An Evaluation of Methods of Indicating Active Words in Hypertext Documents

by

Rani Lea Watkins

Thesis submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

in

Industrial and Systems Engineering

APPROVED:

H. L. Snyder, Chairman

A. G. Hammer

R. J. Beaton

March, 1991

Blacksburg, Virginia

An Evaluation of Methods of Indicating Active Words in Hypertext Documents

by

Rani Lea Watkins

Committee Chairman: Harry L. Snyder
Industrial and Systems Engineering

(ABSTRACT)

Three methods of indicating active (linked) words in hypertext documents were investigated. The Methods consisted of font-manipulation techniques (such as shadowed and outlined text), word-enclosure techniques (variations of a box and pseudo-brackets around the word), and "punctuation-like" icons (located directly above the first letter of an active word). Two Cues were nested within each Method, yielding a total of six Cues. Twenty-four subjects (12 males and 12 females) performed reading tasks and visual search tasks to evaluate each of the three Methods and six Cues. A hierarchical, within-subjects experimental design was used, employing a completely counterbalanced order of treatments.

The study consisted of two experiments. For Experiment 1 (reading tasks), participants read Tinker (1955) passages and identified an inappropriate word in each passage. Passages contained either one of the six Cues or no Cues (control condition). The times taken to read each passage and locate the target word, as well as the number of errors made, were recorded. For Experiment 2 (visual search tasks), participants scanned text fields and located and counted the number of active words contained within each text field; again, the times taken to locate all

active words and the number of errors made were recorded. For both experiments, participants rated Cueing Methods along various dimensions and selected a preferred Cue and Method for the tasks performed in each of the two experiments.

For reading tasks, no single Cueing Method yielded significantly shorter reading times than any other Method. Similarly, no particular Method was preferred by significantly more participants than any other Method. However, user ratings were more favorable for Icons than for Enclosures or Character Styles. In addition, Character Styles were consistently rated as being highly distracting for reading tasks.

For visual search tasks, Character Styles yielded shorter search times ($p < 0.05$) as well as significantly more favorable salience ratings. Participants also preferred the Character Styles significantly more often than either Enclosures or Icons for locating linked words.

The number of errors produced for both experiments was very low (less than 2%) and there were no significant differences in errors across Methods or Cues. This finding is not unexpected considering the low level of difficulty for the reading and search tasks.

Based on the results of both experiments, Icons are the recommended Method for indicating linked words in hypertext documents. Icons provide moderate perceived readability (reflected in subjective ratings, though not in reading times or preferences), whereas Enclosures were not sufficiently salient and Character Styles were perceived to have degraded text readability.

A discussion of a wide variety of techniques which either are currently used in existing hypertext systems or could potentially be used is included in the Literature Review section. The issue of trade-offs between Cue salience and obtrusiveness is addressed in the Discussion.

ACKNOWLEDGEMENTS

The author would like to thank Dr. Harry L. Snyder for his invaluable suggestions regarding this research and his willingness to give the author autonomy in her work. Thanks also to Dr. Robert J. Beaton for his perspective on the study, Dr. Ann G. Hammer for her guidance, support, and flexibility throughout this research, and to the User Systems Engineering Group at Texas Instruments for their help in generating this research idea and for funding my work.

Special gratitude is extended to Isabel Herb and Deborah Cox for their friendship and comic relief during the development of the experimental props, and to Dr. Michael Mohageg for his help in the data analysis stage of this research. I also would like to thank Dr. Richard Nance for shedding light on the elusive Greenhouse-Geisser correction.

Above all I would like to thank Glenn Walthall for his eternal patience and stability, my mother and father for their unwavering support, and Deirdre Royster, who has been a constant source of friendship, advice, and "centeredness" throughout my years at Virginia Tech.

This thesis is warmly dedicated to my grandfather, Col. Robert S. Milner, whose wisdom and love have emboldened me in my academic and personal endeavors.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

The fundamental concept of hypertext was introduced in 1945, when Vannevar Bush envisioned a super-system of information access, called the "memex" (Bush, 1945). The conceptual "memex" system would be surpassed in speed and flexibility only by the human brain but be capable of storing massive amounts of information without the "human" failure of memory decay. This theoretical system is the forerunner of what is now known as "hypermedia," or more specifically, "hypertext."

Eighteen years after Bush's introduction of the memex, Doug Engelbart addressed the issue of human-computer interaction (Engelbart, 1963). He proposed the H-LAM/T (Human using Language, Artifacts, and Methodology, in which he is Trained), a system that "included the human user as an essential element: the user and the computer were dynamically changing components in a symbiosis which had the effect of 'amplifying' the native intelligence of the user" (Conklin, 1987, p. 11).

Over the next five years Engelbart's proposed system evolved into the NLS (oN Line System) (Conklin, 1987). For this new system, Engelbart invented the mouse as an input device, providing convenient user interaction with the computer (Conklin, 1987). In 1968, he introduced the NLS, now recognized as the first hypertext system, at the Fall Joint Computer Conference in Montvale, New Jersey. Engelbart demonstrated NLS by collaboratively writing a document with a colleague located 500 miles away (Engelbart and English, 1968). Since the introduction of these early computer systems several other hypertext systems have been developed, including a more sophisticated version of the NLS system, now called the Augment System (Engelbart and English, 1968). The Augment System includes many types of computer-supported

communication and tools "for document production and control, organizational and project information management..., and software engineering" (Conklin, 1987, p. 12).

Ted Nelson, who receives credit for coining the term "hypertext" (Conklin, 1987), built another hypertext system called Xanadu Hypertext System (Nelson, 1980). Nelson's (1980) objective was to simplify the cumbersome task of conducting literature searches within any profession using his conceptual Xanadu Project. His system was designed to facilitate the retrieval of articles of particular interest to researchers in a given discipline (Nelson, 1980). Numerous other hypertext systems have been developed since the early 1980s (see Conklin, 1987).

One advantage of hypertext systems is the ability to navigate in a non-linear fashion; the ability to skip over extraneous information enables the system to sort through disparate text for the user, making it easier for the user to locate specific information. The technology underlying hypertext systems enables words, phrases, articles, or any other identified "node" of information to be linked to other words, phrases, articles, etc. These links are usually accessible to the user through the mere click of a mouse; once a link has been activated, the user has accessed some node of information that is relevant to the source node the user initially clicked on. The link structures can be modified by users, allowing the information to be "custom linked" to better serve the individual user's needs.

Although hypertext systems have existed for the past 20 years, no consistent, standard format has been adopted for indicating an active word or group of words (i.e., that two nodes of information are linked). A number of cues traditionally used in typesetting, such as bolding, italics, and underlining, have been used. Other more innovative cues also have been used, such as enclosing the active word in a box (Halasz, Moran, and Trigg, 1987). The fact that "links may be depicted in many ways, from highlighted text to button-like objects" (Irish and Trigg,

1989) raises a human factors issue; the lack of standardization, or even guidelines, for designing cues introduces opportunity for inconsistencies across various hypertext systems.

Consistency across computer applications has considerable utility for the transfer of learning from one application to another. Since current hypertext systems have no standards to follow for indicating linked text, users must learn how to discern active words for each of these systems, as opposed to looking for one particular target across all applications. Although system designers have suggested that the issue of indicating active words deserves some attention (Marchionini and Shneiderman, 1988; Shneiderman, 1987), empirical human factors research needed for creating design guidelines has not been devoted to this area of hypertext.

The intent of this research was to evaluate the effectiveness of three methods of designating linked text; the specific cues within the three methods also were evaluated, but the methods are of primary interest. Prior to selecting cueing methods for this study, three factors related to cue effectiveness were identified. To provide an effective cue for active words, these factors must be considered: (1) the cue must be unique to hypertext; (2) the cue must not distract the reader and, thus, degrade the overall readability of text; and (3) the cue must be salient and easily detected. Each of these factors was considered when the various cues were designed within each method to designate linked text.

The first factor, selecting a cue that is unique to hypertext, is addressed in the following section of this thesis. With regard to the second and third factors, cueing methods were evaluated using participants' performance in two separate, controlled experiments. User preference ratings were also used to evaluate each method.

LITERATURE REVIEW

Review of Inappropriate Alternatives

A paper presented at the Hypertext '89 seminar addressed two issues in indicating linked words within hypertext systems. One issue concerned the lack of standardization among some currently used methods of indicating linked words; the other issue involved a critical assessment of the appropriateness and effectiveness of cueing methods that were either in use or could potentially be used (Evenson, Rheinfrank, and Wulff, 1989). According to the authors, most systems use bold face, italics, or underlining to indicate active words. The method used is mentioned at the beginning of the document to inform the user what to look for. This general method is problematic because these typographical cues, which began as font styles, traditionally indicate such content as emphasis and titles. (For a history of bold and italic type forms, see Goudy, 1963.)

Bold, italics, and underlining. Bold and italic type have been used for organizational purposes, emphasizing meaning, and for portraying document and literary titles (Evenson et al., 1989). Underlining began as a way of altering the form of pre-written text to "highlight" the information (Evenson et al., 1989). This handwriting convention has been adopted by the on-line domain of textual information. As Evenson et al. (1989) pointed out, "it may not only be difficult for users to remember a new application of an old convention, but it may also be difficult for the designers themselves" (p. 86).

Blinking text. Another option for formatting linked words is temporal blinking; however, Grice (1989) warned that "it should not take much thought to realize the disastrous effect [that] blinking could have on someone trying to read the text" (p. 42).

Upper- and lower-case letter combinations. Because the case of the letters (upper-case and lower-case) has been used to indicate emphasis (e.g., a message written in all capital letters generally signifies that the message is important), this method fails to provide enough distinction between linked and non-linked words. The text could inherently contain words or sentences written in all capital letters, thereby making the use of capital letters a less unique cue. Therefore, this method is not a viable alternative.

Color coding. Color coding also has been used to indicate linked text. For example, the Hyperties system, developed at the University of Maryland, allows the user to choose from a pallet of colors to indicate active text (Shneiderman, 1987). Another hypertext system, gIBIS, uses color to indicate node- and link-type information, as well as node-state information (Begeman and Conklin, 1988). However, since color coding requires the use of a color monitor and is useless for systems employing monochrome monitors, color coding is not considered in this research as a viable method.

Reverse video. Reverse video is a potential cue for dynamic text; however, it is widely used by on-line systems for indicating that a word has been selected for some manipulation (for deletion, for a change in font size, for cut-and-paste operations, etc.). Therefore, reverse video is not considered in this study.

Spacing cues. Innovative ways of organizing text on the screen is a possible cue for indicating linked words (Evenson et al., 1989). Printed text is usually organized into paragraphs separated by one line or offset by indenting the first word. This spatial arrangement of text conveys a transition to the reader. In the same way, spacing cues could be used to convey linked text. However, there are not many ways in which text can be spaced on the monitor without sacrificing the organizational structure of the information.

The Knowledge Management System (KMS) incorporates this strategy into its design to indicate the *type* of link (annotation or tree-item link) associated with the source node of

information by listing the linked words in a column either on the right or the left side of the screen (Akscyn, McCracken, and Yoder, 1988). Spacing cues might be a viable option for outlines, graphics, and other less structured text form; however, for paragraphs of reading material there is little margin for spacing flexibility. Another problem with this method is that, given the limited amount of space on a monitor screen, designers do not always have sufficient room to capitalize on this technique. In addition, document formats and authoring styles can vary widely, further complicating the added use of spacing cues for indicating links.

Viable Alternatives

As indicated above, there are several problems associated with existing techniques for indicating linked text in hypertext documents. Certain alternative methods also have shortcomings as previously discussed. As a result, these problematic methods are not explored in the present study. Instead, alternative methods utilizing character manipulation (Shadowed and Outlined Character Styles), word enclosures (pseudo-boxes or pseudo-brackets around linked words), and icons (which resemble punctuation marks) are investigated.

Character-style cues. Shadowed or outlined typestyles are not conventional text enhancers and, therefore, these character formats may be viable linked word cues for hypertext applications. Furthermore, these two techniques suggest that there is "something behind the word," thus acting as natural and intuitive cues for portraying linked text (Evenson et al., 1989). There is, however, a possibility that shadowing and outlining will distract the reader and impair the comprehension of the text. Specifically, shadowed or outlined character styles may disrupt the usual flow of text and degrade the readability of the document (Evenson et al., 1989). This possibility can be evaluated experimentally.

Word-enclosure techniques. As was previously mentioned, some hypertext systems (e.g., Xerox PARC's NoteCards) have used the word-enclosure technique to indicate a linked

word. A problem with this method is that *plain boxes* or *plain circles* are used. These boxes and circles have traditionally been used in documents such as job applications, tax forms, questionnaires, and in word processing "spell-checking" utilities.

Punctuation marks also have been used as a word-enclosure technique. In the PlaneText system, developed by Eric Gullichsen in the Software Technology Program at Microelectronics and Computer Technology Corporation, linked words appear in curly brackets ({ }) (Conklin, 1987). However, punctuation marks are not unique to hypertext. (See the following section on "punctuation-like marks.") In keeping with the first criterion, *variations* of a plain box and of brackets will be explored as word-enclosure techniques in this study.

"Punctuation-like marks." Evenson et al. (1989) suggested that punctuation marks have traditionally been used to enhance the meaning of text and that using some similar marks is a viable means for indicating linked words. The Knowledge Management system (KMS) uses this type of link indicator. For this system, a small circle to the left of the linked node represents the link (Akscyn et al., 1988).

Similarly, the use of icons resembling punctuation marks becomes a prospective method of indicating linked words. As Evenson et al. (1989) suggested, a distinctive symbol for hypertext could represent a link, in a manner analogous to the use of the symbol "$" to represent money and "%" to represent a percent value. KMS currently uses the "@" symbol for "at" to indicate a link type (Akscyn et al., 1988).

Related Issues

Designating link type. Using a cue to represent the *type* of link that exists between two words or groups of words is another issue in hypertext. For example, a selected linked word could direct the user to a definition, related information, more detailed information, or more general information. Letting the user know the type of information that a word is linked to is a

helpful tool. The most widely reported difficulty with using a hypertext document is that participants do not know "what was behind the word" (Gordon, Gustavel, Moore, and Hankey, 1988).

Link-type cues are incorporated into Nelson's Xanadu Hypertext System to differentiate among "jump-links," "quote-links," "correlinks" and "equilinks" (Nelson, 1980). **Guide 2®** also provides users with four different symbols, each representing a different type of link (e.g., expansion, reference, note, and command buttons) (**Guide®** Operating Manual, 1988). This feature is designed to indicate the nature of the information of the destination node, thereby allowing the user to choose to follow or not to follow that link.

KMS uses link-type symbols as well as symbols to differentiate among two pairs of link types. One pair of link types can be differentiated by circles to indicate the type of link; a hollow circle represents a link to another "frame" (frames contain text, graphics, or image items), and a filled circle represents a link to a program (which is invoked when the linked word is selected) (Akscyn et al., 1988). Furthermore, the symbol @ indicates that a link is to annotation items rather than to tree items, which are frames at the next lower level of the hierarchy (Akscyn et al., 1988). Spacing cues are also used to distinguish the annotation items from the tree items; annotation items are listed on one side of the screen and tree items are listed on the other side.

Several hypertext system designers have promoted this strategy in hypertext documents (Akscyn et al., 1988; Nelson, 1980). However, this approach introduces problems of cognitive overhead. As the number of symbol cues to be memorized increases, so does the difficulty of using the system. Novice users in particular have no preexisting cognitive model of the system and are less familiar with the cues used in that system.

From a human factors engineering perspective, relying on the user's memory to discriminate among different symbols is not good design. This assumption could result in high

rates of human error and back-tracking through the system, especially for systems with numerous link types. Trigg (1983) identified more than 80 types of links in the Textnet hypertext system. Even experienced users of this system would likely have difficulty keeping track of 80 different symbols for 80 different link types (see Conklin, 1987).

A possible way to avoid cognitive overhead is to provide a menu for each active word. Once a linked word has been selected, a menu could list all paths and link types available to various destination nodes. The user could select the desired path without having to explore all paths to find the most appropriate one. If no path is appropriate, an available menu option would return the user to the original place in the document.

To summarize, indication of informational paths is helpful to users of hypertext systems. However, assigning various types of links to different icons places the responsibility of remembering and differentiating among icons on the user. A menu-driven method may reduce the user's cognitive load.

Link source and destination distinction. Another issue in navigating through hypertext documents is the distinction between "link source" and "link destination" (Conklin, 1987; Yankelovich, Smith, Garrett, and Meyrowitz, 1988). Conklin (1987) specified that a link source is a "point," and the link destination is a "node." Since effective designation of linked words could also effectively designate link source *and* destination, a distinction is not made between the two in this study. All stimuli are considered "source" links.

However, this issue requires further study. On one hand, a link indicator used for source links should be compatible with, yet easily distinguished from, that used for destination links. On the other hand, cue compatibility for source and destination links might not be a crucial criterion since the destination link could be made highly salient, perhaps using reverse video. Reverse video also could indicate the extent of the destination "block" (Yankelovich et al., 1988). The user could turn off the salient highlighting if it is distracting. A toggle mechanism

would be highly effective here since it would allow the user to refer back at will to what is included in the destination block. Using this strategy, a source node could be identified using an effective method, the destination node could be made highly salient to show its location and extent, and the user could turn off the highlighting of the destination node for reading ease.

The extent of a node also is not incorporated into this study. Only single words were designated as linked words. However, in actual hypertext systems phrases, titles, sentences, or even entire paragraphs can be source nodes.

Research Objectives

The objective of this research is two-fold: to offer guidelines in selecting cueing methods for linked words and to provide a basis for further research in cueing annotation for hypertext documents. In this study the effectiveness of several alternative methods of indicating linked text in hypertext documents is evaluated. An assumption underlying the work is that cue effectiveness can be determined from both objective and subjective data. The objective data are performance measures regarding the readability of text and salience of the cues. The subjective data are participant feedback from user preference questionnaires, which contain various dimensions for rating cues on a seven-point scale, and information obtained during the debriefing.

Only character-style manipulation, word-enclosure techniques, and icons (as punctuation-like symbols) are used as methods in this research. The actual *cues* are evaluated primarily to determine how task performance varies from one cue to the second within each method. As a result, the recommendations of this study pertain to general methods or techniques rather than to specific cues.

GENERAL METHOD: EXPERIMENTS 1 AND 2

Both Experiments

This study consists of two separate experiments, both of which were conducted using a Macintosh computer. In Experiment 1, user acceptance of each Cueing Method is examined in the context of text readability. This experiment was designed to determine the effectiveness of each Method with regard to reader distraction or degradation of text readability. In Experiment 2, a visual search task was used to evaluate the salience of each Method. The Cues used within each Method are presented in Figure 1.

Lighting conditions. The illuminance in the room, measured on a plane approximately perpendicular to the computer's monitor screen but at a distance of about 120 cm from the screen, was 25 lux.

Screen luminance. The white background of the screen had a luminance of 38.4 candelas per square meter. A black square covering roughly one square inch on the screen had a luminance reading of 1.5 candelas per square meter. The character luminance modulation was therefore 0.92, or about a 26:1 contrast ratio.

Participant Recruitment

Flyers were posted at many locations frequented by college students on the campus of the Virginia Polytechnic Institute and State University. The flyer advertised the need for participants in this study at a payment of five dollars per hour. Interested persons called to inquire and were screened on several pre-established criteria. Those who met the screening criteria were scheduled for an eye examination and experimental session.

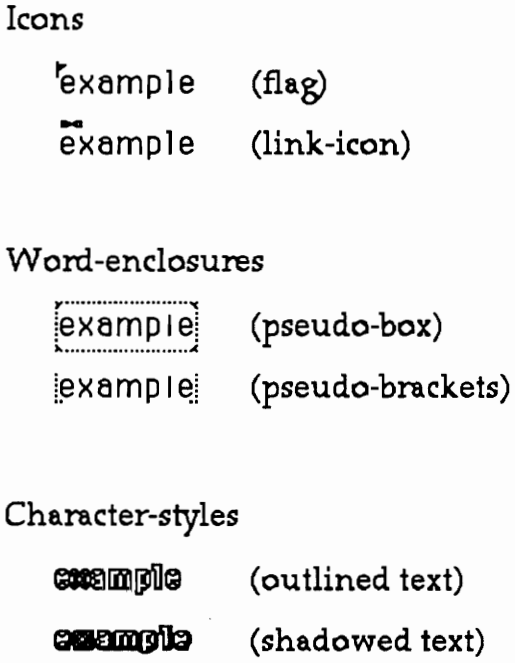


Figure 1. Cues used within each Method for Experiments 1 and 2.

Screening of Participants

Participants were screened for prior experience with hypertext systems, thus preventing any transfer of learning effects to using the Cues in this study. Only persons who spoke English as their primary (native) language were used as participants. All participants were used in both experiments.

EXPERIMENT 1: METHOD

Participants. Twenty-four participants (12 male and 12 female) were used for this experiment. All participants were college students between the ages of 18 and 25. Each participant was required to pass an Orthorater test for near and far visual acuity of 20/22 or better.

Materials and apparatus. An Apple Macintosh IIcx computer and 13-inch Apple RGB monitor were used to present the stimulus screens containing Tinker passages (described in the next section). The monitor displayed 640 horizontal pixels by 480 vertical lines and an active video area of 235 mm (horizontal) by 176 mm (vertical).

The screens (which were presented in black and white) were actually "cards" in a Hypercard stack. The computer's clock was used to time participants in this task. A data file was automatically made for each participant and each trial for that participant. At the completion of each trial, the data file was closed and saved under a label indicating the experiment, the participant's number, and the trial (condition).

Reading task. A reading task was used to evaluate the readability of each Cue and Cueing Method. Reading times for Tinker (1955) passages containing embedded linking notation (the time taken to read a passage and locate the inappropriate word in that passage) were recorded for this task. Each Tinker passage consisted of four lines of text and usually contained about 30 words. One word did not logically correspond to the context of the text. An example of a Tinker passage is presented in Figure 2. Passages were presented one at a time. Each participant was instructed to read each passage, find the inappropriate word, click on a button to go to a blank screen, and then verbalize the word.

Tinker passages. Eighty Tinker passages were used to evaluate the readability of passages containing embedded linking notation using each Cue and Cueing Method. Sixty of the

Please be careful not to disturb the baby for he has just fallen asleep after crying steadily for four hours. The poor little thing must have sung himself to sleep.

*

Figure 2. Sample Tinker passage from Experiment 1.

* The inappropriate word is *sung*.

passages contained three, five, or seven "linked" words (words with associated Cues). These words were not actually linked to other text; they merely possessed Cues as if they were. Twenty of the passages consisted of plain text with no cued words.

Derivation of Tinker passages. The Tinker passages used in Experiment 1 were obtained from the first two groups of 50 (there are 10 groups of 50 in all, totaling 500 passages) from the *Examiner's Manual for Tinker Speed of Reading Test* (1955). To keep the number of words per passage consistent, passages containing roughly 30 words were sought. Since Tinker passages are standardized, each one was methodically, as opposed to randomly, assigned to a specific Cue (and therefore Cueing Method). This assignment was done in the following way: passages were assigned to a Cue condition roughly in the order that they appeared in Tinker's (1955) list of passages. An effort was made, however, to assign an equal number of 30-word passages to each Cue, resulting in some deviation from this method when necessary. Participants read the same groups of passages for each condition.

Tinker passages also were methodically, as opposed to randomly, assigned to "number of linked words" conditions. This was done in the following way: once the passages were assigned to a Cue condition they were considered to be in the same order (first through tenth) in which they were assigned to that condition. The first four passages were then assigned to the "five-word" condition. The next three passages were assigned to the "seven-word" condition, and the last three were assigned to the "three-word" condition. The "number of linked words" conditions were the same for all participants.

Random assignment of words in Tinker passages to possess linking notation. Prior to data collection a random number generator was used to assign words that would contain Cues. For each passage the exact number of words in the passage was determined. A list of numbers between this number (of words in the passage) and one was then generated in random order for each passage. The first three, five, or seven numbers in the list (depending on the "number of

words" condition for that passage) were used to assign Cues to words. For example, if a passage contained 30 words, the numbers from 1 to 30 were randomized. For a "three-word group," the first three numbers to appear in the list corresponded to three words in the passage. Therefore, if those numbers happened to be 23, 19, and 4, then the fourth, nineteenth and twenty-third words in that passage received Cues and posed as linked words. Occasionally, words such as "a" and "the" were randomly selected as linked words. These seemingly unlikely candidates for linked words were allowed to pose as linked words for two reasons. One reason is that these words actually could be linked to some related informational node (e.g., they could be linked to interpretations within foreign language tutorials). A second reason is is that maintaining consistency provided objectivity in linked word assignment. This process was followed for each passage.

Font used in passages. Twelve-point (7 x 9 dot matrix for upper-case letter N) Geneva font was used for Tinker passages. This sans serif font was chosen because Cues were assumed to be more apparent to participants if they were embedded in a plain, unaccentuated font. At a viewing distance of approximately 40 cm, the angle subtended at the eye of an upper-case letter was roughly 26 minutes of arc. The 12-point Geneva font was also used in Experiment 2.

Experimental Procedure

Upon arrival each participant was seated at a desk and asked to read and sign a consent form (consenting to taking the Orthorater exam) before being tested for near and far acuity. The consent form for the vision test is in Appendix A. Once the vision tests were completed, participants who passed the test were asked to read a summary of hypertext, entitled "Hypertext: A Brief Description," to familiarize themselves with the concept and technology of hypertext, and an introduction to the study, which explained why the research was being conducted and instructed them on the tasks they would be performing (Appendix B).

After reading this material each participant was asked to read and sign a second consent form (Appendix A) to participate in the actual experiments. After signing this second form, each participant read supplementary instructions for Experiment 1 (Appendix B) and then began the tasks on the computer.

Task description. Each participant viewed a series of 10 passages for each Cue condition. The control condition consisted of two groups of 10 passages that had no "linked" words. The text fields appeared one at a time on HyperCard® cards; each group of 10 cards made one stack for the corresponding Cue condition.

A button labeled "Next" was located below each passage. The participants clicked on this button to go from a stimulus card (containing a passage) to a non-stimulus card with the following message: "Which word does not belong in the passage?" The participant was instructed to read the passage on each stimulus card, locate the word that did not logically fit in the passage, click on the button labeled "Next," and verbalize the inappropriate word at the message prompt on the following non-stimulus card.

A "Next" button was also located in the same place on the non-stimulus cards. This button was clicked to go to the following stimulus card containing the next passage. In this way, participants responded to a series of 10 passages for each "Cue condition" stack.

The time taken to locate the inappropriate word for each passage was recorded (using the Macintosh's clock) for analysis. For each passage, the clock commenced timing upon presentation of the screen containing the passage (elicited by clicking on the "Next" button from the previous screen), and stopped when the participant hit the button again to go to the succeeding blank screen. Reading times for standard passages containing embedded linking notation were considered to indicate the readability of each Cueing Method. The number of incorrectly identified target words (errors) was also recorded across Cues and Methods.

Practice trials. Participants were given a practice trial (consisting of three practice screens) prior to each block of experimental trials (employing the same Cue). The practice trials served several purposes. The first practice trial exposed the participant to Tinker passages, and the participant was given feedback regarding his or her understanding of the reading task. Each practice trial also familiarized the participant with the Cue that was to be used for the up-coming experimental trial. The same three passages were used in all practice trials; the Cue used in each practice trial depended on which Cue was to be used in the following experimental trial.

User preferences questionnaires. Each participant was asked to complete user preferences questionnaires. Two types of questionnaires were used. The first type was an absolute-rating questionnaire, which was filled out immediately following completion of the trials for the second Cue within each Cueing Method. The second type was a relative-rating questionnaire, which was filled out upon completion of each of the two experiments. Participants rated each Method, using a Likert scale, on various dimensions. Subjects were also asked to choose a preferred Cue within each Cueing Method, and an overall preferred Cue and Cueing Method, for those tasks performed in Experiment 1. The questionnaires are presented in Appendix C.

Design of the cards and instruction. Figure 3 presents a diagram of the path that participants took through the conditions in this study. Note that the same basic procedures were followed for Experiments 1 and 2, and this diagram is representative of both experiments. The Box Cue is used in the diagram, but any other Cue could have been used to illustrate the process.

Participants began by clicking on the "Begin" button located on Card 1 (in Figure 3). The next card they saw was one like Card 2, although the Cue condition varied. This card provided information that the participants would begin the practice trial. Participants would proceed

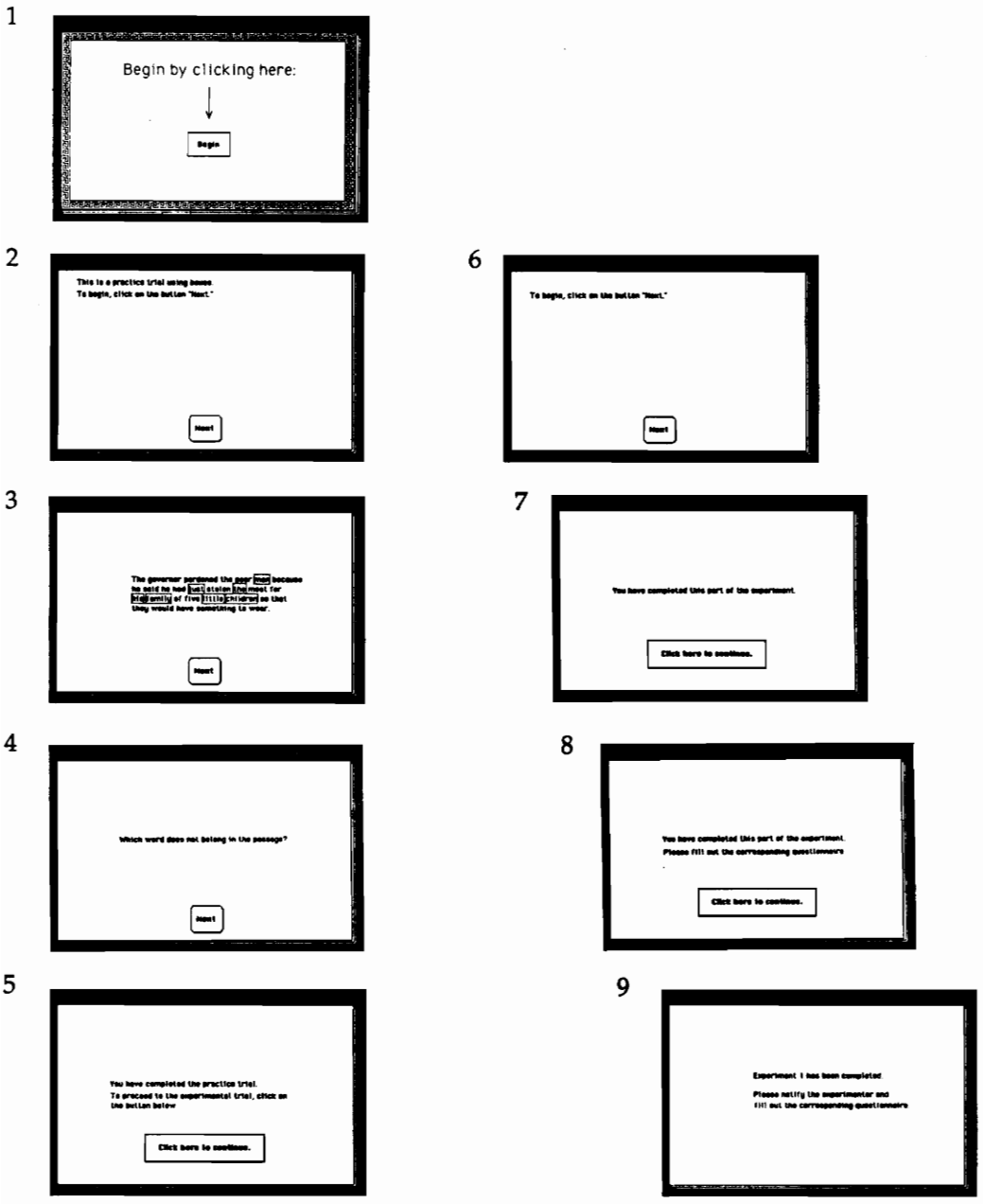


Figure 3. Diagram of the path taken by participants through Experiment 1.

to the following card (Card 3) to view the first passage. After clicking on the "Next" button at the bottom of Card 3, Card 4 appeared and they verbalized the target word (or the number of linked words counted for Experiment 2). Card 5 indicated that the practice trials had been completed and that participants could proceed to the experimental trials by clicking on the button located at the bottom of the card.

Card 6 provided instruction for beginning the experimental trials, and participants passed through the cards within the experimental trials in the same way they had for the practice trials. Card 7 indicated that the experimental trials had been completed, and participants clicked on the button "Click here to continue" to proceed to the practice trials using the other Cue within the same Method. After completing the second experimental trials, participants were notified (Card 8) that they had completed all trials for the first Cueing Method and instructed to fill out the corresponding questionnaire for the passages they had just viewed.

After completing all trials for the remaining Methods, and for the Experiment 1 control condition as well, Card 9 indicated to the participants that they had completed Experiment 1. At this point, participants read the brief paragraph containing instructions for Experiment 2, and the experimenter opened the "Begin" file for Experiment 2. The procedures and cards for Experiment 2 were almost identical to those for Experiment 1.

Experimental Design

This experiment incorporated a within-subjects hierarchical design, consisting of the factor of Cueing Method and a second nested factor of specific Cues within each Method. The experimental design is shown in Figure 4.

Independent variables. Cueing Method (M) consists of four levels or conditions: Icon Cues, Word-Enclosure Cues, Character-Style Cues, and No Cues. Two Cues (C) are nested

	M1		M2		M3		M4	
	Icons		Enclosures		Character Styles		Control - No Cues	
	C1	C2	C3	C4	C5	C6	C7	C8
	Link	Flag	Box	Brackets	Outlined	Shadowed	No Cues 1	No Cues 2
Subjects 1								
2								
3								
4								
5								
6								
7								
8								
9								
10								
11								
12								
13								
14								
15								
16								
17								
18								
19								
20								
21								
22								
23								
24								

Figure 4. Experimental design for Experiment 1.

within each Method, resulting in a total of eight Cues. For the Icon Method, Link and Flag Icons were used; for the Enclosures Method a variation of a Box and a variation of Brackets (psuedo-brackets) were used; for the Character Styles Method, Shadowed and Outlined text were used; and for the control condition, no cues (plain text) were used.

Dependent measures. Two dependent measures were employed: time to find the inappropriate word within passages and the number of errors for each Cue and Method.

Order of treatments for Cueing Methods. There are 24 possible orders for presenting the four Method conditions. One participant was assigned to each order, yielding a sample size of 24 with all possible orders represented in a completely counterbalanced design. The participants were assigned to the conditions in the order that they were scheduled for an experimental session.

Order of treatments for Cues. The particular Cue used (for instance, shadowed versus outlined text) was changed after every 10 trials for all experimental conditions. For instance, the first trial of a condition consisted of 10 passages employing one particular Cue within a Method. The following 10 trials contained passages using the *other* Cue within that same Method. Participants 1-12 received the Link Icon first and the Flag Icon second within the Icons Method, the Box first and the Bracket second within the Enclosures Method, and the Outlined text first and Shadowed text second within the Character Styles Method. Subjects 13-24 received Cues within each Method in the reverse order (Flag Icons first and Link Icons second within the Icons Method, etc.).

Data Analysis

Time to locate target word. The time it took for participants to read the passages and find the inappropriate (target) word was averaged for each participant across the 10 trials within each Cue condition. Since all participants received the same number of passages with

three, five, and seven cued words, task performance times were averaged across all passages within each Cue condition for each participant (i.e., no differentiation was made regarding the "number of cued words" groups since the three groups were equally represented within each of the eight Cue conditions and, hence, four Method conditions).

Mean reading times for Cues and Methods were analyzed using an ANOVA. The Newman-Keuls comparison test was used to indicate significant differences among the Cueing Methods and Cues as well as differences between control and experimental conditions. Comparison test results with $p < 0.05$ were determined to be significant.

Density of linked words. To investigate the point at which the readability of text is significantly degraded by the number of linked words it contains, an ANOVA and subsequent simple-effect F-tests and Newman-Keuls comparison tests were conducted. For each participant, three trials from the No Cues Method were randomly assigned to each of the six Cues, totaling 18 trials from the control condition. The two random points remaining in each participant's No Cues condition were removed from the data set. Each participant's resulting data set for each Cue contained three passages from the control, 3-word, and 7-word conditions, and four passages from the 5-word condition. The cell frequencies are therefore unequal but proportional. The experimental design for the Density evaluation is presented in Figure 5.

Number of errors. The numbers of targets missed within each Cueing Method and Cue were totaled across all trials for each participant. These totals were then analyzed using ANOVA procedures.

Subjective ratings. The rating data obtained from the user-preferences questionnaires were analyzed using separate single-factor ANOVA procedures, and post-hoc comparison tests were used to probe all significant effects that were found. Again, comparison test results with $p < 0.05$ were determined to be significant. Note that these data were transposed to place the "good" ratings consistently toward the lower end of the scale and the "poor" ratings toward the

			M1		M2		M3	
			Icons		Enclosures		Character Styles	
			C1	C2	C3	C4	C5	C6
			Link	Flag	Box	Brackets	Outlined	Shadowed
Ss 1-24	D1	0-word (Control)						
Ss 1-24	D2	3-word						
Ss 1-24	D3	5-word						
Ss 1-24	D4	7-word						

Figure 5. Experimental design for the Density effect, Experiment 1.

higher end of the scale. The anchors for the items from the original questionnaires had been systematically reversed, with a ranking of seven being optimum for some items and a ranking of one being optimum for others, to prevent subjects from patterning their rating responses. The transposition of the data for analysis resulted in a consistently good rating towards "one," and a consistently poor rating towards "seven."

User preferences. Included in the questionnaires were opportunities for participants to select a preferred Cue within each Method (on the Absolute Rating Questionnaires) as well as an overall preferred Method and Cue (on the Relative Rating Questionnaires). These preferences data were analyzed using the Binomial Test and Cochran Q Test (Siegel and Castellan, 1988).

EXPERIMENT 1: RESULTS

Reading Times

The ANOVA summary table for reading times is shown in Table 1. Method and Cue effects are both statistically significant. For Experiments 1 and 2, the values reported in the ANOVA tables reflect the number of *ticks* that had passed before the participants located the target word. (A tick is 1/60 second.) However, the Newman-Keuls tables report reading time in seconds.

Correction for sphericity. Because the experiment consists of two within-subjects factors, a correction for violations of sphericity (Winer, 1971) was calculated for significant main effects. SAS (1985) was unable to compute an epsilon (ε) value (Greenhouse and Geisser, 1959) due to the nesting of Cues. Therefore, a worst case correction for degrees of freedom for Methods was assumed with epsilon at its lower bound: $\varepsilon = 1/(q-1)$, where q is the number of levels of the within-subject factor. The resulting degrees of freedom for Methods (assuming ε is at its lower bound) are 1 and 23. Even with these conservative corrections for sphericity, the Method main effect is still significant at $p < 0.05$. A value for ε was calculated for Cues, and the adjusted degrees of freedom still indicated a significant main effect ($p < 0.05$).

Table 2 presents a Newman-Keuls post hoc analysis (Winer, 1971) across Methods. There are no significant differences among the Cueing Methods (these means are all "joined" by the same bar). However, reading times for the control condition (in which Tinker passages contained no linked words) are significantly shorter (at $p < 0.05$) than those of any of the experimental groups containing linked words.

Table 3 presents the results of a Newman-Keuls comparison test across the nested Cues. While there are no significant differences between the reading times of passages incorporating

Table 1. ANOVA Summary Table for Reading Times

Source of Variance	df	ε*	MS	F	p
Between-Subjects					
Subjects (S)	23		60986.236		
Within-Subjects					
Method (M)	3		31624.347	6.78	< 0.05***
Cue (C/M)	4	0.5997	5826.281	4.32	< 0.05**
M X S	69		4667.253		
C/M X S	92		1347.368		
Total	191				

* Greenhouse and Geisser (1959) epsilon correction for degrees of freedom
** denotes an adjusted p-value
*** significant at $p < 0.05$ with epsilon at its lower bound

Table 2. Newman-Keuls Comparison of Reading Times across Methods

Method	Mean Time (seconds)	
Icons	6.63	|
Enclosures	6.62	|
Character Styles	6.62	|
No Cues (control)	5.77	

* A vertical bar indicates that there are no significant differences among the means to the left of the bar.

Table 3. Newman-Keuls Comparison of Reading Times across Cues

Cue	Mean Time (seconds)	
Shadowed text	6.88	|
Link	6.78	|
Brackets	6.69	|
Box	6.56	|
Flag	6.48	|
Outlined text	6.36	|
No Cues group 1	5.97	
No Cues group 2	5.57	

* A vertical bar indicates that there are no significant differences among the means to the left of the bar.

different Cues for linked words, reading times are significantly shorter for both control groups than for any of the experimental groups ($p < 0.05$). Strangely, reading times for the second control group are significantly shorter than those for the first control group ($p < 0.05$).

Density of linked words. An ANOVA summary table is presented in Table 4. Method and Cue results are not pertinent in this analysis since their effects previously have been determined and compared with a control condition. The main effect of Density and the interaction of Method X Density and Cue X Density are of interest.

The main effect of Density is significant at $p < 0.0001$. The Newman-Keuls comparison test indicates that there are no differences among the reading times for the experimental conditions in which passages contain three, five, or seven linked words. However, all three experimental conditions produced significantly longer reading times than did the control condition. The results of these comparison tests are presented in Table 5.

The Methods X Density interaction (illustrated in Figure 6) is significant ($p < 0.05$) with interpolated, as opposed to rounded, degrees of freedom. Therefore, simple-effect F-tests and Newman-Keuls comparison tests were conducted to investigate any significance within a given level of each of the two factors. Tables 6 and 8 present the results of these simple-effect F-tests, and the results of a Newman-Keuls comparison test are presented in Table 7.

The simple-effect F-tests with Methods held constant reveal that Density is significant for all three Methods ($p < 0.001$). Subsequent comparison tests show that, for both Icons and Enclosures, there are no significant differences among the three Densities (3-, 5-, and 7-word). However, each of these densities produced significantly longer reading times than did the control condition ($p < 0.05$).

For Character Styles the 7-word condition produced significantly longer reading times ($p < 0.05$) than did the 3-word condition. The 7- and 5-word conditions are not significantly different, and the 5- and 3-word conditions are not significantly different. The 7- and 5-word

Table 4. ANOVA Summary Table for Reading Times Incorporating Density (D)

Source of Variance	df	ε*	MS	F	*p*
Between-Subjects					
Subjects (S)	23		581883		
Within-Subject					
Method (M)	2		2065	0.05	> 0.05
Cue (C/M)	3		23931	2.40	> 0.05
Density (D)	3	0.7320	338884	19.69	< 0.0001**
M X D	6	0.4110	33188	3.11	< 0.05**
C/M X D	9	0.3021	41184	5.24	< 0.01**
M X S	46		42509		
C/M X S	69		10214		
D X S	69		17214		
M X D X S	138		10681		
C/M X D X S	207		7861		
Total	575				

* Greenhouse and Geisser (1959) epsilon correction for degrees of freedom
** denotes an adjusted *p*-value

Table 5. Newman-Keuls Comparison of Reading Times across Densities

Density Condition	Mean Time (seconds)	
7-word	6.73	|
5-word	6.64	|
3-word	6.51	|
Control	5.73	

* A vertical bar indicates that there are no significant differences among the means to the left of the bar.

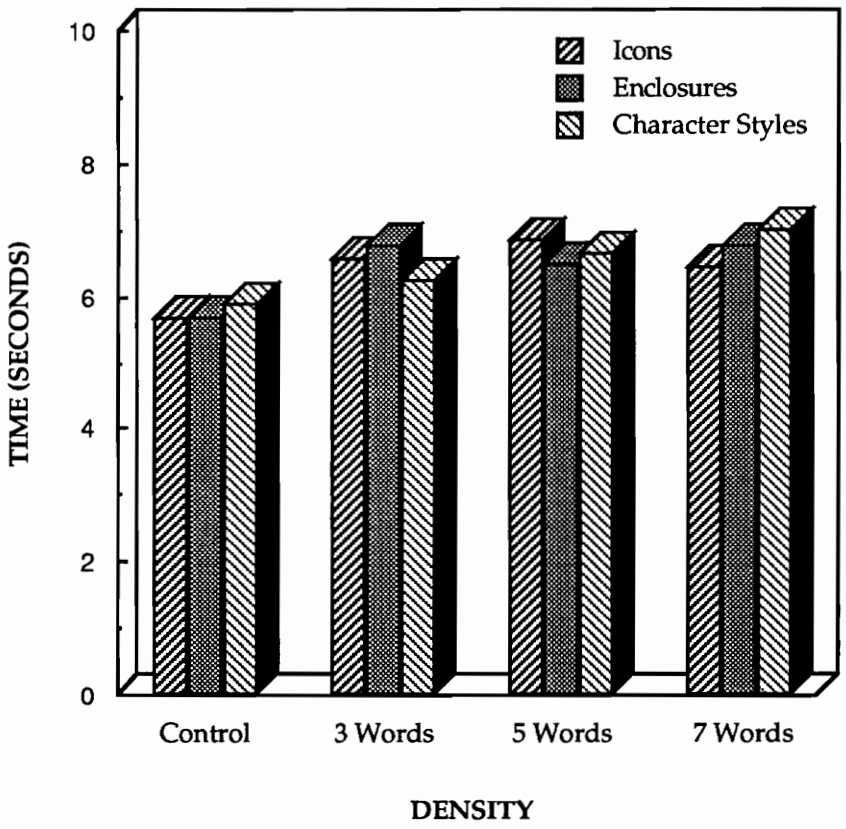


Figure 6. Methods X Density interaction.

Table 6. Simple-Effect F-Tests of Reading Times for Each Method

Source of Variance	df	ε*	MS	F	p
M X D X S	138		10681		
Method = Icons	3	0.4110	142390	13.33	< 0.001**
Method = Enclosures	3	0.4110	134726	12.61	< 0.001**
Method = Character Styles	3	0.4110	128145	12.00	< 0.001**

* Greenhouse and Geisser (1959) epsilon correction for degrees of freedom
** denotes an adjusted *p*-value

Table 7. Newman-Keuls Comparisons of Reading Times for Each Method

	Density Condition	Mean Time (seconds)
Method = Icons	5-word	6.82 |
	3-word	6.56 |
	7-word	6.44 |
	Control	5.65
Method = Enclosures	7-word	6.76 |
	3-word	6.73 |
	5-word	6.48 |
	Control	5.67
Method = Character Styles	7-word	7.00 |
	5-word	6.62 |
	3-word	6.24 |
	Control	5.85 |

* A vertical bar indicates that there are no significant differences among the means to the left of the bar.

Table 8. Simple-Effect F-Tests of Reading Times for Each Density across Methods

Source of Variance	df	ε*	MS	F	p
M X D X S	138		10681		
Density = Control	2		6446	0.60	> 0.25
Density = 3 words	2	0.4110	32749	3.07	> 0.05**
Density = 5 words	2		19919	1.86	> 0.05
Density = 7 words	2	0.4110	42330	3.87	> 0.05**

* Greenhouse and Geisser (1959) epsilon correction for degrees of freedom
** denotes an adjusted *p*-value

conditions produced significantly longer reading times than did the control condition ($p < 0.05$), whereas the 3-word condition produced reading times that are not significantly different from the control condition. Simple-effect F-tests with Density held constant are not significant with interpolated degrees of freedom (Table 8).

The Cues X Density interaction is significant ($p < 0.01$). The results of simple-effect F-tests are presented in Tables 9 and 11, and the results of Newman-Keuls comparison tests are presented in Tables 10 and 12.

The simple-effect F-tests with Cues held constant reveal significant Density effects for all Cues. Within the Link Icon Cue ($p < 0.001$), the 5-word condition produced significantly longer reading times than did either the 3- or 7-word conditions, which produced nondifferent reading times. The control condition produced significantly shorter reading times than did either the 5-, 3-, or 7-word conditions.

Within the Flag Icon Cue ($p < 0.001$), all three Densities produced significantly longer reading times than did the control condition ($p < 0.05$). There are no significant differences among the reading times for the Density conditions.

Within the Box Enclosure Cue ($p < 0.001$), the 7-word condition produced significantly longer reading times than did any of the other conditions, including the control ($p < 0.05$). Both the 3- and 5-word conditions produced significantly longer reading times than did the control condition ($p < 0.05$). The 3-and 5-word conditions produced nondifferent reading times.

As with the Flag Icon Cue, all three Densities within the Brackets Cue produced significantly longer reading times than did the control condition ($p < 0.05$). There are no significant differences among the reading times for the experimental Density conditions.

Within the Outlined text Cue ($p < 0.01$), the 7-word condition produced significantly longer reading times than did either the 3-word or the control condition ($p < 0.05$). The 7- and

Table 9. Simple-Effect F-Tests of Reading Times for Each Cue

Source of Variance	df	ε*	MS	F	p
C/M X D X S	207		7861		
Cue = Link Icons	3	0.3021	143828	18.30	< 0.001**
Cue = Flag Icons	3	0.3021	53710	6.83	< 0.001**
Cue = Box Enclosures	3	0.3021	97319	12.38	< 0.001**
Cue = Bracket Enclosures	3	0.3021	81785	10.40	< 0.001**
Cue = Outlined Text	3	0.3021	32943	4.19	< 0.01**
Cue = Shadowed Text	3	0.3021	119228	15.17	< 0.001**

* Greenhouse and Geisser (1959) epsilon correction for degrees of freedom
** denotes an adjusted *p*-value

Table 10. Newman-Keuls Comparisons of Reading Times for Each Cue

	Density Condition	Mean Time (seconds)	
Cue = Link Icons	5-word	7.32	
	3-word	6.62	|
	7-word	6.22	|
	0-word	5.68	
Cue = Flag Icons	7-word	6.66	|
	3-word	6.51	|
	5-word	6.32	|
	0-word	5.63	
Cue = Box Enclosures	7-word	7.13	
	3-word	6.56	|
	5-word	6.21	|
	0-word	5.68	

* A vertical bar indicates that there are no significant differences among the means to the left of the bar.

Table 10 (continued). Newman-Keuls Comparisons of Reading Times for Each Cue

	Density Condition	Mean Time (seconds)	
Cue = Bracket Enclosures	3-word	6.90	|
	5-word	6.75	|
	7-word	6.39	|
	0-word	5.66	
Cue = Outlined Text	7-word	6.73	|
	5-word	6.38	| |
	0-word	6.02	|
	3-word	5.96	|
Cue = Shadowed Text	7-word	7.27	|
	5-word	6.86	| |
	3-word	6.51	|
	0-word	5.69	

* A vertical bar indicates that there are no significant differences among the means to the left of the bar.

Table 11. Simple-Effect F-Tests of Reading Times for Each Density across Cues

Source of Variance	df	ε*	MS	F	p
C/M X D X S	207		7861		
Density = Control	3		4738	0.60	> 0.25
Density = 3 words	3	0.3021	18303	2.33	> 0.05**
Density = 5 words	3	0.3021	87247	11.10	< 0.001**
Density = 7 words	3	0.3021	44811	5.70	< 0.01**

* Greenhouse and Geisser (1959) epsilon correction for degrees of freedom
** denotes an adjusted *p*-value

Table 12. Newman-Keuls Comparisons of Reading Times for Density

	Cue	Mean Time (seconds)	
Density = 5 words	Link	7.32	
	Shadowed text	6.86	|
	Brackets	6.75	| |
	Outlined text	6.38	| |
	Flag	6.32	|
	Box	6.21	|
Density = 7 words	Shadowed text	7.27	|
	Box	7.13	|
	Outlined text	6.73	| |
	Flag	6.66	| |
	Brackets	6.39	|
	Link	6.22	|

* A vertical bar indicates that there are no significant differences among the means to the left of the bar.

5-word groups yielded nondifferent reading times, as did the 5-word, 3-word, and control conditions.

The 7-word condition yielded significantly longer reading times than did either the 3-word or the control condition ($p < 0.05$) within the Shadowed text Cue ($p < 0.001$). The 5-word condition produced significantly longer reading times than did the control condition ($p < 0.05$). The 7- and 5-word conditions produced nondifferent reading times, as did the 3-word and control conditions.

The simple-effect F-tests of Cues with Density held constant yield no significant results for the control ($p > 0.25$) and 3-word ($p > 0.05$) conditions. (Degrees of freedom were interpolated for the 3-word condition.) The 5-word condition is significant ($p < 0.001$), as is the 7-word condition ($p < 0.01$). A Newman-Keuls analysis of the 5-word condition indicates that Box Enclosures produced significantly shorter reading times than did either the Link Icons or the Shadowed text Cues ($p < 0.05$). All other differences among Cue conditions are non-significant. Within the 7-word condition, reading times produced by the Shadowed text and Box Enclosures Cues are significantly longer than those produced by the Bracket Cues and the Link Icons ($p < 0.05$). The Shadowed text and Box Cues produced nondifferent reading times, as did the Bracket Cues and Link Icons. Reading times for Outlined text and Flag Cues are not different from all other Cue conditions.

Number of Errors

Less than 1% of all responses were errors. Across all participants, three errors were made within Flag and Link Icons and Outlined Text. Two errors were made within Shadowed Text and Brackets, one within the first control condition, and no errors were made within either the Box Enclosures or second control condition. These low error frequencies are expected due to

the simple nature of the reading tasks; Tinker passages were designed to produce consistent reading times, and not high numbers of erroneously selected inappropriate words.

Table 13 presents an ANOVA summary table for the number of errors made within each Cue as well as within each Cueing Method. The type of error made for the reading tasks was a failure to select the inappropriate word for a Tinker passage. There is no significance for Cueing Method or for the nested Cues.

Subjective Data

Absolute ratings. An ANOVA was performed on absolute ratings for each dimension on the Absolute Rating Questionnaire; the results incorporate Greenhouse and Geisser (1959) corrections for degrees of freedom and are summarized in Tables 14, 15, 16, 18, 20. Newman-Keuls comparison tests for significant Method effects are presented in Tables 17, 19, and 21.

Items #1 and #2. The ANOVA for item #1 on the Absolute Rating Questionnaire indicates that there are no significant differences among the Methods in the ability to distinguish punctuation marks from Cues. Likewise, there are no significant differences among the Methods in having to read passages more than once (item #2).

Item #3. The ANOVA for item #3 reveals a significant result ($p = 0.0411$), and the Newman-Keuls comparison indicates that Character Styles are significantly more noticeable than either Icons or Enclosures ($p < 0.05$). There is no significant difference between Icons and Enclosures for item #3.

Item #4. The ANOVA for item #4 also reveals a significant Method effect ($p = 0.0190$), and the Newman-Keuls comparison indicates that Character Styles are significantly more distracting than either Icons or Enclosures ($p < 0.05$). Of the Icon and Enclosures Methods, one is no more distracting than the other.

Table 13. ANOVA Summary Table for Number of Errors

Source of Variance	df	MS	F	*p*
Between-Subjects				
Subjects (S)	23	0000.140		
Within-Subject				
Method (M)	2	0000.118	0.97	0.4119
Cue (C/M)	3	0000.031	0.26	0.9001
M X S	46	0000.122		
C/M X S	69	0000.118		
Total	143			

Table 14. ANOVA Summary Table for Absolute Questionnaire Item Number 1: Could you distinguish punctuation marks from cues?

Source of Variance	df	ε*	MS	F	*p*
Between-Subjects					
Subjects (S)	23		1.882		
Within-Subject					
Method (M)	2	0.7827	0.681	0.922	0.3856**
M X S	46		0.739		
Total	71				

* Greenhouse and Geisser (1959) epsilon correction for degrees of freedom
** denotes an adjusted *p*-value

Table 15. ANOVA Summary Table for Absolute Questionnaire Item Number 2: Did you have to read passages more than once?

Source of Variance	df	ε*	MS	F	*p*
Between-Subjects					
Subjects (S)	23		3.420		
Within-Subject					
Method (M)	2	0.8904	2.167	1.954	0.1588**
M X S	46		1.109		
Total	71				

* Greenhouse and Geisser (1959) epsilon correction for degrees of freedom
** denotes an adjusted *p*-value

Table 16. ANOVA Summary Table for Absolute Questionnaire Item Number 3: Could you notice the "linked" words in the passages?

Source of Variance	df	ε*	MS	F	*p*
Between-Subjects					
Subjects (S)	23		1.659		
Within-Subject					
Method (M)	2	0.8491	6.292	3.675	0.0411**
M X S	46		1.712		
Total	71				

* Greenhouse and Geisser (1959) epsilon correction for degrees of freedom
** denotes an adjusted *p*-value

Table 17. Newman-Keuls Comparison of Subjective Ratings of Methods for Absolute Questionnaire Item Number 3: Could you notice the "linked" words in the passages?

Method	Mean Rating	
Icons	2.13	|
Enclosures	1.96	|
Character Styles	1.17	

* A vertical bar indicates that there are no significant differences between the means to the left of the bar.

Table 18. ANOVA Summary Table for Absolute Questionnaire Item Number 4: Were the cues distracting?

Source of Variance	df	ε*	MS	F	*p*
Between-Subjects					
Subjects (S)	23		6.123		
Within-Subject					
Method (M)	2	0.9957	14.014	4.337	0.0190**
M X S	46		3.231		
Total	71				

* Greenhouse and Geisser (1959) epsilon correction for degrees of freedom
** denotes an adjusted *p*-value

Table 19. Newman-Keuls Comparison of Subjective Ratings of Methods for Absolute Questionnaire Item Number 4: Were the cues distracting?

Method	Mean Rating	
Character Styles	4.88	
Enclosures	3.75	|
Icons	3.42	|

* A vertical bar indicates that there are no significant differences between the means to the left of the bar.

Table 20. ANOVA Summary Table for Absolute Questionnaire Item Number 5: Did you experience any eye strain while reading the passages as a result of the cue present?

Source of Variance	df	ε*	MS	F	*p*
Between-Subjects					
Subjects (S)	23		4.879		
Within-Subject					
Method (M)	2	0.8910	15.125	10.801	0.0003**
M X S	46		1.400		
Total	71				

* Greenhouse and Geisser (1959) epsilon correction for degrees of freedom
** denotes an adjusted *p*-value

Table 21. Newman-Keuls Comparison of Subjective Ratings of Methods for Absolute Questionnaire Item Number 5: Did you experience any eye strain while reading the passages as a result of the cue present?

Method	Mean Rating	
Character Styles	3.63	
Icons	2.25	|
Enclosures	2.25	|

* A vertical bar indicates that there are no significant differences between the means to the left of the bar.

Item #5. Finally, for item #5 a significant result ($p = 0.0003$) was obtained from the ANOVA, and the Newman-Keuls comparison reveals that participants reported significantly more eye strain while reading passages containing Character Style Cues than while reading those containing either Icons or Enclosures ($p < 0.05$). Again, there is no significant difference between Icons and Enclosures.

Relative ratings. An ANOVA was conducted for each dimension on the Absolute Rating Questionnaire. ANOVAs are summarized in Tables 22, 24, 26, 28, and 30. Corresponding Newman-Keuls comparison tests across Methods and Cues are presented in Tables 23, 25, 27, and 29.

Item #1. The ANOVA for item #1 on the Relative Rating Questionnaire reveals a significant result ($p = 0.0036$). A Newman-Keuls comparison indicates that passages containing Character Styles are significantly more difficult to read than those containing Icons or Enclosures ($p < 0.05$).

Item #2. Item #2 is also determined from the ANOVA to have a significant result ($p = 0.0001$). The Newman-Keuls comparison indicates that Character Styles are significantly more distracting than Icons or Enclosures ($p < 0.05$), supporting the result of this same dimension from the Absolute Rating Questionnaire. There is no significant difference between the Icons and the Enclosures along this dimension.

Item #3. An ANOVA shows item #3 to have a significant result ($p = 0.0115$). A Newman-Keuls comparison further reveals that both Icons and Enclosures are more comfortable to use than Character Styles ($p < 0.05$).

Item #4. An ANOVA indicates a significant result for item #4 ($p = 0.0136$). The Newman-Keuls comparison reveals that participants experienced significantly more problems reading punctuation in passages containing Enclosure Cues than those containing either Icons or Character Style Cues ($p < 0.05$).

Table 22. ANOVA Summary Table for Relative Questionnaire Item Number 1: Easy to read?

Source of Variance	df	ε*	MS	F	p
Between-Subjects					
Subjects (S)	23		2.232		
Within-Subject					
Method (M)	2	0.9870	23.625	6.415	0.0036**
M X S	46		3.683		
Total	71				

* Greenhouse and Geisser (1959) epsilon correction for degrees of freedom
** denotes an adjusted p-value

Table 23. Newman-Keuls Comparison of Subjective Ratings of Methods for Relative Questionnaire Item Number 1: Easy to read?

Method	Mean Rating	
Character Styles	4.46	
Enclosures	2.96	|
Icons	2.58	|

* A vertical bar indicates that there are no significant differences between the means to the left of the bar.

Table 24. ANOVA Summary Table for Relative Questionnaire Item Number 2: Distracting?

Source of Variance	df	ε*	MS	F	p
Between-Subjects					
Subjects (S)	23		2.725		
Within-Subject					
Method (M)	2	0.8710	44.042	12.722	< 0.0001**
M X S	46		3.462		
Total	71				

* Greenhouse and Geisser (1959) epsilon correction for degrees of freedom
** denotes an adjusted p-value

Table 25. Newman-Keuls Comparison of Subjective Ratings of Methods for Relative Questionnaire Item Number 2: Distracting?

Method	Mean Rating	
Character Styles	5.71	
Enclosures	3.63	|
Icons	3.17	|

* A vertical bar indicates that there are no significant differences between the means to the left of the bar.

Table 26. ANOVA Summary Table for Relative Questionnaire Item Number 3: Comfortable to use?

Source of Variance	df	ε*	MS	F	*p*
Between-Subjects					
Subjects (S)	23		2.848		
Within-Subject					
Method (M)	2	0.9865	20.792	4.971	0.0115**
M X S	46		4.183		
Total	71				

* Greenhouse and Geisser (1959) epsilon correction for degrees of freedom
** denotes an adjusted *p*-value

Table 27. Newman-Keuls Comparison of Subjective Ratings of Methods for Relative Questionnaire Item Number 3: Comfortable to use?

Method	Mean Rating	
Character Styles	4.96	
Enclosures	3.63	|
Icons	3.17	|

* A vertical bar indicates that there are no significant differences between the means to the left of the bar.

Table 28. ANOVA Summary Table for Relative Questionnaire Item Number 4: Problems reading punctuation?

Source of Variance	df	ε*	MS	F	*p*
Between-Subjects					
Subjects (S)	23		3.507		
Within-Subject					
Method (M)	2	0.9712	11.347	4.805	0.0136**
M X S	46		2.362		
Total	71				

* Greenhouse and Geisser (1959) epsilon correction for degrees of freedom
** denotes an adjusted *p*-value

Table 29. Newman-Keuls Comparison of Subjective Ratings of Methods for Relative Questionnaire Item Number 4: Problems reading punctuation?

Method	Mean Rating	
Enclosures	3.50	
Character Styles	2.54	|
Icons	2.17	|

* A vertical bar indicates that there are no significant differences between the means to the left of the bar.

Table 30. ANOVA Summary Table for Relative Questionnaire Item Number 5: Problems finding the target "wrong" word?

Source of Variance	df	ε*	MS	F	*p*
Between-Subjects					
Subjects (S)	23		2.603		
Within-Subject					
Method (M)	2	0.9772	4.667	1.695	0.1957**
M X S	46		2.754		
Total	71				

* Greenhouse and Geisser (1959) epsilon correction for degrees of freedom
** denotes an adjusted *p*-value

Item #5. There are no significant results for item #5, regarding problems finding the target "wrong" word.

Preferred Cue within each Method. For each Absolute Rating Questionnaire, participants were asked to select a preferred Cue within each Method for the tasks they had performed in Experiment 1. Although most participants reported a preferred Cue within a Method, some did not indicate a preference. Therefore, those participants who failed to make a distinction were removed from the data pool, and the preferences of remaining participants were analyzed. The Binomial Test (Siegel and Castellan, 1988) was used to determine if one particular Cue was preferred by significantly more participants (excluding those who indicated "no preference") than the other Cue. The results of these tests are presented in Tables 31, 32, and 33.

The results of the Binomial Tests indicate no significant differences between the number of participants preferring the Link (8) and Flag (12) Icons (Table 31). Similarly, nondifferent numbers of participants preferred the Box (8) and Bracket (13) Enclosures (Table 32). For the Character Styles Method (Table 33), the number of participants who preferred Outlined text (15) significantly exceeds the number preferring Shadowed text (4) ($p = 0.0095$).

Overall Method and Cue preferred. For each Relative Rating Questionnaire, participants were asked to select an overall preferred Method and Cue for reading passages containing linked words. The Cochran Q Test was used to analyze the preferences data. Among Methods there are three possible choices, and among Cues there are six possible choices. The results (presented in Table 34) indicate that there are no significant differences in Method or Cue preferences.

All participants responded to this item. However, the number of participants who preferred the Cues within a Method does not necessarily equal the number of persons who preferred that Method. There are two reasons for this: one is that a single participant chose

Table 31. Binomial Test of Cue Preferences within the Icons Method for the Absolute Rating Questionnaire

Cues	n	*p*
Link / Flag	8 / 12	0.2517
No Preference	4	
Total Subjects	24	

Table 32. Binomial Test of Cue Preferences within the Enclosures Method for the Absolute Rating Questionnaire

Cues	n	*p*
Box / Brackets	8 / 13	0.1916
No Preference	3	
Total Subjects	24	

Table 33. Binomial Test of Cue Preferences within the Character Styles Method for the Absolute Rating Questionnaire

Cues	n	*p*
Outlined / Shadowed	15 / 4	0.0095
No Preference	5	
Total Subjects	24	

Table 34. Cochran Q Tests of Method and Cue Preferences for the Relative Rating Questionnaire

Factor	df	Q	*p*
Method	2	2.25	> 0.05
Cue	5	1.67	> 0.05

both Cues within the preferred Method as the preferred Cues, and the other reason is that several participants erroneously selected a preferred Cue that was not nested in their preferred Method. For example, a participant may have preferred the overall Method of Icons, but in particular liked Boxes as Cues. The implications of these anomalies are addressed in the Discussion section.

EXPERIMENT 1: DISCUSSION

Reading Times

No single Cue or Cueing Method produced longer reading times than any other Cue or Cueing Method. However, the presence of linked words did lead to a significant difference between control and experimental conditions within both Methods (with the average reading times of Icons, Enclosures, and Character Styles being 6.6 seconds, each, and for the control condition being 5.8 seconds) and Cues.

Number of Errors

The number of erroneous responses was very low (less than 1%) across all experimental and control conditions. This result is expected, since Tinker passages rarely yield errors. There are no significant differences in the numbers of errors made within the Method or Cue conditions, including experimental versus control conditions. Therefore, the presence of linked words in passages, regardless of the Cue used, does not degrade the readability of the text to the extent that errors increase.

Subjective Ratings

Ratings of the Methods and Cues are more variable than task performance times and errors. The absolute ratings indicate more favorable responses to Icons and Enclosures than to Character Styles along dimensions of distraction and eye strain. The relative ratings conform to the absolute ratings, with passages incorporating Icons and Enclosures being rated as significantly easier to read than those incorporating Character Styles. According to the

relative ratings, Icons and Enclosures are significantly less distracting and more comfortable to use than Character Styles.

There are no significant differences among the three Methods in distinguishing punctuation marks from Cues or having to read passages more than once. The relative user ratings do indicate, however, that Enclosures are significantly more problematic for reading punctuation than are the Icons or Character Styles. This item was included in the questionnaire because interference with punctuation was considered to be as detrimental to readability as interference with text.

Since Enclosures do interfere with punctuation relative to the other two Methods, this Method is considered to have a more obtrusive presence than the Icons or Character Styles. The fact that Enclosures are not rated as significantly more problematic along this dimension of punctuation on an absolute basis may suggest that the interference of this Method is significant only in comparison with the other two Methods.

While Icons and Enclosures are rated as more comfortable than Character Styles, the readability of text incorporating Character Style Cues is not degraded to the extent that reading time is affected. In other words the readability of the text containing Character Style Cues is degraded *only* with regard to user preference and comfort, and not with regard to reading time.

Lack of Cue Compatibility within Methods for Subjective Ratings

It is difficult to make assessments about Character Styles as a group based on subjective ratings since the two Cues within that group yield significantly different preference frequencies. For example, the Character Styles Method might have received higher ratings or been preferred by more participants if a Cue more equivalent to Outlined text, and not Shadowed text, had been used. In other words, since a low number of participants preferred

Shadowed text, this Cue may have degraded the appeal for the entire Method of Character Styles.

On the other hand, Character Styles might receive the same ratings relative to the other Methods *regardless* of the inclusion of Shadowed text or an alternative Cue, and still produce more preferences for the Outlined text than for the other Cue used. It is impossible to assess the extent to which the lack of appeal for Shadowed text affects the overall ratings and preferences for Character Styles as a group.

User Preferences

The preferences for Methods and Cues overall (from the Relative Rating Questionnaire) did not yield significant differences among either of the two factors. It can therefore be concluded that one Cue or Cueing Method was not preferred by significantly more participants than any of the others for readability.

Density of Linked Words

All three of the experimental conditions produced significantly longer reading times than did the control conditions across all Cues and Cueing Methods, with the smallest difference being between the 3-word and control conditions at 0.8 second. The mere presence of linked words in general degraded the readability of the text, as indicated in the reading times previously discussed. Since the only significant differences exist between the control and the three Density conditions, no recommendations can be made with regard to a maximum proportion of linked to non-linked words within a hypertext document.

The Methods X Density interaction reveals that, unlike within the Icon and Enclosure Methods, as the Density increases from the 3-word to the 7-word condition within Character Styles, reading times increase significantly. Therefore, although Character Styles appear to

be the most appropriate Cueing Method at low densities, readability may be degraded in areas of a hypertext document where linked words are more densely located.

Results of analyses within the Cues X Density interaction are difficult to explain within certain Cues. Although the simple-effect F-tests are significant for all Cues, the variability among density conditions does not suggest a meaningful trend. For example, within the Link Icons Cue the 5-word condition produced significantly longer reading times than did the 3-word, 7-word, and control condition. In addition, the reading times of the 7-word condition are equivalent to those of the control condition.

One would not expect the highest Density condition to produce reading times comparable to those of the control group and significantly shorter than the 5-word condition. Similarly, one would not expect the lowest Density condition to produce significantly longer reading times than the control in light of the comparability of the 7-word and control conditions. Link Icons appear to be appropriate for low or high Densities but not for moderate Densities. Results of the Newman-Keuls analyses within Box Cues conversely suggest that Box Cues may be appropriate for moderate Densities but not for high or low Densities.

Comparison tests suggest that since the reading times yielded by the 3- and 5-word conditions are comparable to those of the control condition within Outlined text, and since only the 7-word condition produces reading times that are significantly longer than those of the 3-word and control conditions, Outlined text is most appropriate with regard to all density levels. Conversely, since all experimental levels of the Shadowed condition produced significantly longer reading times than did the control condition, and since text readability continued to degrade as the Density levels increased, Shadowed text is the least appropriate Cue for all levels of density.

These conclusions ironically place the two Cues within one Method (Character Styles) on opposing extremes with regard to Density. For this reason, and in light of the Methods X

Density interaction results (indicating that Character Styles are inappropriate for high density levels and that Icons are appropriate for both high and low density levels), Character Styles are not recommended for eliminating Density effects within hypertext documents. The results of the Density analyses, in fact, are too complex to recommend a particular Method over the others as most suitable for balancing Density effects. Therefore, the effect of Density is not considered in determining a most appropriate Cueing Method within hypertext documents. The Density issue is most appropriately addressed once a standard cueing notation has been adopted.

EXPERIMENT 2: METHOD

Participants. The same 24 participants used in Experiment 1 were also used for Experiment 2. Experiment 2 commenced as soon as participants had completed filling out the Relative Rating Questionnaire for Experiment 1.

Materials and apparatus. The same apparatus used in Experiment 1 was used to present the stimulus screens containing text fields (described in the next section). These screens also were black-and-white cards in a Hypercard stack. Again, the computer's clock was used to time participants in this task, and data files were made in the same manner as in Experiment 1.

Visual search task. A visual search task was used to evaluate the salience of each Cue and Cueing Method. Search times for text fields containing embedded linking notation (the time taken to locate all linked words, as indicated by specific Cues, in a text field) were recorded. Each text field was eight lines long and covered roughly the same amount of screen space. An example of a text field is presented in Figure 5. Passages were presented one at a time. Each participant was instructed to scan each text field, locate and count all linked words, click on a button to go to a blank screen, and then verbalize the number counted.

Text fields. Sixty text fields were used to evaluate the salience of the Cues and Cueing Methods. Each text field contained one to five "linked" words (words with associated Cues). As in Experiment 1, these words were not actually linked to other text; they merely possessed Cues as if they were. There was no control group, since a participant would likely have taken *longer* trying to locate a cued word that simply was not there.

Derivation of text fields. The text fields used in Experiment 2 were obtained from three *New York Times* Sunday newspapers, dated March 11, March 25, and April 15, 1990. Text fields containing minimal quotes, and few numerical or monetary values were selected to reduce the amount of clutter in the text. Furthermore, the content of each text field was different from

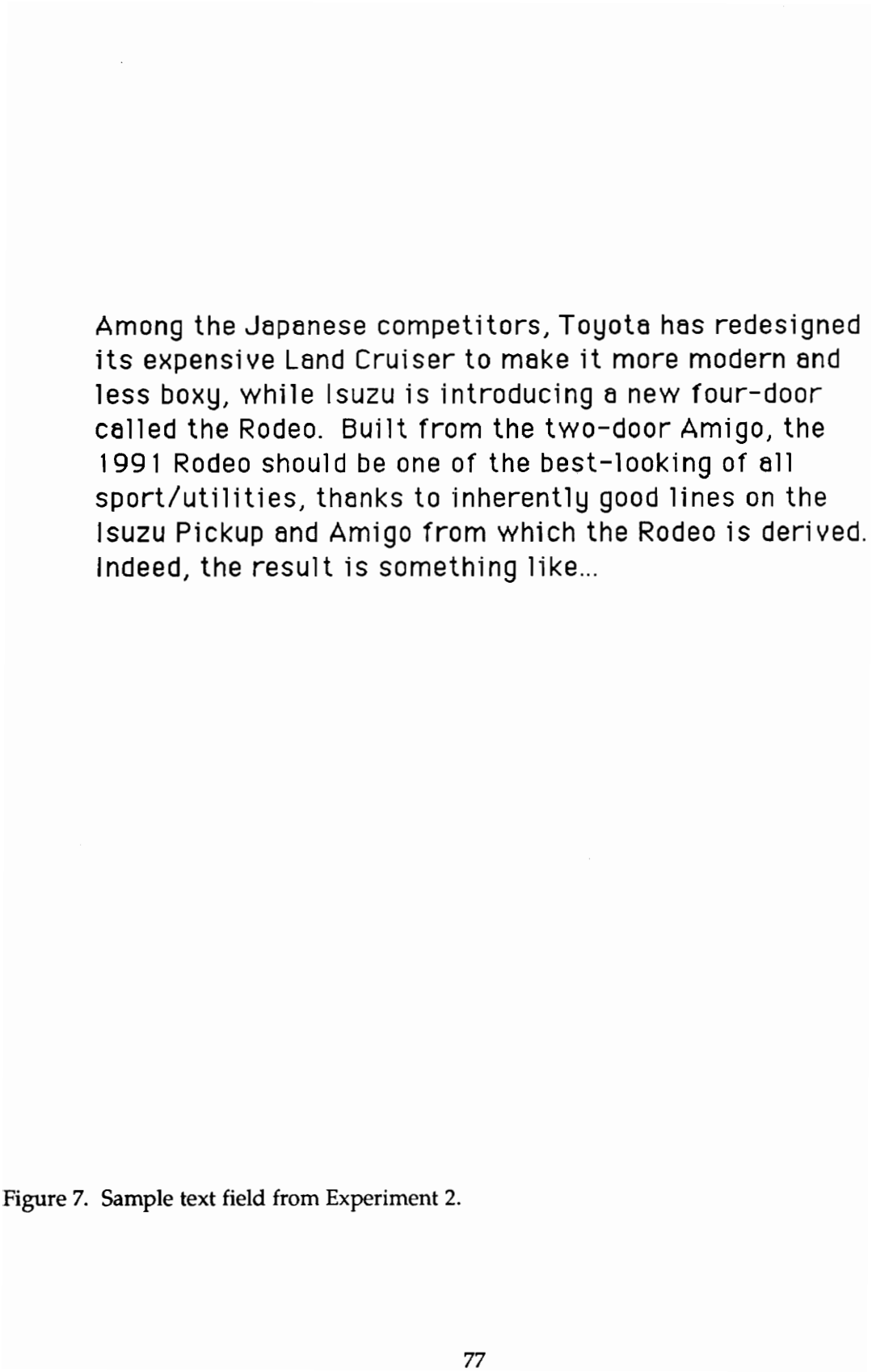


Figure 7. Sample text field from Experiment 2.

that of the other text fields and of a fairly innocuous nature. Since these text fields were not standardized, they were randomly assigned to each Cue (and Cueing Method) for each participant; each participant viewed a different group of text fields for each condition.

Text fields were methodically, as opposed to randomly, assigned to "number of linked words" conditions, which were the same for all participants. The following procedure was used to assign text fields to conditions: once a text field was taken from a newspaper article it was placed in an envelope with the other text field clippings. Once 60 text fields had been obtained they were taped, in no particular order, into notebook pages in groups of six, resulting in 10 pages of text fields.

Once this had been done, each text field was numbered 1 through 60 in the order it had been placed on the page. Then groups of text fields were assigned to conditions based on their numerical orders. For example, text fields one through four were assigned to the "one-word" condition, five through eight were assigned to the "two-word" condition, et cetera. Once a group of four text fields had been assigned to the "five-word" condition, the process started again with the next four text fields being assigned to the "one-word" condition. This method was repeated until 12 text fields had been assigned to each of the five "number of linked words" conditions.

Random assignment of words in text fields to possess linking notation. Prior to data collection, a random number generator was used to assign words that would contain Cues. For each text field the exact number of words in the text field was determined. A list of numbers between this number (of words in the text field) and one was then generated in random order for each text field. The first one, two, three, four, or five numbers in the list (depending on the "number of words" condition for that text field) were used to assign Cues to the corresponding words in the text field. This process, which is much like the one used in Experiment 1, was followed for each passage.

Experimental Procedure

Since participants had already passed the vision test, signed the required consent forms, and read the introductory material prior to performing tasks in Experiment 1, they read supplementary instructions for Experiment 2 and then began the tasks on the computer.

Task description. Each participant viewed a series of 10 text fields for each Cue condition. The text fields appeared one at a time on HyperCard® cards; each group of 10 cards made one stack for the corresponding Cue condition.

As in Experiment 1, a button labeled "Next" was located below each text field. The participants clicked on this button to go from a stimulus card (containing a text field) to a non-stimulus card with the following message: "How many linked words are there in the passage?" The participant was instructed to locate and count all (one to five) "linked" words within each text field, click on the button labeled "Next," and verbalize the number of linked words that were counted at the message prompt on the following non-stimulus card. As in Experiment 1, participants viewed a series of 10 text fields for each "Cue condition" stack by clicking the "Next" buttons on the stimulus and non-stimulus cards.

The time taken to locate all linked words for each text field was recorded (using the Macintosh's clock) for analysis. The same timing mechanism designed for Experiment 1 was used for Experiment 2. Target search time for "linked words" (words containing embedded linking notation) was considered to indicate the salience of each Cueing Method. The number of missed target words (errors) was also recorded across Cues and Methods.

Practice trials. Practice trials served the same basic purpose as in Experiment 1. Participants were given a practice trial (consisting of three practice screens) prior to each block of experimental trials. This was done to expose the participant to the text fields and to provide feedback regarding that participant's understanding of the visual search task. Each practice trial also familiarized the participant with the Cue that was to be used for the up-

coming experimental trial. The same three passages were used in all practice trials, incorporating the Cue used in the experimental trial that followed.

User preferences questionnaires. Each participant was asked to complete user preferences questionnaires. As with Experiment 1 an absolute-rating questionnaire was filled out immediately following completion of the trials for the second Cue within each Cueing Method, and a relative-rating questionnaire was filled out upon completion of each of the two experiments. Participants rated each Method, using a Likert scale, on various dimensions. Subjects were also asked to choose a preferred Cue within each Cueing Method, and an overall preferred Cue and Cueing Method, for those tasks performed in Experiment 2. The questionnaires are presented in Appendix C.

Debriefing. After participants had completed both experiments they were asked the following two questions:

(1) If you were browsing through an actual hypertext document to find out what types of information you could access, which would be most important to you - to be able to read the document without being distracted, or to be able to locate linked words quickly and easily?

(2) Given your answer to the first question, which Cueing *Method* would you feel most comfortable using?

These questions provided the experimenter with information regarding the relative importance and subsequent weightings of the salience and readability of Cueing Methods.

Experimental Design

This experiment incorporated the same within-subjects hierarchical design used in Experiment 1. Again, this design consisted of the factor of Cueing Method and the second nested factor of Cues within each Method. The experimental design is the same as that which was used for Experiment 1 (Figure 3), except the control condition has been eliminated.

Independent variables. Cueing Method (M) consists of three levels or conditions: Icons, Word Enclosures, and Character Styles. There are two Cues within each Cueing Method. The same Cues used in Experiment 1 were used in Experiment 2.

Dependent measures. Two dependent measures were employed: time to locate all linked words within text fields and the number of errors for each Cue and Method.

Order of treatments for Cueing Methods. There are six possible orders for presenting the three Methods. Four participants (two males and two females) were assigned to each order, yielding a sample size of 24 with all possible orders represented four times in a completely counterbalanced design. Participants were assigned to order conditions roughly in the order that they were scheduled for an experimental session (some deviation from this was required to ensure that an equal number of males and females was assigned to each of the six order conditions).

Order of treatments for Cues. As in Experiment 1, the particular Cue within each Method was changed after every 10 trials across all experimental conditions for each subject. As with Experiment 1, participants 1-12 received the Link Icon first and the Flag Icon second within the Icons Method, the Box first and the Bracket second within the Enclosures Method, and the Outlined text first and Shadowed text second within the Character Styles Method. Subjects 13-24 received Cues within each Method in the reverse order.

Data Analysis

Time to count target words. The amount of time it took for participants to locate and count linked words was averaged for each participant across the 10 trials within each Cue. No differentiation was made regarding the "number of cued words" groups since the five groups were represented equally within each of the six Cues.

Mean search times were analyzed using an ANOVA, and Newman-Keuls post-hoc comparison tests indicated any significant differences among Methods and Cues. Comparison test results with $p < 0.05$ were determined to be significant.

Number of errors. The numbers of targets missed within each Cueing Method and Cue were totaled across all trials for each participant. These totals were then analyzed using ANOVA procedures.

Subjective ratings. The rating data obtained from the user preferences questionnaires were analyzed using separate single-factor ANOVA procedures, and post-hoc comparison tests were used to probe all significant effects that were found. Again, comparison test results with $p < 0.05$ were determined to be significant. As with the first experiment, these rating data also were transposed for consistency.

User preferences. These questionnaires also provided opportunities for participants to select a preferred Cue within each Method (on the Absolute Rating Questionnaires) as well as an overall preferred Method and Cue (on the Relative Rating Questionnaires) with regard to tasks performed in Experiment 2 (finding linked words quickly). These data were analyzed using the Binomial Test and Cochran Q Test.

EXPERIMENT 2: RESULTS

Search Times

Assuming the worst case for violations of sphericity had no effect on the significance of the main effects or comparison tests. The ANOVA summary table for task completion times is shown in Table 35. Method ($p < 0.0001$) and Cue ($p < 0.0001$) are both significant. Table 36 presents a Newman-Keuls comparison across Methods. Search times across the three Cueing Methods are all significantly different from one another, with search times for Character Styles being the shortest and search times for Enclosures the longest.

Table 37 presents a Newman-Keuls comparison across Cues. Search times for Boxes, Link Icons, and Shadowed and Outlined text are significantly shorter than those for Flag Icons and Brackets ($p < 0.05$), with Brackets yielding the longest search times of all the Cues ($p < 0.05$).

Number of Errors

Less than 2% of all responses were errors. Brackets produced 12 errors, Outlined Text produced 6 errors, Box Enclosures yielded 3, Flag Icons and Shadowed Text each yielded 2, and no errors were made within Link Icons. Since locating and counting linked words is a simple task, these low error frequencies are not unexpected.

Table 38 presents an ANOVA summary table for the number of errors made within each Cue and Cueing Method. The type of error made for the visual search tasks was counting either too many or too few linked words in a text field. The number of *errors* as opposed to the number of *text fields* in which errors were made comprised the data. For instance, if a participant

Table 35. ANOVA Summary Table for Search Times

Source of Variance	df	MS	F	p
Between-Subjects				
Subjects (S)	23	4083.908		
Within-Subject				
Method (M)	2	167655.674	125.80	< 0.0001*
Cue (C/M)	3	168934.014	176.30	< 0.0001*
M X S	46	1332.688		
C/M X S	69	958.202		
Total	143			

* significant at $p < 0.05$ with epsilon at its lower bound

Table 36. Newman-Keuls Comparison of Search Times across Methods

Methods	Mean Time (seconds)
Enclosures	3.11
Icons	1.81
Character Styles	1.18

Table 37. Newman-Keuls Comparison of Search Times across Cues

Cues	Mean Time (seconds)
Brackets	4.79
Flag	2.15
Link	1.47 |
Box	1.44 |
Outlined text	1.26 |
Shadowed text	1.10 |

* A vertical bar indicates that there are no significant differences among the means to the left of the bar.

Table 38. ANOVA Summary Table for Number of Errors

Source of Variance	df	ε*	MS	F	*p*
Between-Subjects					
Subjects (S)	23		0000.908		
Within-Subject					
Method (M)	2	0.6775	0000.674	3.14	> 0.05**
Cue (C/M)	3	0.4896	0000.014	2.78	> 0.05**
M X S	46		0000.688		
C/M X S	69		0000.202		
Total	143				

* Greenhouse and Geisser (1959) epsilon correction for degrees of freedom
** denotes an adjusted *p*-value

missed two linked words in a single text field, this counted as two errors. The ANOVA indicates that neither main effect is significant at $p < 0.05$.

Subjective Data

Absolute ratings. ANOVA summary tables for absolute ratings, also incorporating Greenhouse and Geisser (1959) corrections for degrees of freedom, are contained in Tables 39 and 41. A Newman-Keuls comparison test for item #1 is presented in Table 40.

Item #1. An ANOVA indicates significance ($p < 0.0001$) for item #1. The Newman-Keuls comparison reveals that all three Methods are significantly different from each other. Character Styles were rated as significantly easier to locate than were Icons and Enclosures ($p < 0.05$). Likewise, Icons were rated as significantly easier to locate than were Enclosures ($p < 0.05$).

Item #2. The ANOVA does not reveal any significant result for item #2, indicating that the location of linked words in text fields did not affect the participants' ability to locate them.

Relative ratings. ANOVA summary tables for relative ratings are presented in Tables 42, 44, 46, and 48. Corresponding Newman-Keuls comparison tests for significant Method and Cue effects are presented in Tables 43, 45, 47, and 49.

Item #1. The ANOVA for item #1 reveals a significant result ($p < 0.0001$). According the the Newman-Keuls comparison, Character Styles are significantly easier to locate than Icons and Enclosures ($p < 0.05$), and Icons are significantly easier to locate than Enclosures ($p < 0.05$). These results are in agreement with the results for item #1 of the Absolute Rating Questionnaire for Experiment 2.

Table 39. ANOVA Summary Table for Absolute Questionnaire Item Number 1: Were the cues easy to locate?

Source of Variance	df	ε*	MS	F	*p*
Between-Subjects					
Subjects (S)	23		1.666		
Within-Subject					
Method (M)	2	0.9913	45.795	34.118	< 0.0001**
M X S	46		1.342		
Total	71				

* Greenhouse and Geisser (1959) epsilon correction for degrees of freedom
** denotes an adjusted *p*-value

Table 40. Newman-Keuls Comparison of Subjective Ratings of Methods for Absolute Questionnaire Item Number 1: Were the cues easy to locate?

Method	Mean Rating
Enclosures	3.98
Icons	2.83
Character Styles	1.23

Table 41. ANOVA Summary Table for Absolute Questionnaire Item Number 2: Did the location of the "linked" word in the passage affect your ability to locate the word?

Source of Variance	df	ε*	MS	F	*p*
Between-Subjects					
Subjects (S)	23		6.633		
Within-Subject					
Method (M)	2	0.8673	5.469	2.075	0.1445**
M X S	46		2.635		
Total	71				

* Greenhouse and Geisser (1959) epsilon correction for degrees of freedom
** denotes an adjusted *p*-value

Table 42. ANOVA Summary Table for Relative Questionnaire Item Number 1: Easy to find linked words?

Source of Variance	df	ε*	MS	F	*p*
Between-Subjects					
Subjects (S)	23		2.057		
Within-Subject					
Method (M)	2	0.7797	76.514	39.858	< 0.0001**
M X S	46		1.920		
Total	71				

* Greenhouse and Geisser (1959) epsilon correction for degrees of freedom
** denotes an adjusted *p*-value

Table 43. Newman-Keuls Comparison of Subjective Ratings of Methods for Relative Questionnaire Item Number 1: Easy to find linked words?

Method	Mean Rating
Enclosures	4.54
Icons	3.58
Character Styles	1.08

Table 44. ANOVA Summary Table for Relative Questionnaire Item Number 2: "Linked" words seemed to stand out from the rest of the text?

Source of Variance	df	ε*	MS	F	*p*
Between-Subjects					
Subjects (S)	23		1.463		
Within-Subject					
Method (M)	2	0.6942	91.764	42.294	< 0.0001**
M X S	46		2.170		
Total	71				

* Greenhouse and Geisser (1959) epsilon correction for degrees of freedom
** denotes an adjusted *p*-value

Table 45. Newman-Keuls Comparison of Subjective Ratings of Methods for Relative Questionnaire Item Number 2: "Linked" words seemed to stand out from the rest of the text?

Method	Mean Rating	
Enclosures	4.63	|
Icons	4.08	|
Character Styles	1.00	

* A vertical bar indicates that there are no significant differences between the means to the left of the bar.

Table 46. ANOVA Summary Table for Relative Questionnaire Item Number 3: If you had been reading the passages more thoroughly, would the cues have been distracting?

Source of Variance	df	ε*	MS	F	*p*
Between-Subjects					
Subjects (S)	23		2.751		
Within-Subject					
Method (M)	2	0.9841	33.514	9.233	0.0005**
M X S	46		3.630		
Total	71				

* Greenhouse and Geisser (1959) epsilon correction for degrees of freedom
** denotes an adjusted *p*-value

Table 47. Newman-Keuls Comparison of Subjective Ratings of Methods for Relative Questionnaire Item Number 3: If you had been reading the passages more thoroughly, would the cues have been distracting?

Method	Mean Rating	
Character Styles	5.04	
Enclosures	3.21	|
Icons	2.83	|

* A vertical bar indicates that there are no significant differences between the means to the left of the bar.

Table 48. ANOVA Summary Table for Relative Questionnaire Item Number 4: Did you experience any difficulty finding the "linked" words?

Source of Variance	df	ε*	MS	F	*p*
Between-Subjects					
Subjects (S)	23		0.896		
Within-Subject					
Method (M)	2	0.7629	110.722	55.029	< 0.0001**
M X S	46		2.012		
Total	71				

* Greenhouse and Geisser (1959) epsilon correction for degrees of freedom
** denotes an adjusted *p*-value

Table 49. Newman-Keuls Comparison of Subjective Ratings of Methods for Relative Questionnaire Item Number 4: Did you experience any difficulty finding the "linked" words?

Method	Mean Rating
Enclosures	5.58
Icons	4.00
Character Styles	1.33

Item #2. Item #2 also reveals a significant result in the ANOVA ($p < 0.0001$). The Newman-Keuls comparison indicates that Character Styles seem to stand out from the rest of the text more than Icons and Enclosures ($p < 0.05$). Icons and Enclosures are not significantly different for this item (item #2).

Item #3. The ANOVA for item #3 indicates a significant effect ($p < 0.0005$). The Newman-Keuls comparison shows that Character Styles are rated as being significantly more distracting than Icons or Enclosures ($p < 0.05$). There is no significant difference between Icons and Enclosures.

Item #4. Another significant result is obtained from the ANOVA ($p = 0.0001$) for item #4. The Newman-Keuls comparison shows that all three Methods are significantly different from each other. Participants felt that they experienced significantly less difficulty finding linked words with Character Styles than they did finding them with Icons or Enclosures ($p < 0.05$). Likewise, Icons were rated as being significantly less difficult to find than were Enclosures ($p < 0.05$).

Preferred Cue within each Method. For each Absolute Rating Questionnaire, participants were again asked to select a preferred Cue within each Method for the tasks they had performed in Experiment 2. Again, although most participants reported a preferred Cue within a Method, some did not indicate a preference. As with Experiment 1, those who failed to make a distinction were removed from the data pool, and the preferences of remaining participants were analyzed. The preference data were analyzed using the Binomial test. The results are presented in Tables 50 through 53.

The results of the Binomial Tests indicate that, within the Icon Method, Link Icons were preferred by significantly more participants (22) than were Flag Icons ($p < 0.0001$), which were preferred by only 1 participant. The remaining participant indicated no preference. All

Table 50. Cochran Q Tests of Cue Preferences within each Method for the Absolute Rating Questionnaire

Method	df	Q	*p*
Icons	1	19.17	<0.001
Enclosures	1	24.00	<0.001
Character Styles	1	20.00	<0.001

Table 51. Binomial Test of Cue Preferences within Icons for the Absolute Rating Questionnaire

Cues	n	*p*
Link / Flag	22 / 1	< 0.0001
No Preference	1	
Total Subjects	24	

Table 52. Binomial Comparison Tests of Cue preferences within the Enclosures Method for the Absolute Rating Questionnaire

Cues	n	*p*
Box / Brackets	24 / 0	< 0.0001
No Preference	0	
Total Subjects	24	

Table 53. Binomial Comparison Tests of Cue preferences within the Character Styles Method for the Absolute Rating Questionnaire

Cues	n	*p*
Outlined / Shadowed	0 / 20	0.0001
No Preference	4	
Total Subjects	24	

24 participants preferred the Box Cue to the Brackets within the Enclosures Method, yielding a significant difference between the two Cues ($p < 0.0001$). Twenty participants preferred the Shadowed text and none preferred the Outlined text within the Character Styles Method (significant at $p < 0.0001$); the four remaining participants indicated no preference.

Overall Method and Cue preferred. For each Relative Rating Questionnaire, participants were asked to select an overall preferred Method and Cue for locating linked words. The Cochran Q Test was used to analyze these data, and significant results were further analyzed using the Binomial Test. The results are presented in Tables 54 through 56. The Binomial test compared the *actual* number of preferences to the expected number for each Method (and Cue), assuming Methods (and Cues) are equally likely to be preferred. (The expected number to prefer any Method is eight, and to prefer any Cue is four.) Therefore the values for α have been adjusted to reflect the number of comparisons within Methods (3) and Cues (6), accounting for the increased probability of making a Type I error and controlling the Type I error rate experimentwise.

According to the Cochran Q Test there are significant differences among both Methods ($p < 0.001$) and Cues ($p < 0.001$). Icons were not preferred significantly more or less than expected. However, Enclosures were preferred by significantly fewer participants (2) than expected ($p = 0.0041$), and Character Styles were preferred by significantly more (19) than expected ($p < 0.0001$). Among Cues, only preferences for Shadowed text (preferred by 17 participants) significantly differ from the expected number ($p < 0.0001$), leaving Shadowed text as preferred by significantly more participants than any other Cue.

Table 54. Cochran Q Tests of preferences among Method and Cue for the Relative Rating Questionnaire

Factor	df	Q	*p*
Method	3	22.75	<0.001
Cue	6	47.46	<0.001

Table 55. Binomial Test of Method Preferences for the Relative Rating Questionnaire

Cue	n	p	α^*
Icons	3	0.0200	0.0085
Enclosures	2	0.0041	0.0085
Character Styles	19	< 0.0001	0.0085
Total Subjects	24		

* corrected for a constant experiment-wise Type I error rate

Table 56. Binomial Test of Cue Preferences for the Relative Rating Questionnaire

Cue	n	p	α^*
Link	3	0.4156	0.0043
Flag	0	0.0126	0.0043
Box	1	0.0730	0.0043
Brackets	2	0.2119	0.0043
Outlined	3	0.4156	0.0043
Shadowed	17	< 0.0001	0.0043
Total Number of Preferences	26		

* corrected for a constant experiment-wise Type I error rate

All participants responded to this item. Again, the number of participants who preferred the Cues within a Method does not necessarily equal the number of persons who preferred that Method. For this item, two participants chose both Cues within one Method as the preferred Cues, and three participants (including one of the previously mentioned participants who selected both Cues within a Method) selected preferred Cues that were not nested in the preferred Method. These anomalies also are addressed in the Discussion section.

Debriefing Results

According to the information gathered in the informal debriefing session, 10 participants felt that the readability of the text was most important, 10 felt that the ability to locate linked words quickly and easily was most important, 2 felt that both were equally important, and 1 participant felt that the importance of one over the other would depend on the task. One (the first) participant was not asked to indicate which was most important because the experimenter had not yet decided to ask for this information.

Of those participants who felt that the salience of linked words was most important, five preferred Character Styles, four preferred Icons, and one preferred Enclosures. Of those participants who felt that the readability was most important, seven preferred Icons, two preferred Character Styles, and one preferred Enclosures.

EXPERIMENT 2: DISCUSSION

Search Times and Lack of Cue Compatibility within Methods

Character Styles are the most salient Cueing Method, with Icons and Enclosures being moderately and least salient, respectively. However, Cues nested within Enclosures and Icons produced significantly different search times, thereby complicating the ability to draw conclusions about these Methods as a whole.

Specifically, participants spent an average of 4.8 seconds locating linked words in text fields employing Brackets, and 1.4 seconds locating linked words cued with Boxes. Similarly, 2.2 seconds was the average time spent locating linked words cued with Flag Icons, compared to 1.5 seconds spent locating those cued with Link Icons. These differences between the Cues within Enclosure and Icons are significant. (Outlined and Shadowed text yielded more similar search times.)

Since Cues within the Icon and Enclosure Methods produced different search times, they are assumed not to be equally salient. The variability that exists within these Methods may have caused the results of comparison tests across Methods to be less conclusive. More comparable Cues that better represent the Methods of Enclosures and Icons might yield more conclusive results in determining which Cueing Method is most quickly and easily located by users of hypertext systems.

Number of Errors

No single Method or Cue is so subtle that it is consistently overlooked by participants (there are no significant differences in the numbers of errors made among Methods). However, the "subtleness" of a Cue (or Method) may be evident from the search times discussed

previously. In other words, more subtle Cues are not overlooked any more often than other Cues, but participants do take significantly longer to locate them. For example, the search times for Bracket Cues significantly exceed those of all other Cues, although no more errors are made in this Cueing condition than in any other. The low percentage of erroneous responses is not surprising, considering the simple nature of the search tasks.

Subjective Ratings

On a relative basis, subjects rated Character Styles (average rating of 1.23 on a seven-point scale) as significantly easier to locate than Icons (average rating of 2.83) or Enclosures (average rating of 3.98) and Icons as significantly easier to locate than Enclosures. This finding supports the results of the search times analyses; the Methods are ranked in the same order for search times and ratings, indicating that the salience of a Cue affects *both* performance measures.

Item #2 of the Absolute Rating Questionnaire, "Did the location of the linked word in the passage affect your ability to locate the word?" was included to determine if Cues were more visible when located in the middle versus the periphery of a text field. Since this item did not produce a significant result, the effectiveness of a Cueing Method is assumed not to be contingent on its location in text.

Lack of Cue Compatibility within Methods for Subjective Ratings

Ratings from the Relative Questionnaire consistently reveal that Character Styles are more salient and, therefore, easier to locate than the other two Methods. Conversely, Enclosures were rated much more difficult to locate than the other Methods and, thus, are not salient Cues. The low ratings for the Enclosures, however, are likely attributable to the lack of appeal for the Brackets, which were much more difficult to locate than were the other Cues (as

evidenced by the search times). Again, a Cue equally as salient as the Box may have yielded more favorable ratings for the Enclosures Method as a whole.

The third item from the Relative Questionnaire reinforces previous results indicating an obtrusive quality of the Character Styles. Even though Character Styles are more salient than Icons and Enclosures, participants indicated that these Cues are also the most distracting. This finding is consistent with the assertion that Cues which are highly salient also tend to degrade the readability of the text. The point must be made, however, that reading time is not affected; only ratings (and preferences, discussed in the following paragraph) of the Methods indicate any degradation of readability.

User Preferences

Within Methods, Link Icons, Boxes, and Shadowed text are preferred by significantly more participants than could be attributed to chance. The Cues within each Method are not at all equally preferred; there is a "significantly more salient" Cue within each Method, again indicating a lack of similarity among the Cues and suggesting the need for a study using more equivalent Cues.

Lack of Cue Compatibility within Methods for User Preferences

It is difficult to determine how Methods might compare if more similar Cues that more accurately represent the Methods were used. When the Cues for this study were originally designed, a slightly more salient variation of the Brackets was chosen. However, the more salient Brackets variation was also more distracting than the variation used in this study because it appeared to blend in with the text (it was difficult to differentiate letters from this Cue). An alternative Cue could have been more "box-like"; however, a more varied sampling of

ways to enclose the word was desired. As a result, the more discernable, though slightly more subtle, variation of brackets was chosen.

The Box Cues provided sufficient salience to render this Cue the unanimously preferred Cue within Enclosures; however, the overall Method of Enclosures produced the longest search times and received the lowest ratings of all three Methods. If an alternative Cue had been used rather than the Brackets, then the Box Cue may not have been the preferred Cue. Furthermore, and more importantly, an alternative Cue might have produced more favorable search times and ratings for the overall Method of Enclosures. This same argument can be applied to the Icons Method, in which Link Icons were preferred by significantly more participants than the Flag Icons. If more comparable Cues had been used within this Method, then the overall Method may have faired more positively in the visual search tasks and the user ratings.

In summary, Character Styles are the preferred Cueing Method for salience. In addition, within each Method there is a strong preference for one Cue in particular. The search times for Cues within the Enclosure and Icon Methods complement their respective preference scores in that the Cues preferred within these Methods (Boxes and Link Icons) also produced significantly shorter search times than did the other Cues (Brackets and Flag Icons) within the same Methods.

GENERAL DISCUSSION AND CONCLUSIONS: EXPERIMENTS 1 AND 2

Debriefing Probe for Importance of Readability versus Salience

The participants' responses regarding the importance of reading without being distracted versus quickly locating linked words were evenly split, with 10 participants placing emphasis on readability, 10 on salience, and 2 placing equal importance on both. This "importance" issue was probed to help the experimenter determine whether the results of the two experiments should be weighted accordingly. The information resulting from this probe indicates that the results from Experiment 1 and Experiment 2 should be equally weighted; readability is no more important than salience, and salience is no more important than readability.

Difference In Reading Times for the Two Control Groups in Experiment 1

An interesting result that was not expected is that the two control conditions containing no linked words (from Experiment 1) yielded significantly different task completion times, with the first condition producing longer times than the second. One participant in particular took much longer to read passages from the first control condition. The data belonging to this participant were removed, and ANOVAs and comparison tests were conducted using the data from the remaining participants. The results reveal no significant differences between the control conditions; however, all other results of the original analyses are unchanged. In other words, the only result of the second analyses that differs from the original analyses is that the two control conditions are no longer significantly different.

Implications of the Anomalies within the User Preferences Data

As was briefly mentioned in the Results sections of Experiments 1 and 2, participants occasionally preferred a Cue overall (Relative Questionnaires for both experiments) that is not nested within the overall Method preferred by that subject. This occurrence implies that those participants did not actually understand the relationship between Cueing Methods and their nested Cues. Therefore, the preference data belonging to these participants were removed and analyses were conducted for the data belonging to the remaining participants. In addition, the data of three different participants (one from Experiment 1 and two from Experiment 2) who selected both Shadowed and Outlined text as the overall preferred Cues within the Character Styles Method (as opposed to just one Cue) were removed.

The results of these parsed analyses are not different from the original results; in other words, eliminating the questionable subjects from the subject pool did not make a difference in the significance (or lack of significance) of the results. For Experiment 1, four subjects were removed for further analysis of the preference for Method. As with the original test, the Cochran Q-statistic is not significant. For the Cue preference the same four subjects were removed, plus a fifth subject who failed to choose a single preferred Cue, and again the results are not significant.

For Experiment 2, three subjects were removed for a follow-up analysis of the Method preferences. The results of the corresponding Cochran Q test and subsequent Binomial test of the most preferred Method (Character Styles) are still significant at $p < 0.001$ and $p < 0.0001$, respectively. The same three subjects plus two others who selected two preferred Cues were removed from the subject pool for analysis of the Cue preferences. Again, the results are still significant for the Cochran Q Test ($p < 0.001$) and the subsequent Binomial test ($p < 0.0001$) for the most preferred Cue (Shadowed text).

Lack of Cue Compatibility within Methods, Experiment 2

The search times (for Icons and Enclosures) and preferences (for all three Methods) from Experiment 2 indicate that there are significant salience differences between Cues within each Method. Because the effectiveness of a Method is contingent upon the effectiveness of its nested Cues, this issue deserves attention in further studies.

Sensitivity and Appropriateness of Measures Used

Tinker passages. For tests of readability (Experiment 1), Tinker passages contained three, five, or seven linked words, resulting in a density of up to 7 linked words for every 30 plain words in the text. This density, perhaps, is higher than would commonly be found in an actual hypertext document. However, the number of linked words per passage was intentionally varied to prevent subjects from anticipating the appearance of passages. Using seven linked words provided a density high enough to ensure that any effect of their presence would appear in the dependent measures. In other words, if each passage had contained only one to three linked words, then their presence may not have been obtrusive enough to elicit any substantial difference in reading times from those of the control condition. Therefore, varying the density of linked words in the passages enhanced the sensitivity of the measure (reading time). Since Tinker passages are standardized (Tinker, 1955), any differences among the reading times across conditions can be attributed to the presence of the linked words or the Cues themselves as opposed to the inherent qualities of the text.

Text fields. The text fields used in this research were not selected from a collection of standardized passages; they were obtained from the Sunday edition of three *New York Times* newspapers. An effort was made, however, to select passages that contained few or no numerical figures, titles, and acronyms (to reduce the number of capital letters in the text), and symbols (such as % and $). Therefore, the text fields that were selected were similar in

appearance. In addition, passages of highly varied informational content and relatively common verbiage (few or no technical or highly irregular words were contained in the passages) were selected and were truncated as necessary to maintain a consistent passage size. The issue of standardization is not as pertinent to the tasks performed in Experiment 2 as it is for those in Experiment 1, since participants did not read the text; they essentially scanned the text fields to locate and count the number of linked words present. Therefore, the selection of passages from newspapers is considered to be a viable procedure for the tasks performed in Experiment 2.

Questionnaires. Both Absolute and Relative Questionnaires were used to obtain information about the Cueing Methods independently and comparatively. Independent ratings gave participants an opportunity to rate all three Methods with theoretically the same freedom to express their evaluations with regard to the merits of the Cues within each Method (without having to compare the Methods to one another). Conversely, relative ratings allowed participants to compare Methods and express their evaluations in terms of "best" or "worst." Both types of ratings provide complementary information regarding the effectiveness of a Method.

Trade-off Issue for Readability and Salience

The primary issue surrounding this research concerns trade-offs; Cues which tend to be highly salient also are likely to degrade the readability of text. By the same token, Cues that do not greatly affect the readability of text are also likely to be non-salient. Even within the Character Styles Method, this trade-off issue is evident; one Cue (Outlined text) is preferred over the other (Shadowed text) for readability, and yet the reverse is true for salience. The most appropriate Cueing Method should provide at least moderate readability and salience.

Icons and Enclosures yielded more favorable ratings (though not reading times or preferences) than Character Styles for readability, and Character Styles yielded more

favorable search times, ratings, and preferences than both Icons and Enclosures for Cue salience. Hence, since highly salient Cues tend to degrade perceived readability, and more subtle Cues tend to be difficult to find, a Cue (or Cueing Method) that provides moderate salience and little distraction is a comfortable and appropriate compromise.

Icons fall into the two categories of moderate salience and minor obtrusiveness. Icons (and the Enclosures Method) were favorably rated for readability, whereas Character Styles were not. Further, since Icons did not introduce problems in reading punctuation (as Enclosures did), it is a superior Method to Enclosures as well as Character Styles. Although Icons are not as salient as Character Styles, they are more salient than Enclosures and thus provide moderate salience.

Conclusions

The following general observations are made with regard to task performance times, participant ratings, and preferences:

Task performance times:

Readability	No single Method was best
Salience	Character Styles were most salient
	Icons were moderately salient
	Enclosures were least salient

Participant ratings:

Readability	Icons were most highly rated
	Enclosures were moderately highly rated
	Character Styles were least highly rated

Salience	Character Styles were most highly rated
	Icons were moderately highly rated
	Enclosures were least highly rated

Participant preferences:

Readability	No single Method was most preferred
Salience	Character Styles were most preferred

Based on these observations, the following conclusions are drawn. Enclosures are not a preferred Cueing Method for either readability or salience. They are, therefore, not considered a best Method. Although Character Styles are the most salient Cues according to search times, participant ratings, and preferences, they are also the most distracting and least highly rated with regard to readability. Because of their highly distracting nature, they also are not considered a best Cueing Method. Note that this judgment is based on readability ratings, since Character Styles are no worse than the other Cueing Methods for reading times (Experiment 1).

The Icons, however, are most highly rated for readability, and are moderately salient according to both timing data and rating data. Therefore, since Icon Cues are comfortable for reading and provide at least moderate salience, this Cueing Method is considered best for indicating linked words within hypertext documents. Further research is needed, however, to determine how Enclosure Cues (particularly Brackets) might be improved to provide more salience to the users.

A heuristic of the intuitive relationship between readability and salience (essentially, trade-offs between the two) is presented in Figure 8. The graph illustrates the negative correlation between text readability and Cue salience. Character Styles tend to be highly salient, yet detrimental to perceived readability (though not reading time). Enclosures do not

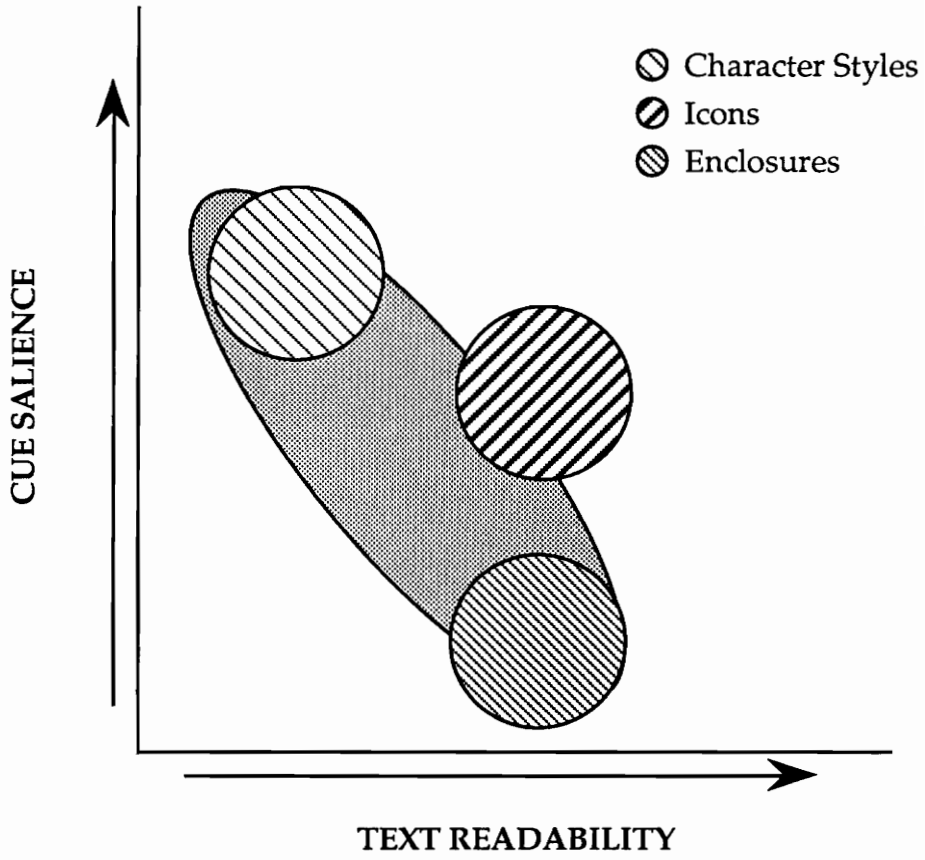


Figure 8. Trade-off relationship between readability and salience.

degrade perceived readability, but they also are not salient. Icons do not greatly degrade perceived readability, nor are they difficult to locate. The location of Icons in this conceptual diagram is most desirable among the three Methods. Note, however, that this graph is not based on actual data points, but merely represents the area that each Method tends to dominate within the salience/readability continuum based on a composite of the results of the two experiments.

Advantages of Icon Cues

Icons provide several advantages. The design of the Icon is flexible; there is an unlimited number of possible Icon designs that can be used and evaluated in future studies. In addition, the placement of an Icon relative the the linked word is flexible. Icons can be symbolic of link types (such as annotation versus reference links), and might therefore serve as mnemonic devices for discriminating among link types. Unlike Enclosures, Icons do not obscure punctuation when they are appropriately placed.

The results of this study provide some direction for further research into appropriate cueing notation for hypertext systems. There is a current need for design guidelines and standards for developers of hypertext systems in this area. Any standards that come out of the research in cueing notation will in turn generate consistency across hypertext systems.

Relationship to Other Studies

Since this is the first known study to investigate alternative ways of indicating linked text within hypertext documents, there are no other research findings either to concur with or to dispute. There is a need for further research in the entire field of hypertext and hypermedia, and the designation of linked text is but one of the issues that deserves attention. In the following section, suggestions are included regarding possible areas of future research.

Future Research

Several related issues discussed in the Literature Review were not addressed within this study and require investigation. Future research is needed to determine if Cues used for single words are also effective for groups of words, or if a secondary form of highlighting could be used in conjunction with the primary Cue to designate lengthy source nodes.

Further research also is needed to determine which specific Cues within a Method are effective. For example, within the Enclosures Method, evaluation of alternatives to the Brackets used in this study (which would provide higher salience for ease of locating linked text) is an area that requires further investigation.

The density of linked words within text is another area that requires further research. Conclusive results from a study of this nature would provide hypertext system designers with some guidelines in the creation of future systems.

REFERENCES

Akscyn, R. M., McCracken, D. D., and Yoder, E. A. (1988). KMS: A distributed hypermedia system for managing knowledge in organizations. *Communications of the ACM, 31*(7), 820-835.

Begeman, M. L., and Conklin, J. (1988). The right tool for the right job. *BYTE*, October 1988, 255-266.

Bush, V. (1945). As we may think. *Atlantic Monthly, 176*(7), 101-108.

Conklin, E. J. (1987). *Hypertext: An introduction and survey*. MCC Technical Report, Number STP-356-86, Rev. 2.

Engelbart, D. C., and English, W. K. (1968). A research center for augmenting human intellect. In *Proceedings of the 1968 Fall Joint Computer Conference* (pp. 395-410). Montvale, NJ:AFIPS Press.

Engelbart, D. C. (1963). A conceptual framework for the augmentation of man's intellect. In W. Howerton and D. C. Weeks (Eds.) *Vistas in Information Handling, Vol 1* (pp. 1-29). London: Spartan Books.

Evenson, S., Rheinfrank, J., and Wulff, W. (1989). Towards a design language for representing hypermedia cues. In *Proceedings of Hypertext '89* (pp. 83-92). New York: ACM Press.

Gordon, S., Gustavel, J., Moore, J, and Hankey, J., (1988). The effects of hypertext on reader knowledge representation. *Proceedings of the Human Factors Society 32nd Annual Meeting*, 296-300. Santa Monica, CA: Human Factors Society.

Goudy, F. W. (1963). *The alphabet and elements of lettering*. Dover, NY: Dover Publications.

Greenhouse, S. W., and Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrica, 24*, 95-112.

Grice, R. A. (1989). Online Information: What do people want? What do people need? In E. Barrett (Ed.), *The society of text: Hypertext, hypermedia, and the social construction of information* (pp. 22-44). Cambridge, MA: MIT Press.

Guide 2 Operating Manual. (1988). Bellevue, WA: OWL International.

Halasz, F. G., Moran, T. P., and Trigg, R. H. (1987). NoteCards in a nutshell. In J. M. Carroll and P. P. Tanner (Eds.) *Proceedings of the ACM conference on human factors in computing systems and graphics interface* (pp. 45-52). Toronto, Canada: Association for Computing Machinery.

Irish, P. M., and Trigg, R. H. (1989). Supporting collaboration in hypermedia: Issues and experiences. In E. Barrett (Ed.), *The society of text: Hypertext, hypermedia, and the social construction of information* (pp. 90-106). Cambridge, MA: MIT Press.

Marchionini, G., and Shneiderman, B. (1988). Finding facts vs. browsing knowledge in hypertext systems. *IEEE Computer, 21*(1), 70-80.

Nelson, T. H. (1980). Replacing the printed word: A complete literary system. In S. H. Lavington (Ed.) *Information Processing '80 Proceedings* (pp. 1013-1023). Amsterdam: North-Holland.

SAS® User's Guide: Statistics, Version 5 Edition. (1985). Cary, N. C.: SAS Institute, Inc.

Shneiderman, B. (1987). User interface design for the Hyperties electronic encyclopedia. In *Hypertext '87 Proceedings* (pp. 189-194). New York: ACM Press.

Siegel, S., and Castellan, N. J. (1988). *Nonparametric statistics for the behavioral sciences, 2nd Edition.* New York: McGraw-Hill.

Tinker, M. A. (1955). *Examiner's manual for Tinker speed of reading test.* Minneapolis: University of Minnesota Press.

Winer, B. J. (1971). *Statistical principles in experimental design, 2nd Edition.* New York: McGraw-Hill.

Yankelovich, N., Smith, K., Garrett, N., and Meyrowitz, N. (1988). Issues in designing a hypermedia document system. In S. Ambron and K. Hopper (Eds.) *Multimedia in education* (pp. 35-87). Redmond, WA: Microsoft Press.

APPENDIX A

CONSENT FORMS

PARTICIPANT'S INFORMED CONSENT

Vision Screening Test

The purpose of this screening exam is to determine whether your vision meets the criteria we have established for participating in our experiment. This is not a professional eye exam; therefore, results should not be considered an accurate description of your visual capability. A professional eye care practitioner should be consulted for an accurate assessment of your vision.

This eye test consists of two parts. If your performance meets the criteria for both parts you will be asked to participate in the experiment. This screening will take approximately 15 minutes to complete.

This research is being conducted by the Human Factors Laboratory of the Department of Industrial and Systems Engineering by:

Rani L. Watkins	Student, Master of Science (231-9092)
Dr. Harry L. Snyder	R. H. Bogle Professor and Principal Investigator (231-7527)

If you have any questions regarding this research, please contact the researchers at the numbers indicated. You may also contact Dr. E. R. Stout, Chairman of the Institutional Review Board (231-5281). Dr. Stout oversees research involving human participants at Virginia Tech.

As a participant in this screening exam, you have certain rights. The purpose of this form is to make you aware of these rights and to obtain your informed consent.

You have the right to do the following:

1) To stop participating in the exam at any time. If you decide to do so, you must notify the experimenter immediately.

2) To see the results of your eye exam. You can ask the experimenter for the results of your eye exam. However, keep in mind that these results were not obtained by a professional eye doctor.

Your signature below indicates that you have read the above stated rights and that you consent to participate.

PARTICIPANT'S INFORMED CONSENT

Title of the Study: An Investigation into Methods of Indicating of Indicating Active Words in Hypertext Documents

Researcher: Rani L. Watkins

You have read the description of this experiment and understand that the purpose of this experiment is to examine your response to information that will be presented to you on a display screen using various different cueing methods. If you have any questions regarding this research please ask the experimenter now. If she cannot answer your questions you may contact the Faculty Research Advisor for this project, Dr. Snyder, at 231-7527, or the University Internal Review Board Chairman, Dr. Stout, at 231-5281.

You will be paid $5.00 per hour for your participation in this experiment. The study will take approximately one and a half hours to complete. If for some reason you cannot complete the research you will be paid for the time you have participated.

As a participant in this experiment you have certain rights. The purpose of this form is to make you aware of these rights and to obtain your informed consent.

You have the right to do the following:

1) To stop participating in the experiment at any time. If you decide to do so, you must notify the experimenter immediately.

(2) To see your data and to withdraw them from the experiment. All data are treated with anonymity; therefore, if you wish to withdraw your data you must do so immediately after your final task has been completed.

(3) To be informed of the overall results of the experiment. If you wish to receive information about the results, please include your address with your signature on the accompanying informed consent form. A summary will be sent to you in approximately three months. For further information you may contact the Human Factors Laboratory and a full report will be made available to you.

Your signature below indicates that you have read the above stated rights and that you consent to participate.

_____	_____	_____
Print name	Sign name	Date

To receive a summary of the results:

_____

_____
Address

APPENDIX B

SUMMARY OF HYPERTEXT

INTRODUCTION TO THE STUDY

PARTICIPANTS' INSTRUCTIONS

Hypertext

A Brief Description

Hypertext is a technology used with computers. It allows a user of the computer to read an online document in a "non-linear" fashion. If the user reads something of particular interest in the document and wishes to find more information about that topic, a key word can be selected (usually with a mouse click), and the reader is taken directly to an area of the document that contains related or more detailed information.

One of the many applications of hypertext is in education. Suppose you were taking a course entitled "Social Problems," and that your professor had a series of course-related articles in hypertext format available to students in one of the computer centers. While browsing through the tutorial you may come across an article about the Civil Rights Movement. The author may have briefly mentioned Affirmative Action in the article. If Affirmative Action is somehow cued to indicate that it is "linked" to another area of the online system, you could click on the the words Affirmative Action and go directly to a more detailed discussion of this topic in the same article or in another article within the system.

The purpose of this study is to investigate various cues for indicating linked words within hypertext documents. Your performance on the tasks involved will help the experimenters determine the effectiveness of alternative methods identifying these active words.

An Investigation into Methods of Indicating Active Words in Hypertext Documents

Introduction

The purpose of this study is to evaluate three alternative ways of indicating that a word is linked within a hypertext document. The three methods include character styles, word enclosures, and icons.

This study is being conducted by the Controls and Displays Laboratory, Department of Industrial and Systems Engineering, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061 (telephone number: (703) 231-5358). The research assistant conducting this study, Rani Watkins, is a graduate student in Industrial and Systems Engineering, under the direction of Dr. Harry L. Snyder, principal investigator and R. H. Bogle Professor of Industrial and Systems Engineering (telephone number (703) 231-7527).

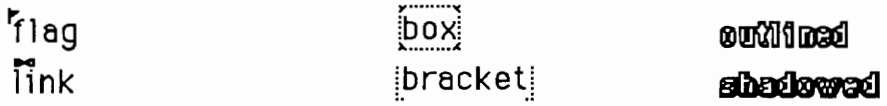
In this study you will be asked to perform two tasks in two different experiments. You will receive written instructions regarding your tasks for these two experiments.

After reading this introductory material, you will be given an informed consent form. If you understand what this experiment entails and agree to participate in it, you must sign this form. You will also be asked to take a visual acuity test as a screening procedure. This test poses no threat to you as a participant. Assuming that you meet all of our standardization criteria, you will then proceed with the experiment. The following is a brief outline of the experiment you will participate in.

Experimental Procedure

You will first be seated in front of a Macintosh computer. Feel free to adjust your seat to the height that is most comfortable for you. For this experiment you will see various passages

containing "linked" words. As you have read read in "Hypertext - A Brief Description," these linked words can be activated to access related information. All linked words will be indicated using one of the following cues:



You will participate in two different experiments. For each experiment you will read a description of your tasks so you will know exactly what you need to do.

Experiment 1:

The first message you will see will ask for your participant number. Simply type the number given to you by the experimenter into the space provided and hit "Return." The screen you will see contains the following message: "To begin, click on the button 'Next.'" **Please click only once on this button.** Upon clicking on "Next" the first passage will appear. You will notice that the passage contains a number of "linked" words. You will read the entire passage and locate the word that does not logically correspond to the rest of the passage. **Immediately after** you have located the word you will click on the "Next" button. This will take you to a screen containing the following question, "Which word does not belong in the message?" At this point you will say the word out loud so the experimenter can hear and record it. When you are ready to go to the next passage, you will again click on the "Next" button. You will repeat this process for the next nine screens you will see. After the tenth passage you will come to a screen with the message "You have completed this part of the experiment." When you are ready you may click on the button "Click here to continue" located below this question. The next screen you will see should look familiar, and the process will repeat itself. After every other trial you will be asked to fill out a questionnaire. The experimenter will provide these to you at the appropriate times.

You will continue this process until all passages have been read, at which point the following message will appear on the screen: "Experiment 1 has been completed." You will have read a total of 80 passages at this point. Next, you will be asked to complete a questionnaire pertaining to this first experiment.

Experiment 2:

The first message you will see will ask for your participant number. Simply type the number given to you by the experimenter into the space provided and hit "Return." The screen you will see contains the following message: "To begin, click on the button 'Next.'" **Please click only once on this button.** Upon clicking "Next" the first text field will appear. You will scan the passage and count all "linked" words, identified by one of the six cues used. **Immediately after** you have counted all "linked" words, you will click on the "Next" button. This will take you to a screen containing the following question, "How many linked words are there in the passage?" At this point you will say the number of words you counted out loud so the experimenter can hear and record it. When you are ready to go to the next passage, you will again click on the "Next" button. You will repeat this process for the next nine screens you will see. After the tenth passage you will come to a screen with the message "You have completed this part of the experiment." When you are ready you may click on the button "Click here to continue" located below this message. The next screen you will see should look familiar, and the process will repeat itself. After every other trial you will be asked to fill out a questionnaire. The experimenter will provide these to you at the appropriate times.

You will continue this process until you have seen all text fields, at which point the following message will appear on the screen: "Experiment 2 has been completed." You will have seen a total of 60 text fields at this point. Next, you will be asked to complete a questionnaire regarding this first experiment.

Throughout the experiment, an experimenter will be seated beside you to record your verbal responses.

Practice:

You will go through several practice screens prior to each experimental trial. Experiment 1 consists of eight experimental trials, and Experiment 2 consists of six experimental trials.

To practice, you will follow the same procedure as you would with an experimental trial (enter your participant number, click on the "Next" button, etc.). You will begin each experimental trial immediately after you have completed the practice trial. At the end of each practice trial, you will be informed that the experimental trial will begin. You will be asked to relax for about one minute before the experimenter tells you that you may begin. Note that the cues

used for the practice trial will also be used for the experimental trial. The practice trials should give you an idea of what to look for in the experimental trials and how to respond to the passages.

If you have any questions, please ask the experimenter before you begin the experimental trials.

Additional Information

You have the right to terminate your participation at any time during the study, if you no longer wish to continue. You will be compensated for the length of your participation up to that point.

If you have any questions about the experiment or your rights as a participant after reading the attached informed consent form, please do not hesitate to ask. We will answer your questions honestly, and as openly as possible. We ask that you do not discuss the details of this experiment with any person, particularly with those who might participate, since prior familiarity of seemingly incidental information could compromise the data. This study will be completed by August 9, after which time you can freely discuss the experiment. All data will be analyzed with anonymity, i.e., immediately upon completion of your experimental session, your data will be identified only by a randomly assigned serial number.

Experiment 1

For Experiment 1, your task will be to read a series of four-line passages and locate the word that does not logically belong. Once you have found the word, you will click on the "Next" button to go to the next screen containing the question "Which word does not belong in the passage?" You will then verbalize the inappropriate word. You will do this for each passage that appears on the screen. You will begin simply by clicking on the "Next" button when the experimenter indicates that you may start. You will notice that some of the words you will see in the passages will be "linked" words. Try to complete your task as quickly and accurately as you can, despite the presence of these words. The experimenter will record the words you have verbalized.

Experiment 2

For Experiment 2, you will be asked to scan a series of text fields and count the number of "linked" words contained within each passage. Once you have counted all words, you will click on the "Next" button to go to the next screen containing the question "How many linked words are there in the passage?" You will then verbalize the number of "linked" words you counted. You will do this for each text field that appears on the screen. You will begin simply by clicking on the "Next" button when the experimenter indicates that you may start. Try to complete your task as quickly and accurately as you can. The experimenter will record the number you have verbalized.

APPENDIX C

QUESTIONNAIRES

Experiment 1
Absolute

Rate the **icons** (symbols) you used on the following dimensions by circling the appropriate number.

1) Could you distinguish punctuation marks from icons?

Very easily	1	2	3	4	5	6	7	Not at all

2) Did you have to read passages more than once?

Very often	1	2	3	4	5	6	7	Never

3) Could you notice the "linked" words (words with icons) in the passages?

Very easily	1	2	3	4	5	6	7	Not at all

4) Were the icons distracting?

Very	1	2	3	4	5	6	7	Not at all

5) Did you experience any eye strain while reading the passages as a result of an icon being present?

Not at all	1	2	3	4	5	6	7	A great deal

Did you prefer one *icon* (or symbol) over the other for this task?

If so, which one? (circle)

Experiment 1
Absolute

Rate the **enclosures** you used on the following dimensions by circling the appropriate number.

1) Could you distinguish punctuation marks from enclosures?

Very easily	1	2	3	4	5	6	7	Not at all

2) Did you have to read passages more than once?

Very often	1	2	3	4	5	6	7	Never

3) Could you notice the "linked" (enclosed) words in the passages?

Very easily	1	2	3	4	5	6	7	Not at all

4) Were the enclosures distracting?

Very	1	2	3	4	5	6	7	Not at all

5) Did you experience any eye strain while reading the passages as a result of an enclosure?

Not at all	1	2	3	4	5	6	7	A great deal

Did you prefer one *enclosure* over the other?

If so, which one? (circle)

Experiment 1
Absolute

Rate the **character styles** you used on the following dimensions by circling the appropriate number.

outlined

shadowed

1) Could you distinguish punctuation marks from character styles?

Very easily	1	2	3	4	5	6	7	Not at all

2) Did you have to read passages more than once?

Very often	1	2	3	4	5	6	7	Never

3) Could you notice the "linked" words (words either outlined or shadowed) in the passages?

Very easily	1	2	3	4	5	6	7	Not at all

4) Were the character styles distracting?

Very	1	2	3	4	5	6	7	Not at all

5) Did you experience any eye strain while reading the passages as a result of the character style?

Not at all	1	2	3	4	5	6	7	A great deal

Did you prefer one *character-style* over the other?

If so, which one? (circle)

outlined

shadowed

Experiment 1
Relative

Use this key to rate the passages on the dimensions that follow as a function of the cueing method used. METHODS: I = Icons E = Enclosures C = Character-styles

Simply write the letter of the corresponding method in the appropriate space above the ranking. Place only one letter in a space.

Example: Quality of the text?

Very High	X 1	__ 2	__ 3	Y 4	Z 5	__ 6	__ 7	Very Low

---

1) Easy to read?

Very Easy	__ 1	__ 2	__ 3	__ 4	__ 5	__ 6	__ 7	Very Difficult

2) Distracting?

Very Distracting	__ 1	__ 2	__ 3	__ 4	__ 5	__ 6	__ 7	Not Noticable

3) Comfortable to use?

Very Comfortable	__ 1	__ 2	__ 3	__ 4	__ 5	__ 6	__ 7	Very Uncomfortable

4) Problems reading punctuation?

Often	__ 1	__ 2	__ 3	__ 4	__ 5	__ 6	__ 7	Never

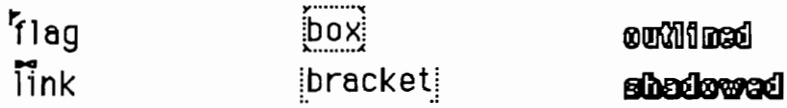
5) Problems finding the target "wrong" word?

Often	__ 1	__ 2	__ 3	__ 4	__ 5	__ 6	__ 7	Never

Did any particular *method* seem to be more intuitive or logical for representing "linked" words?

Icons Enclosures Character-styles

Did any particular *cue* seem to be more intuitive or logical for representing "linked" words?



Experiment 2
Absolute

Rate the **icons** you used on the following dimensions by circling the appropriate number.

1) Were the icons easy to locate?

Very easy	1	2	3	4	5	6	7	Very difficult

2) Did the location of the "linked" word in the passage affect your ability to locate the word?

Absolutely	1	2	3	4	5	6	7	Not at all

Was one icon easier to locate than the other?

If so, which one? (circle)

Experiment 2
Absolute

Rate the **enclosures** you used on the following dimensions by circling the appropriate number.

box

bracket

1) Were the enclosures easy to locate?

Very easy	1	2	3	4	5	6	7	Very difficult

2) Did the location of the "linked" word in the passage affect your ability to locate the word?

Absolutely	1	2	3	4	5	6	7	Not at all

Was one enclosure easier to locate than the other?

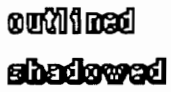
If so, which one? (circle)

box

bracket

Experiment 2
Absolute

Rate the **character-styles** you used on the following dimensions by circling the appropriate number.



1) Were the character styles easy to locate?

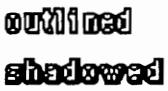
Very easy	1	2	3	4	5	6	7	Very difficult

2) Did the location of the "linked" word in the passage affect your ability to locate the word?

Absolutely	1	2	3	4	5	6	7	Not at all

Was one character style easier to locate than the other?

If so, which one? (circle)



Experiment 2
Relative

Use this key to rate the passages on the dimensions that follow as a function of the cueing method used. METHODS: I = Icons E = Enclosures C = Character-styles

Simply write the letter of the corresponding method in the appropriate space above the ranking. Place only one letter in a space.

Example: Quality of the text?

Very High	X 1	__ 2	__ 3	Y 4	Z 5	__ 6	__ 7	Very Low

---

1) Easy to find linked words?

Very Easy	__ 1	__ 2	__ 3	__ 4	__ 5	__ 6	__ 7	Very Difficult

2) "Linked" words seemed to stand out from the rest of the text?

Very Noticable	__ 1	__ 2	__ 3	__ 4	__ 5	__ 6	__ 7	Not Noticable

3) If you had been reading the passages more thoroughly, would the cues have been distracting?

Very Distracting	__ 1	__ 2	__ 3	__ 4	__ 5	__ 6	__ 7	Not at all

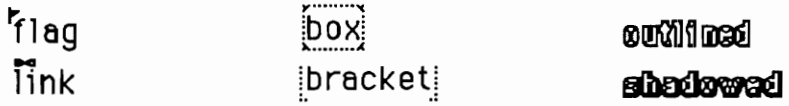
4) Did you experience any difficulty finding the "linked" words?

Often	__ 1	__ 2	__ 3	__ 4	__ 5	__ 6	__ 7	Never

Was a particular *method* better overall for representing "linked" words?

Icons Enclosures Character-styles

Was a particular *cue* better overall for representing "linked" words?



VITA

Rani Lea Watkins was born in Springfield, Massachusettes on May 7, 1966. She received her B. S. degree in Psychology from Virginia Polytechnic Institute and State University in May, 1988, and her M. S. in Industrial and Systems Engineering from Virginia Polytechnic Institute and State University in May, 1991. She is currently an intern with the Office of Nuclear Regulatory Research within the United States Nuclear Regulatory Commission in Rockville, Maryland. Rani's research and design interests include human information processing, human-computer interaction, and hypermedia systems.