicate those found in Experiment 1. However, when complex problems included definition targets, categorical organizations resulted in faster problem solving times than did alphabetical organizations. Thus as problems become increasingly complex, a categorical organization may facilitate response time even when the categories are not highly familiar.

Why was the advantage of a categorical organization in this experiment only significant when the task was highly complex, whereas other studies found this effect using a simple search task? Perhaps the simplicity of the subjects' task in past studies enabled them to use a categorical organization with little or no effort. The categories employed in these studies were usually distinct and highly familiar thereby reducing the amount of learning required to effectively narrow search. As argued previously, the categories used in Experiments 1 and 2 were less distinct and familiar (even categories from the distinct condition) which required more effort and learning time to effectively use. In condition where the task was less complex (i.e., term targets and simple problems) it is possible that subjects employed a more efficient search strategy than using the categories which required time and effort to learn. When the task became highly complex, however, the benefit of using the categorical organization to assist problem solving outweighed the cost of learning time and effort.

The fact that these effects are not dependent on category distinctiveness is somewhat puzzling. The lack of a four-way interaction between category distinctiveness, organization, problem type and target type would suggest that for complex problems with implicit targets categorical organizations have an advantage even when categories are overlapping. In addition, accuracy measures indicated that performance with alphabetical organizations worsened with increased overlap between categories, whereas performance in categorical organizations remained constant. A similar pattern of results was found in the organization by category distinctiveness by target type interaction from the Experiment 1 accuracy data (Figure 1). This result is again opposite of predictions and counter intuitive to a point, but supports conclusions drawn in Experiment 1.

As suggested in the discussion of Experiment 1, subjects are likely to search through the display in a serial fashion and select the first item they find that might fit the definition. Because items from overlapping categories are more likely to have similar definitions, it is more likely in these conditions that a plausible match will be found before the correct target is seen thereby resulting in an error. When the display is arranged categorically, search is reduced to one fourth of the screen thereby reducing the number of false matches that might be seen. In this situation, subjects probably search the target's category exhaustively and then select the best match (Naus, 1972).

The above explanation is logical only when subjects are searching for definition targets. The fact that the category distinctiveness by organization interactions is not mediated by target type as in Experiment 1 is somewhat puzzling, but the direction of the means suggest that this counter intuitive pattern is largely present when targets are definitions. When targets are terms, the search time and accuracy advantage of categorical organizations for overlapping categories tends to be reduced.

The type of problem subjects were asked to solve also had a major impact on problem solving performance as well. Simple problems were always solved faster and more accurately than either of the complex problems. This, of course, is logical because in these problems subjects were only required to buy one product as opposed to two or three in the complex problems. The two types of complex problems that required subjects to buy items from the same category or two different categories only affected performance differently when the task was made more difficult with random organizations, definition targets or overlapping categories. The interactions between problem type and the other factors were not additive, rather, as the problem type became more difficult random organizations, definition targets, and overlapping categories had greater detrimental effects on performance compared to non-random organizations, term targets and distinct categories. This result could be due to the increased cognitive difficulty associated with the complex problems.

The complex problems required subjects to hold multiple targets in memory, forcing them to direct more resources to retaining information which left fewer resources for performance of the problem solving task. It is possible the multiple items from the complex2 problems required even more resources for the retention task because the items were less similar than those from complex1 problems which were from the same category. This explanation assumes that cognitive resources are limited and that dual task performance may require more resources than are available, thereby reducing performance of at least one of the tasks (in this case the problem-solving task (see Wickens, 1984).

The results of this study suggest a reoccurring pattern. Effects of organization, target type, problem type, and category distinctiveness are all enhanced as the task is make more difficult. The addition of each of these complicating factors to a display design (e.g., complex problems, overlapping categories or definition targets) adds additional difficulty to the task.

General Discussion

The results of Experiment 1 and Experiment 2 look similar, but due to the increased complexity of the task in Experiment 2 several interesting new results were found. The effect of problem type in Experiment 2 interacted with the other effects of organization, target type and category distinctiveness manipulated in Experiment 1, but the pattern of results was similar. The effects were generally enhanced as problem complexity increased. Consequently, seemingly minor effects on the order of a few seconds may seem trivial in the visual search paradigm, but may become even larger when task difficulty increases. For instance, in Experiment 2 the reaction time differences between term and definition targets were 7.6 seconds for simple problems, 14.5 seconds for complex1 problems and 48.6 seconds for complex2 problems. The fact that type of problem interacted with each of the other manipulated factors in Experiment 2 should serve as a warning to designers that the type of task is critical when designing information displays.

Other factors, such as the the individual items used in each display, also need to be considered in display design. This was pointed out by the analysis of the identical items from both category distinctiveness conditions. The significant effects and interactions found in the analysis of the full data differed from the results found in the analysis of the data which controlled for individual item effects. This suggests that individual items of a display have some effect on search time and accuracy. This argument should not be interpreted as a justification for terminating display research or the comparison of results from studies which use different display items. Despite the fact that items appear to influence the effects of some factors on performance, several consistent results have been found across studies (i.e., term targets are superior to definition targets). Therefore, research should continue in this area, but attention should be given to display items employed and how they influence the factors studies. In addition, item features such as familiarity, complexity and type (action vs. object) should also be studied. It is possible generalizations could be made across these features as well.

The inconsistent results found for organization effects might be explained by individual item differences. When a simple visual search task was performed, alphabetical organizations did not result in significantly different response times from categorical organizations in either experiment. However, a categorical advantage was found for complex problems when the target was a definition. The fact that a categorical advantage was not found for simpler tasks may be a result of item effects or subjects' unfamiliarity with the categories. Because individual items influence how factors such as display organization affect performance, it is likely this is one of the reasons that a categorical advantage was not found in this study but was in others. If individual items have an effect within one experiment, they probably affect comparisons of results made across studies as well.

The various categories employed in the experimental displays might also impair the ability to compare results with those of another study. It is likely that the categories and items used in these two studies were too different from those used in other studies to allow comparisons to be made. However, it is interesting that despite the apparent difficulty of the items, a categorical organization was beneficial for tougher problems. This was as predicted.

When subjects are asked to solve difficult problems an organization that helps build accurate mental models of the display or provides semantic information should prove beneficial. It is likely that a user's normal interactions with a display are more similar to these more difficult situations and involve complex tasks. Therefore, a categorical organization should be considered in display design.

Although guidelines can be generated based on these studies, care should be taken to treat them only as guidelines and not as laws. Designers of displays should remember that each display's items are unique and the organization which will enhance performance will probably vary depending on other factors such as target type and problem type. If time and money will allow, a usability test should be performed on each information display to determine its optimal organization. Usability tests are only useful, however, if several performance measures are evaluated.

Results from both studies indicate various factors studied here affect accuracy as well as response time. In both studies, several interactions were significant for errors, but not for search time. The pattern of these effects was not contradictory to the search time effects eliminating the possibility of a speed-accuracy tradeoff. They do indicate, however, that categorical distinctiveness, although it had little effect on search time, did had an effect on error rate. Interestingly, the effect was greater for alphabetical than for categorical organizations, particularly for definition targets. The presence of items from overlapping categories is much more confusing in an alphabetical organization, particularly when the user does not know the explicit target. As mentioned earlier, determining which factors affect accuracy should be of particular interest to display designers because the consequences of making a mistake are much less trivial than the consequences of a slowed visual search (a number of minutes vs. a number of seconds).

Several general guidelines follow from these two studies. Briefly, they are: (a) use distinct categories if possible; (b) provide an explicit target if possible; (c) consider the display items and the task difficulty in design; (d) if usability tests are conducted, measure accuracy as well as search time; (e) use either alphabetical or categorical organization when targets are explicit; (f) use a categorical organization when targets are implicit, especially if problems are complex; (g) the use of categorical organizations may be even more important in cases in which items overlap across categories.

These results indicate that several factors interact with display organization causing the optimal organization to change as these factors change. It is likely these studies did not test all the factors that influence organization effectiveness. Each display should be evaluated individually for factors that may affect performance before choosing an organization.

These two studies suggest the need for several other studies. First, the effect of item familiarity should be tested by manipulating category familiarity and distinctiveness to tease out any differential effects these two factors might have on search time and accuracy. In addition, manipulating item and category familiarity might determine whether familiarity can explain the lack of categorical organization advantage found in these studies. Second, as the second study has shown, the type of task significantly effects how display factors such as organization and category distinctiveness interact with each other to affect reaction time and accuracy. Only one problem solving domain was studied. Several other problem tasks should be studied as it is likely the effects of these factors will vary somewhat. Finally, this study or one similar should be performed using command names or menu items from existing systems or applications. The assumption was made in these studies that categories used in most systems are overlapping and unclear, especially to novice or casual users. This assumption should be tested empirically.

References

Battig., W. F., & Montague, W. E. (1969). Category norms for verbal items in 56 categories: A replication and extension of the Connecticut category norms. Journal of Experimental Psychology Monograph, 80 No. 3, part 2, 1-46.

Bødker, S. (1989). A human activity approach to user interfaces. Human-Computer Interaction, 4, 171-195.

Boehm-Davis, D., Holt, R., Koll, M., Yastrop, G., & Peters, R. (1989). Effects of different database formats on information retrieval. Human Factors, 31(5), 579-592.

Card, S. K. (1982). User perceptual mechanisms in the search of computer command menus. Proceedings of the Conference on Human Factors in Computer Systems, Gaithersburg, MD, 190-196.

Card, S. K. (1984). Visual search of computer command menus. In H. Bourma and D. Bouwhuis (Eds.). Attention and performance X. Hillsdale, NJ: Erlbaum, 97-108.

DeRosa, D. V., & Tkacz, S. (1976). Memory scanning of organized visual material. Journal of Experimental Psychology: Human Learning and Memory, 2 (6), 688-694.

Dumais, S.T. & Landauer, T. K. (1984). Describing categories of objects for menu retrieval systems. Behavior Research Methods, Instruments and Computers, 16 (2), 242-248.

Giroux, L. & Belleau, R. (1986). What's on the menu? The influence of menu content on the selection process. Behaviour and Information Technology, 5 (2), 169-172.

Guralnik, D. B. (1984). Webster's New World Dictionary of the American Language. New York, NY: Warner Books.

Hammond, E. & Green, D. (1982). Detecting targets in letter and non-letter arrays. Canadian Journal of Psychology, 36(1), 67-82.

Hollands, J. G., & Merikle, P. M. (1987). Menu organization and user expertise in information search tasks. Human Factors, 29(5), 577-586.

Hunt, K. P. & Hodge, M. H. (1971). Category-item frequency and category-name meaningfulness (m'): Taxonomic norms for 84 categories. Psychonomic Monograph Supplements, 4 ,6 (Whole No. 54), 97-121.

Jonides, J. & Gleitman, H., (1976). The benefit of categorization in visual search: Target location without identification. Perception and Psychophysics, 20(4), 289-298.

Krueger, L. E., & Hettinger, (1984). Alphanumerical order and categorical effects on visual search. Bulletin of the Psychonomic Society, 22(6), 521-524.

Lachman, R., Lachman J. L., & Butterfield, E. C., (1979). Cognitive Psychology and Information Processing: An Introduction. Hillsdale, NJ: Lawrence Erlbaum Association, Publishers.

Mandler, G. & Pearlstone, Z. (1966). Free and constrained concept learning and subsequent recall. Journal of Verbal Learning and Verbal Behavior, 5, 126-131.

McCloskey, M. & Glucksberg, S. (1979). Decision processes in verifying category membership statements: Implications for models of semantic memory. Cognitive Psychology, 11, 1-37.

McDonald, J., Dayton, T. & McDonald, D. (1986). Adapting menu layout to tasks. (Report No. MCCS-86-78). Las Cruces, NM: New Mexico State University.

McDonald,J., Stone, J., & Liebelt, L. (1983). Searching for items in menus: The effects of organization and type of target. Proceedings of the Human Factors Society- 27th Annual Meeting.

Mehlenbacher, B., Duffy, T.M., & Palmer, J. (1989). Finding information on a menu: Linking menu organization to users goal's. Human-Computer Interaction, 4, 231-251.

Miller, G. A., (1969). A psychological method to investigate verbal concepts. Journal of Mathematical Psychology, 6, 169-191.

Miller, G. A., (1984). Dictionaries of the mind. In Proceedings of the 23rd Annual Meeting of the Association for Computational Linguistics. Melano Park, CA: Association for Computational Linguistics, 305-314.

Naus, M. J., Glucksberg, S., & Ornstein, P.A. (1972). Taxonomic word categories and memory search. Cognitive Psychology, 3, 643-653.

Neisser, U. (1963). Decision-time without reaction time: Experiments in visual scanning. American Journal of Psychology, 76, 376-385.

Neisser, U. (1964). Visual Search. Scientific American, 210, 94-102.

Norman, D. (1983). Some observations about mental models. In D. Getner & A. Stevens (Eds.), Mental Models. Hillsdale, NJ: Lawrence Erlbaum Association.

Paap, K. & Roske-Hofstrand, R. (1986). The Handbook of Human Factors Research on Menu Interfaces. IBM Technical Report.

Perlman, G. (1989). Making the right choices with menus. Human-Computer Interaction - INTERACT '84, 317-321.

Rosch, E., (1978). Principles of Categorization. In Rosh, E. & Lloyd, B. (Eds.) Cognition and Categorization. Hillsdale, NJ: Lawrence Erlbaum Association, Publishers.

Schwartz, J.P. & Norman, K.L. (1986). The importance of item distinctiveness on performance using a menu selection system. Behaviour and Information Technology ,5 (2), 173-182.

Shiffrin, R. M. & Schneider, W., (1977). Controlled and automatic information processing: II. Perceptual learning, automatic attending, and a general theory. Psychological Review, 84, 269-276.

Sisson, N., Parkinson, S. & Snowberry, K. (1986). Considerations of menu structure and communication rate for the design of computer menu displays. International Journal of Man-Machine Studies, 25 (5). 479-487.

Sternberg, S. (1966). High-speed scanning in human memory. Science, 153, 652-654.

Teitelbaum, R. C. & Granda, R. E. (1983). The effects of positional constancy on searching menus for information. CHI Proceedings.

Vandierendonck, A., Van Hoe, R., & De Soete, G. (1988). Menu search as a function of menu organization, categorization and experience. Acta Psychologica, 69, 231-248.

Webster's Dictionary. (1980). New York, NY: Ottemheimer Publisher, Inc.

Wickens, C. D, (1984). Processing Resources in Attention. In R. Parasuraman & D. R. Davies (Eds.), Varieties of Attention, New York: Academic Press, Inc.

Appendix A

Distinct Categories and Instances

Clothing	Vehicles	Tools	Furniture
jeans	jet ski	hammer	chair
raincoat	baby stroller	saw	table
socks	motor scooter	nail	bed
ski jacket	ice skates	screwdriver	end table
wool shirt	roller skates	level	desk
scarf	bicycle	crowbar	lamp
running shoes	trailer	chisel	couch
windbreaker	wagon	ruler	dresser
scarf	canoe	wrench	stool
mittens	skateboard	pliers	rocker
wool gloves	tricycle	drill	rug
belt	unicycle	screw	coffee table
sweat pants	skis	sandpaper	book case
leotard	jeep	pencil	cabinet
long underwear	surfboard	sander	love seat
turtleneck	sailboat	sawhorse	footstool

Overlapping Categories and Instances

Clothing	Vehicles	Sports Equipment	Camping Equipment
jeans	jet ski	bat	tent
raincoat	baby stroller	basketball	sleeping bag
socks	motor scooter	tennis racquet	fishing pole
ski jacket	ice skates	sweat band	insect spray
wool shirt	roller skates	hockey mask	tent stakes
scarf	bicycle	golf tees	hatchet
running shoes	trailer	bowling ball	hiking boots
wind breaker	wagon	swimming goggles	matches
scarf	canoe	shoulder pads	wood
mittens	skateboard	hurdle	jack knife
wool gloves	tricycle	shot put	camper-trailer
belt	unicycle	volleyball net	back pack
sweat pants	skis	floor mat	flashlight
leotard	jeep	golf clubs	cooler
long underwear	surf board	stopwatch	thermos
turtleneck	sail boat	barbell	canteen

Appendix B

Term Target	Definition Target
baby stroller	a light carriage for wheeling a baby about
back pack	a knapsack often on a light frame worn by campers or hikers
barbell weight	a metal bat with weights attached at each end used for lifting exercises
basketball	an inflated rubber ball thrown through a basket in a popular court game
bat	a club to hit the ball in baseball
bed	a piece of furniture for sleeping on
belt	a band of leather worn around the waist
bicycle	a vehicle consisting of a metal frame with two wheels - a seat - and pedals
bookcase	a set of shelves for holding books
bowling ball	a heavy ball which is bowled along a wooden lane
cabinet	a case with drawers or shelves
camper-trailer	a trailer equipped for camping out
canoe	a narrow and light boat moved by pedals
canteen	a small flask for carrying water
chair	a piece of furniture with a back for one person to sit on
chisel	a sharp-edged tool for cutting or shaping wood or stone
coffee table	a small low table for serving refreshments
cooler	a portable box used for keeping food and beverages cool

couch	an article of furniture on which one may sit or lie down
crowbar	a long metal bar used as a lever for prying
desk	a table for writing or drawing or reading
dresser	a chest of drawers for clothes-usually with a mirror
drill	a tool for boring holes
end table	a small table placed at the end of a sofa (etc.)
fishing pole	a slender pole with an attached line and hook used in fishing
flashlight	a portable electric light
floor mat	a thickly padded floor covering used for wrestling (etc.)
foot stool	a low stool for supporting the feet of a seated person
golf clubs	a stick used to hit a golf ball
golf tees	a small peg from which a golf ball is driven
hammer	a tool with a medal head and a handle used for pounding
hatchet	a small axe with a short handle
hiking boots	special foot gear used for long walks across rugged terrain
hockey mask	a covering to protect the face used in the sport of hockey
hurdle	a frame-like barrier which horses and runners must leap in a race
ice skates	metal runners in frames fastened to shoes for gliding on ice
insect spray	an aerosol spray spray used to repel insects
jack knife	a large pocket knife
jeans	trousers made of denim
jeep	a small rugged military or recreational automobile
jet ski	a motorized vehicle designed for one person to ride on the water

lamp	any device for producing light or therapeutic rays
leotard	a tight fitting garment for an acrobat or dancer
level	an instrument for determining the horizontal
long underwear	tight-fitting garments worn under clothing to add warmth
love seat	a small sofa for two people
matches	pieces cardboard tipped with a composition that catches fire by friction
mittens	a glove with a thumb but no separately divided fingers
motor scooter	a motor-driven bicycle
nail	a slender pointed piece of metal driven into wood to hold it together
pencil	a pointed rod-shaped instrument with a core of graphite used for writing
pliers	small pincers for gripping small objects or bending wire
raincoat	a water-repellent coat
rocker	a chair mounted on two pieces of curved wood on which it can rock
roller skates	frames with four small wheels fastened to shoes for gliding on a floor
rug	a piece of thick fabric used as a floor covering
ruler	a strip of wood with a straight edge, used in drawing lines or measuring
running shoes	footwear worn to jog or run races
sail boat	a boat that is propelled by means of a sail or sails
sander	an electric appliance equipped with a disk of sandpaper to smooth or polish

sandpaper	paper coated on one side with sand, used for smoothing and polishing
sawhorse	a rack on which wood is placed while being sawed,43
saw	a cutting tool consisting of a thin metal blade with sharp teeth
scarf	a piece of cloth worn around the neck or head,45
screw	a cylindrical or conical medal piece for fastening things by being turned
screwdriver	a tool used for turning screws
shot put	a metal ball which is propelled with an overhand thrust from the shoulder
shoulder pads	hard foam secured to a fitted piece of plastic used to protect shoulders
skateboard	a short and oblong board with four wheels ridden as down an incline
ski jacket	a warm jacket often filled with down and is water-repellant
skis	a pair of long runners of wood (etc.) fastened to shoes for gliding over snow
sleeping bag	a warmly lined zippered bag for sleeping outdoors
socks	foot covering worn under shoes
stool	a single seat having no back and no arms
stopwatch	a watch that can be started and stopped instantly-used for timing
surfboard	a long narrow board used in the sport of surfing
sweat pants	heavy cotton pants worn to absorb sweat as after exercise
swimming goggles	large spectacles to protect the eyes against water

table	a piece of furniture having a flat top set on legs
tennis racket	a stringed frame for tennis
tent	a portable shelter made of canvas (etc.) stretched over poles
tent stakes tent	pointed lengths of wood or metal for securing the ropes of a to the ground
thermos	a bottle or jug for keeping liquids at almost their original temperature
trailer	a wagon or van designed to be pulled by an automobile or truck
tricycle	a child's three-wheeled vehicle operated by pedals
turtleneck	a shirt or sweater with a high snug turned down collar
unicycle	a one-wheeled bicycle which has no handlebars,59
volleyball net	a meshed fabric that the ball must be hit over in the sport of volleyball
wagon hauling	a four-wheeled vehicle (usually with a long handle) for heavy loads
wind breaker	a light coat serving as protection from the wind
wood	a hard fibrous substance beneath the bark of trees and shrubs
wool gloves	a warm covering for the hand with separate sheaths for the fingers
wool shirt	a garment worn on the upper part of the body made of wool
wrench	a tool for turning nuts or bolts (etc.)

Appendix C

Learning Trends

As past research has shown, practice influences how display organization affects search speed and accuracy (Card, 1982; McDonald et. al., 1983). This was also confirmed in this experiment when effects of not only organization, but also target type and category distinctiveness washed out with practice in the second block of trials. Finding that display features influence user performance differently in the first few interactions than in later interactions, could indicate a need for adaptive interfaces. Therefore, these factors were examined on a trial-by-trial basis.

Subjects' data from the two categories in both distinctiveness conditions were split into two groups; those who had term targets for the first 32 trials and then received definitions, and those who received definitions followed by terms. For each subject group, search times were averaged over every eight trials for each organization by category distinctiveness condition. A plot of these means across the first block of trials is shown in Figures 1 and 2.

When term targets are presented first (Figure 1) it appears alphabetical organizations facilitate search across both category distinctiveness conditions. When this group searches for definitions, however, categorical organizations appear to be advantageous to alphabetical organizations after a short period of time. The fact that categorical organizations are not immediately superior to alphabetical organizations, supports the hypothesis that the categories within the categorical organization need to be learned before this type of display can be advantageous. Even though the subject's in this condition had been using the same categorical display for the first 32 trials when searching for term targets, it is possible the categories were not fully learned because the task did not require the subject to think about the targets in a semantic fashion. When searching for implicit, semantic targets, the subject is forced to think about the meaning of each word and perhaps it is not until this time that the categorical organization is consciously used to help narrow search.

When definition targets were searched for first as in Figure 2, a learning period appears to occur again, especially in the distinct categorical and overlapping alphabetical conditions. However, subjects in these two conditions quickly learned the organizations and their searched times reduced to match the distinct alphabetical condition's times in subsequent blocks. These three conditions continue to have similar search times throughout the term targets as well. The overlapping categorical condition, on the other hand, takes over 40 trials before it consistently produces similar search times as the other alphabetical and categorical conditions. Apparently, subjects in the overlapping categorical condition were able to learn the categories and used them to help narrow search faster if term targets were searched for first. Subjects who searched for definition targets first were not able to learn the overlapping categories until over half of the trials were performed.

This finding would suggests that learning a difficult display, such as overlapping categorical, is retarded when difficult targets such as definitions are searched for. Subjects who were allowed to "ease" into the task by searching for term targets first, apparently learned the overlapping targets faster and continued to use to display to narrow search even when required to search for the more difficult definition targets.

Figure 1

Experiment 1 learning curves across organization and category distinctiveness conditions- Term targets first

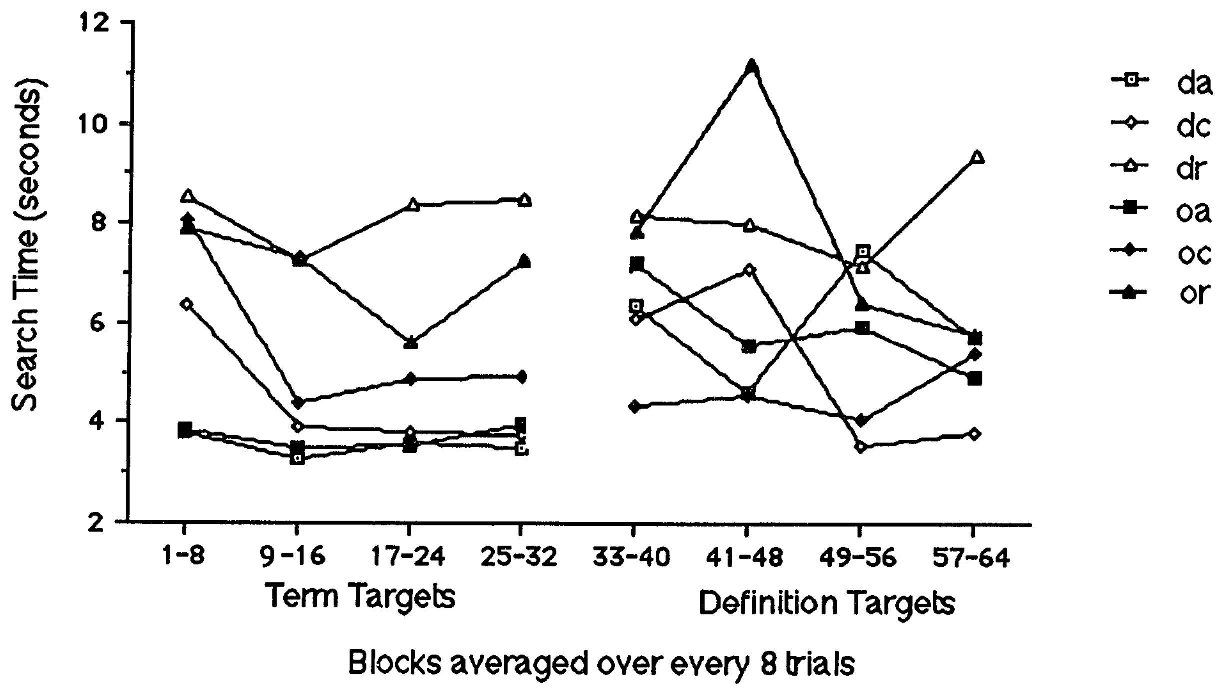
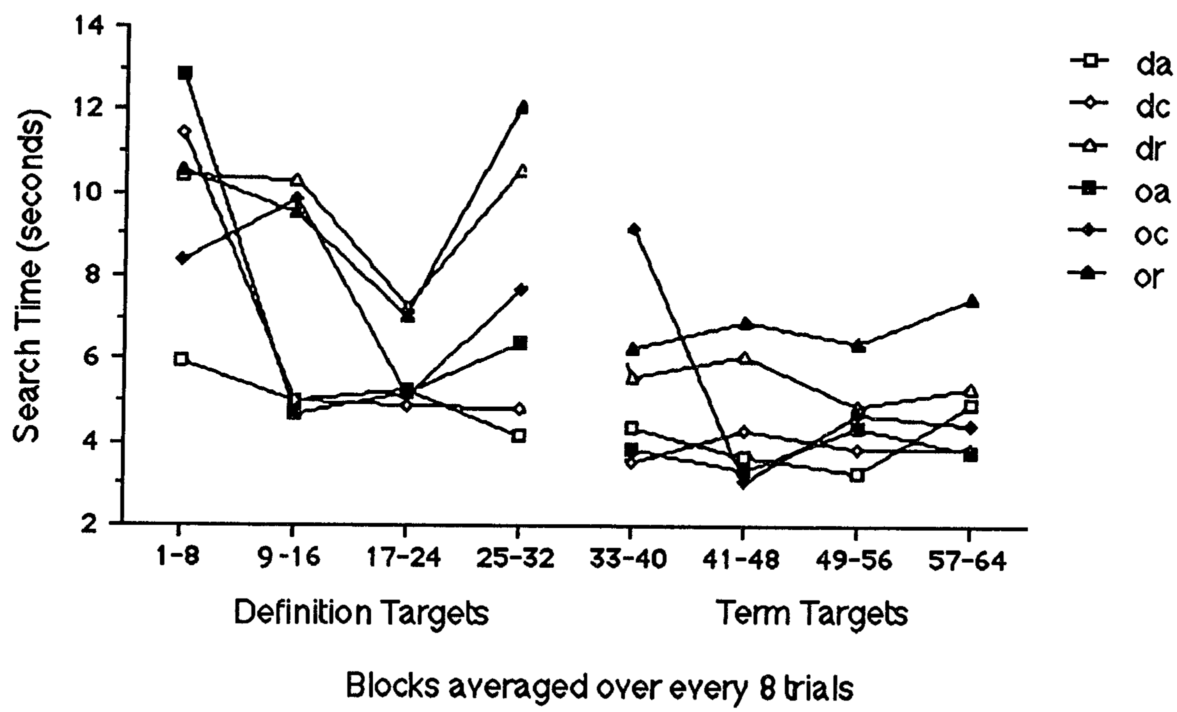


Figure 2

Experiment 1 learning curves across organization and category distinctiveness conditions - Definition targets first



Appendix D

SUBMIT ORDER

baby stroller
bed
belt
bicycle
bookcase
cabinet
canoe
chair
chisel
coffee table
couch
crowbar
desk
dresser
drill
earmuffs
end table
foot stool
hammer
ice skates
jeans
jeep
jet ski
lamp
leotard
level
long underwear
love seat
mittens
motor scooter
nail
pencil
pliers
raincoat
rocker
roller skates
rug
ruler
running shoes
sail boat
sander
sandpaper
saw
sawhorse
scarf
screw
srewdriver
skateboard
ski jacket
skis
socks
stool
surfboard
sweatpants
table
trailer
tricycle
turtleneck
unicycle
wagon
wind breaker
wool gloves
wool shirt
wrench

(●) BROWSE ( ) PURCHASE

Main menu screen from catalog program

BABY STROLLER

Price per unit: $75.00

Colors available: pink, blue

Weight in pounds: 10

Shipping cost: $5.50

Return to Menu

Screen seen after baby stroller is selected from main menu while in browse mode

Appendix E

The goal of the Experiment 2 was to provide subjects with a task that was more complex than a simple visual search task. Problems were developed which required subjects to search for one or more targets and then use information gained about the targets to determine if they should be purchased. Two classes of problems were developed - problems that required the purchase of one product only, and problems that required the purchase of multiple products. It was decided that whether the multiple products were from the same or different categories should be controlled; therefore, both types of multiple target problems were developed.

Twenty-four problems were developed for each category distinctiveness condition - eight simple problems (requiring the purchase of one product), eight complex1 problems (requiring purchase of multiple products from the same category), and eight complex2 problems (requiring purchase of multiple products from two different categories). Problems requiring purchases exclusively from the clothing or vehicles categories were used in both distinctiveness conditions.

For each distinctiveness conditions, both the simple and complex1 problems were equated across the different categories such that each of the four categories were searched an equal number of times. The complex2 problems were created so that each pair of categories was represented in one problem. Four of the eight simple and complex1 problems could be used in both distinctiveness conditions, but this was the case for only two of the eight complex2 problems. All problems created are listed below broken down by problem type, target type and distinctiveness condition.

DISTINCT PROBLEMS

Simple Information Retrieval Problems:

Explicit:

1. Buy a lamp if it comes in gold. If gold is not one of your options, buy a rug instead.

2. Buy a sander if it weighs less than 50 pounds, if not, buy sandpaper.

Definition:

1. You've just built a picnic table and need to make sure it has an even surface to eat on. Buy an instrument that will help you make sure your surface is not uneven or tilted.

2. You've just moved into an apartment and need something to sleep on. Buy something to serve this purpose, as long as it doesn't exceed your $300.00 budget.

Complex Integration Problems (within one category):

Explicit:

1. Buy a saw, pliers and a screwdriver if their combined shipping cost doesn't exceed $10.00. If it does, buy only the saw and screwdriver.

2. Purchase a couch and a love seat if they both come in the same color (it doesn't matter what color) and their combined cost plus shipping isn't over $500.00. If the cost is too much or colors don't match, just buy the couch.

Definition:

1. Purchase three pieces of furniture for your living room. They must go with your new grandfather clock which is made of mahogany wood (make sure the three items come in mahogany as well).

2. You've just purchased a great new poster that you would like to hang on the wall. It already has a hook attached. Buy the two items you need to hang the poster (your landlord doesn't care if you put holes in the wall). Combined, they cannot weigh more than 10 pounds.

Complex Integration Problems (between two categories):

Explicit:

1. You decide to devote a corner of your garage to a workshop area. Buy a cabinet to hold all your tools. In addition, buy a drill, crowbar and a stool to place in your new cabinet, providing their combined weight isn't more than 20 pounds. If the tools weigh too much, buy all the above items minus the drill.

2. You are responsible for buying your 7-year-old niece her birthday presents. Buy her roller skates, but only if they come in blue. In addition, buy her a desk and chair if their combined shipping cost is under $15.00.

3. It's almost Halloween and you decide to be that weird guy on public television who does the carpenter do-it-yourself show. Buy a wool shirt, a ruler and a pencil to tuck behind your ear. Make sure these items don't exceed $35.00. (You will have to buy a box of pencils).

Definition:

1. You would like to buy your 3-year-old son a present. He's too young for a bicycle, but something as mobile would be perfect. Buy the most

appropriate vehicle for him and a tool to adjust the seat to his height. The price for the two items cannot exceed $70.00.

2. You have signed up for horseback riding lessons. You decide you need some new rugged denim pants to wear on the trial. Purchase the pants. Now you realize your running out of closet space. Purchase some piece of furniture to keep your pants in as long as it costs less than $70.00.

3. The rod that holds up your bedroom curtains is coming loose from the wall. You'll need to buy something for standing on to reach the rod overhead and a tool (other than pliers) you can use to tighten the nuts holding the rod to the wall. The combined cost of these two items cannot exceed $30.00.

OVERLAPPING CATEGORIES:

Simple Information Retrieval Problems:

Explicit:

1. Buy a hockey mask if it comes in your team color - blue.

2. Purchase a cooler for the beach if the price including shipping doesn't exceed $30.00.

Definition:

1. You play tennis frequently and need something to carry cold water to the courts. Buy something that you could use for this purpose, but only if it comes in green to match your tennis outfit.

2. Buy a piece of equipment you can use for baseball. Don't buy the item if it's weight is over 10 pounds.

Complex Integration Problems (within one category):

Explicit:

1. Buy a sleeping bag, backpack and a tent for your trip to the mountains. If the combined price of all three items if over $1,000, then buy the two cheapest items.

2. You have decided it's time to join the fitness craze. Buy barbells and a floor mat if their combined weight isn't over 30 pounds.

Definition:

1. Buy items necessary to build a fire on your camping trip planned for next weekend. In addition, buy another source of light, just in case you luck at fire building is running low. Make sure the combined cost of these items is not over $45.00.

2. You've just purchased a membership at Cheap Workout Fitness Center. The cost of the membership is kept low because the members must bring all of their own equipment. Buy the equipment available for two of your favorite sports, volleyball and basketball. If the combined cost of these two pieces of equipment are over $70.00 purchase the items necessary for golf instead.

Complex Integration Problems (between two categories):

Explicit:

1. You need some miscellaneous items. Buy swimming goggles, a hatchet, insect spray, and a stopwatch providing the combined weight of the four items does not exceed 7 pounds. If the items weigh too much, buy the two lightest items.

2. You have planned a fishing trip for next month. Buy a fishing pole, long underwear, and a jack knife if their combined shipping cost isn't over $8.00. If the cost is too much, buy only the long underwear and the fishing pole.

3. To prepare for your beginning tennis lessons, buy a tennis racquet, sweat pants, and a sweat band, but only if they all come in the same color (you wouldn't dream of not being color coordinated!). It doesn't matter what color, as long as all three items are available in the same color.

Definition:

1. Buy foot gear and equipment necessary participating in track and field events. The combined cost of these items cannot exceed $100.00. If the items cost too much, buy the foot gear and the cheapest piece of equipment (you already own a stopwatch so buy something else).

2. Buy footwear and warm clothing for your backpacking trip next fall. Purchase the footwear and two items of warm clothing good for camping. The combined cost of all three items cannot exceed $115.00.

PROBLEMS FOUND IN BOTH OVERLAPPING AND DISTINCT CONDITIONS:

Simple Information Retrieval Problems:

Explicit:

1. Buy a turtleneck if it comes in red. If red isn't available, buy a raincoat.

2. Buy a unicycle if it's cost including shipping is not over $200.00.

3. Purchase the cheapest mode of transportation you could use on a camping trip. The vehicle must hold more than one person and it must operate on land.

Definition:

1. Buy a water vehicle that will hold more than one person and costs less than $1,000.

2. Your jeans are getting loose around the waist (that new diet is really working). Buy something to hold them up as long as it's under $15.00.

Complex Integration Problems (within one category):

Explicit:

1. Buy a ski jacket and either wool gloves or mittens. To decide whether to buy gloves or mittens, determine what colors the jacket comes in and choose whatever will come in a matching color (it doesn't matter what color you choose, as long as the jacket color and hand covering color is the same).

2. Buy as many of the following as possible, providing their cost doesn't exceed $200.00: bicycle , motor scooter , skateboard
It is best to have as many items as possible, rather than just the most expensive item.

Definition:

1. Purchase three items to wear jogging. Of course, you are quite fashion conscious (in a monochromatic sort of way), these items must all come in the same color - it doesn't matter what color.

2. You have a 1-year-old nephew you like to take for walks in the park. Buy two items you can put him in for your walks. Make sure they don't cost over $150.00 combined.

Complex Integration Problems (between two categories):

Explicit:

1. Buy ice skates, earmuffs and a scarf if it's possible to have all three in the same color. If not, just buy the skates.

Definition:

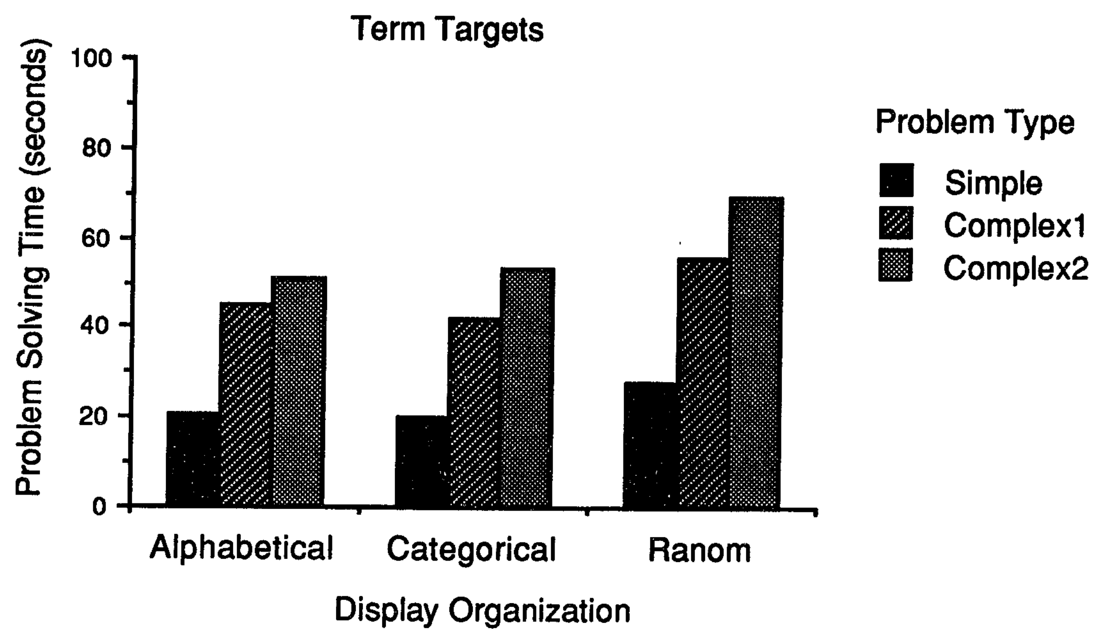
2. Your moving to Chicago, the windy city. Buy two items that were made just for such conditions (windy). Hint, these two items combined shipping cost is $154.00.

Appendix F

Three-way interactions from the Experiment 2 full data

Figure 1

Mean problem solving times across organization, target type and problem type



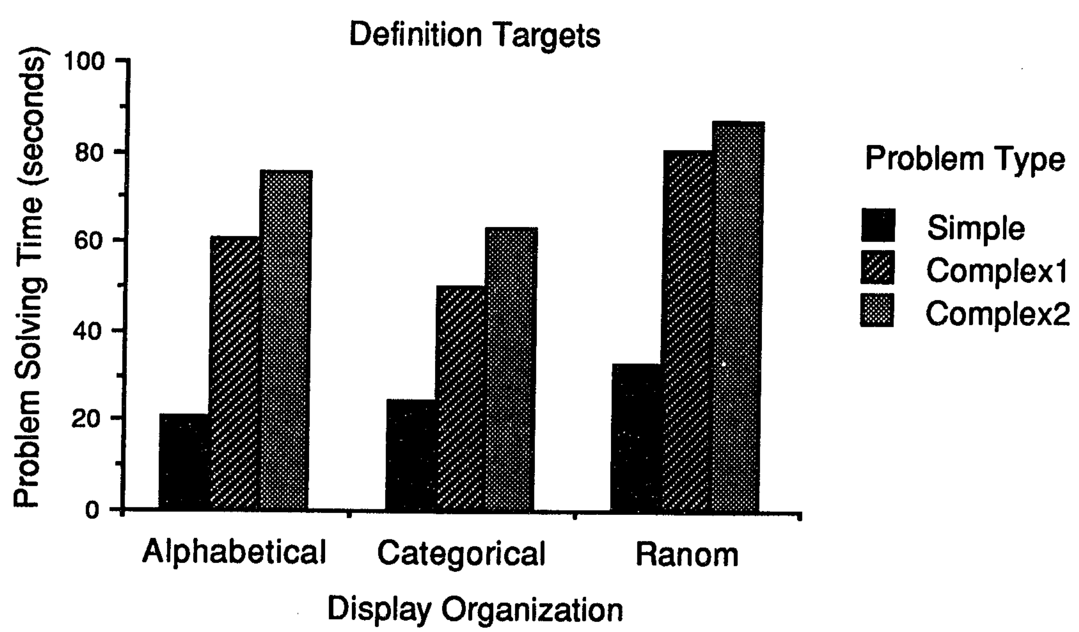
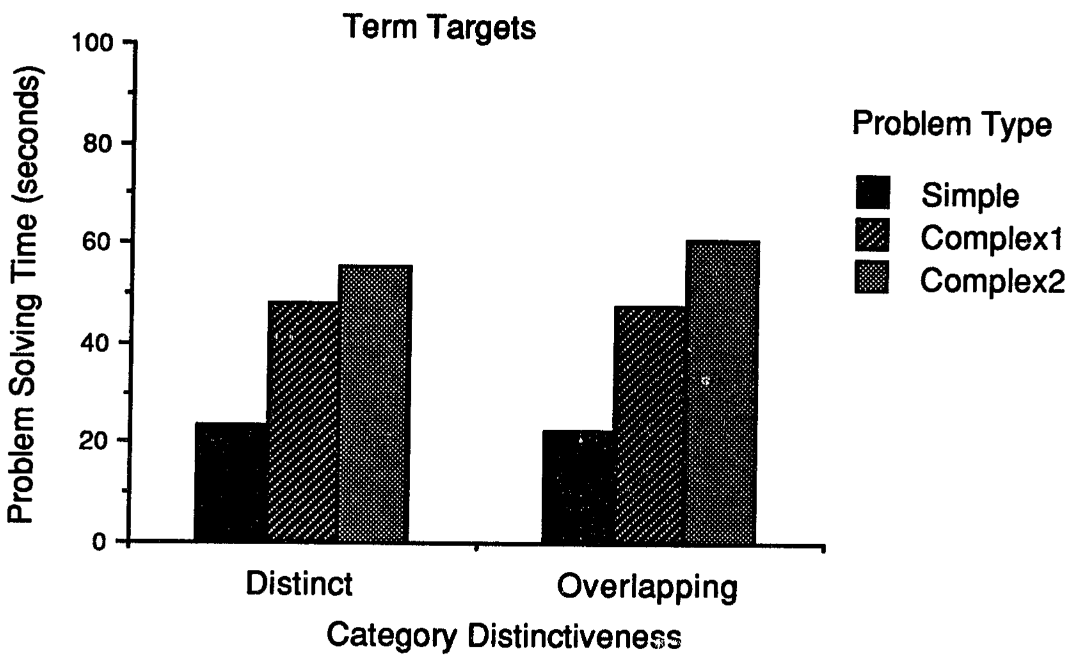
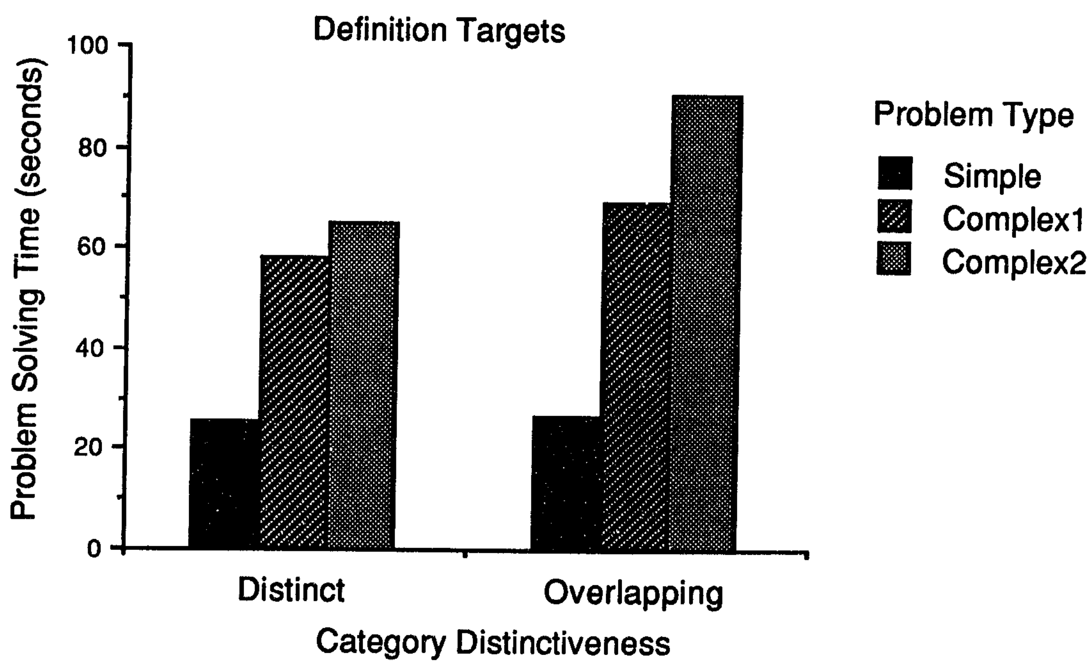


Figure 2

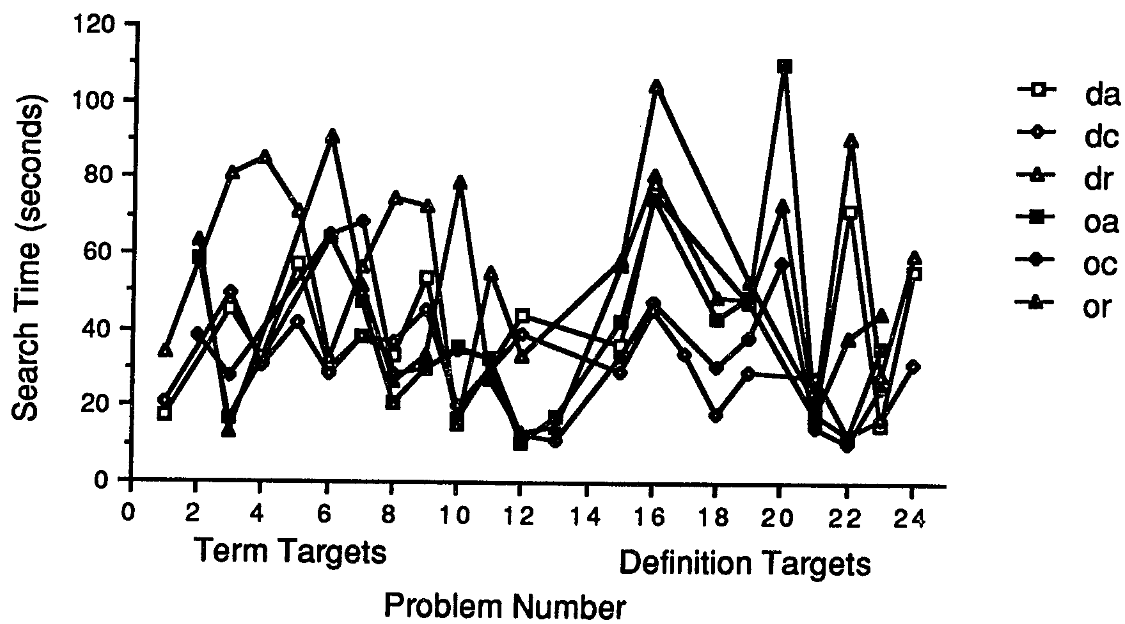
Mean problem solving times across category distinctiveness, target type and problem type





Appendix G

Experiment 2 learning curves across organization and category distinctiveness conditions- Term targets first



Experiment 2 learning curves across organization and category distinctiveness conditions - Definition targets first