AD-A012 591

ANALYSIS OF WORD PROPERTIES AND LIMITED READING TIME IN A SENTENCE COMPREHENSION AND VERIFICATION TASK

David Edward Kieras

Michigan University

Prepared for:

Air Force Office of Scientific Research
Advanced Research Projects Agency

August 1974

010588-22-T

HUMAN PERFORMANCE CENTER
DEPARTMENT OF PSYCHOLOGY
The University of Michigan, Ann Arbor

Analysis of Word Properties and Limited Reading Time in a Sentence Comprehension and Verification Task

DAVID E. KIERAS

Technical Report No. 52
August 1974

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER AFOSR - TR - 75 - 0953	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER AD-A Ø12 591
4. TITLE (and Subtitle) Analysis of word properties and limited reading time in a sentence comprehension and verification task		5. TYPE OF REPORT & PERIOD COVERED Interim Technical Report
		6. PERFORMING ORG. REPORT NUMBER Technical Report No. 52
7. AUTHOR(s) David Edward Kieras		8. CONTRACT OR GRANT NUMBER(s) F44620-72-C-0019
9. PERFORMING ORGANIZATION NAME AND ADDRESS Human Performance Center 330 Packard Road Ann Arbor, Michigan 48104		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 61101D/1949/681313
11. CONTROLLING OFFICE NAME AND ADDRESS Advanced Research Projects Agency 1400 Wilson Blvd. Arlington, Va. 22209		12. REPORT DATE August 1974
		13. NUMBER OF PAGES 126
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) Air Force Office of Scientific Research (NL) 1400 Wilson Blvd. Arlington, Va. 22209		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release;
distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

imagery value
association value
acquisition
dual-representation memory model
semantic network memory model

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This study contains an experimental investigation and theoretical analysis of the effects of limited reading time and the effects of imagery value, association value, and frequency of sentence words, in a task involving comprehension and subsequent verification of meaningful sentences. The theoretical analysis was concerned with how these word properties could be represented in a semantic network memory model and how their effects could be explained in terms of the memory structure and related processes.

PRICES SUBJECT TO CHANGE

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

The experiment used sentence materials that varied othogonally in imagery value, association value, and frequency of the constituent words. A set of sentences that fit together to describe a set of distinct complete ideas, or events, were presented to subjects during acquisition; the sentences were presented one at a time in random order. The reading time manipulation consisted of either allowing subjects to read at their own rate, or forcing them to read the acquisition material at a fast rate. The time self-paced subjects took to read each acquisition sentence was recorded as a measure of comprehensibility.

After reading all the sentences describing the set of events, subjects answered a series of true-false questions about the events; the speed and accuracy of sentence verification was recorded. One type of true question could be answered only if the subject recognized that the meaning of the question sentence question agreed with the events described during acquisition. A second type of true question was a repetition of an acquisition sentence, and thus could be answered on the basis of either meaning recognition or verbatim recognition of the sentence. The process of answering the two question types, which allowed comparison of memory for meaning and memory for wording, was described with a formal model that produced estimates for the parameters governing subprocesses in the verification task. These estimates assisted the theoretical interpretation of the effects of the word properties and the reading time limitation.

The limitation of reading time increased verification speed, due to forced-paced subjects acquiring a speed-reading set, and impaired recognition of meaning by causing the acquisition information to be stored in a less complete and less accessible form. Higher sentence word frequency facilitated comprehension but lowered verification accuracy because acquisition information expressed in high frequency words was confused with pre-existing information and was thus harder to locate in the memory network. Higher sentence word association value improved acquisition storage, as shown by shorter reading time and higher verification accuracy, by facilitating the re-use of pre-existing memory information. Higher sentence word imagery value produced much faster and more accurate sentence processing during acquisition and verification. These effects were tentatively explained in terms of a dual-representation memory model. Interactions of imagery with other factors were used to argue the merits of several explanations for the imagery effects.

The general conclusion was that the effects of reading time, and the word properties of imagery value, association value, and frequency, could be accounted for with extensions of current information processing models of memory.

THE HUMAN PERFORMANCE CENTER

DEPARTMENT OF PSYCHOLOGY

The Human Performance Center is a federation of research programs whose emphasis is on man as a processor of information. Topics under study include perception, attention, verbal learning and behavior, short- and long-term memory, choice and decision processes, and learning and performance in simple and complex skills. The integrating concept is the quantitative description, and theory, of man's performance capabilities and limitations and the ways in which these may be modified by learning, by instruction, and by task design.

The Center issues two series of reports. A Technical Report series includes original reports of experimental or theoretical studies, and integrative reviews of the scientific literature. A Memorandum Report series includes printed versions of papers presented orally at scientific or professional meetings or symposia, methodological notes and documentary materials, apparatus notes, and exploratory studies.

THE UNIVERSITY OF MICHIGAN

COLLEGE OF LITERATURE, SCIENCE AND THE ARTS

DEPARTMENT OF PSYCHOLOGY

ANALYSIS OF THE EFFECTS OF WORD PROPERTIES AND LIMITED READING TIME IN A SENTENCE COMPREHENSION AND VERIFICATION TASK

David Edward Kieras

HUMAN PERFORMANCE CENTER--TECHNICAL REPORT NO. 52

August, 1974

This research was supported by the Advanced Research Projects Agency, Department of Defense, and monitored by the Air Force Office of Scientific Research, under Contract No. F44620-72-C-0019 with the Human Performance Center, Department of Psychology, University of Michigan.

Approved for public release;
distribution unlimited.

PREFACE

This report is an independent contribution to the program of research of the Human Performance Center, Department of Psychology, on human information processing stress factors, supported by the Office of the Advanced Research Projects Agency, under Order No. 1949, and monitored by the Air Force Office of Scientific Research under Contract No. F44620-72-C-0019.

This report was also a dissertation submitted by the author in partial fulfillment of the degree of Doctor of Philosophy (Psychology) in the University of Michigan, 1974. The doctoral dissertation committee was Drs. E. Martin, Chairman, W. M. Kincaid, R. W. Pew, and J. R. Anderson.

TABLE OF CONTENTS

ABSTRACT

This study contains an experimental investigation and theoretical analysis of the effects of limited reading time and the effects of imagery value, association value, and frequency of sentence words, in a task involving comprehension and subsequent verification of meaningful sentences. The theoretical analysis was concerned with how these word properties could be represented in a semantic network memory model and how their effects could be explained in terms of the memory structure and related processes.

The experiment used sentence materials that varied othogonally in imagery value, association value, and frequency of the constituent words. A set of sentences that fit together to describe a set of distinct complete ideas, or events, were presented to subjects during acquisition; the sentences were presented one at a time in random order. The reading time manipulation consisted of either allowing subjects to read at their own rate, or forcing them to read the acquisition material at a fast rate. The time self-paced subjects took to read each acquisition sentence was recorded as a measure of comprehensibility.

After reading all the sentences describing the set of events, subjects answered a series of true-false questions about the events; the speed and accuracy of sentence verification was recorded. One type of true question could be answered only if the subject recognized that the meaning of the question sentence question agreed with the events described during acquisition. A second type of true question was a repetition of an acquisition sentence, and thus could be answered on the basis of either meaning recognition or verbatim recognition of the sentence. The process of answering the two question types, which allowed comparison of memory for meaning and memory for wording, was described with a formal model that produced estimates for the parameters governing subprocesses in the verification task. These estimates assisted the theoretical interpretation of the effects of the word properties and the reading time limitation.

The limitation of reading time increased verification speed, due to forced-paced subjects acquiring a speed-reading set, and impaired recognition of meaning by causing the acquisition information to be stored in a less complete and less accessible form. Higher sentence word frequency facilitated comprehension but lowered verification accuracy because acquisition information expressed in high frequency words was confused with pre-existing information and was thus harder to locate in the memory network. Higher sentence word association value improved acquisition storage, as shown by shorter reading time and higher verification accuracy, by facilitating the re-use of pre-existing memory information. Higher sentence word imagery value produced much faster and more accurate sentence processing during acquisition and verification. These effects were tentatively explained in terms of a dual-representation memory model. Interactions of imagery with other factors were used to argue the merits of several explanations for the imagery effects.

The general conclusion was that the effects of reading time, and the word properties of imagery value, association value, and frequency, could be accounted for with extensions of current information processing models of memory.

CHAPTER I

THEORETICAL CONSIDERATIONS

Introduction

Two Approacnes to Human Memory

This thesis is motivated by the results of two characterizations of human memory which describe the memory information that is related to individual words. The two approaches, or models, will be termed semantic and associative.

The associative model has a long history in psychology, underlying the many studies of the learning of lists of isolated or paired words in the field of verbal learning. Such learning was viewed as the process of forming associations between words; hence a person's knowledge about a word or the meaning of a word was in the form of associations with other words. Some psychologists, such as Deese (1962), studied large sets of words associatively related to each other; such sets may be represented as a structure consisting of associations interconnecting words in a network-like fashion. Other psychologists, such as Kintsch (1970), used structures of this form to account for performance in free-recall learning where the effects of associative organization are especially strong. The FRAN model of Anderson (1972) successfully simulated many aspects of free recall performance using a memory containing a network of associatively interconnected words.

The semantic model of memory originated more recently as a result of developments in linguistics and artificial intelligence. Katz and Fodor (1963) described the information required to generate alternate meanings of a sentence as a structure of features and distinguishers attached to individual words. Features represented general primitive properties that capture the regularities underlying word usage; distinguishers represented the specific meanings of the words and were not treated in any detail by Katz and Fodor since they were unnecessary to account for the regularities. Quillian (1968, 1969) devised a natural language processing simulation in which the memory information corresponding to both features and distinguishers was represented in the form of a structure of concepts interconnected by various types of logical and semantic relations, again in a network-like form. Quillian's contention was that this type of knowledge, which is, broadly speaking, knowledge of the world, is essential to language comprehension. This semantic network conception of memory had a strong impact on psychology, leading to a large amount of experimental and theoretical work, represented by Collins and Quillian (1969, 1972), Bobrow (1970), Anderson and Bower (1971, 1972, 1973), Kintsch (1972), and Rumelhart, Lindsey, and Norman (1972). In line with current usage, the term network model will often be used to refer to semantic network models.

The two approaches to memory have some interesting relations to each other. Both semantic and associative information can be represented by network structures; the associative structure can be viewed as a simplification of the semantic structure if the concepts are represented by words, and undifferentiated semantic relations play the role of associative connections. Despite the close relationship of the two approaches, in most semantic models the nature of the memory structure is based on semantic, logical, or efficiency considerations; the variety of work properties studied in verbal learning plays no obvious role in the semantic network. That associative word properties can be a factor in semantically-oriented tasks is vividly demonstrated by the great importance of association frequency effects in the Collins and Quillian (1969) semantic memory retrieval task (Conrad, 1972; Wilkins, 1971), which is seemingly under the control of semantic relations between words.

Since semantic information is essential to language comprehension, the semantic network conception of memory will probably become a permanent fixture in psychological theory. On the other hand, the associative approach has yielded much information about memory in a learning context and includes the extensive study of several word properties. There has been little effort to combine the predominantly simulation-based semantic notions with the data-based associative approach, despite the fact that both can be represented naturally with a network model.

The Problem

The problem attacked by this thesis is to combine the associative and semantic approaches by developing ways in which some important associative word properties can be represented in a semantic network model and determining whether in a language comprehension task these word properties have effects that are consistent with their network representations. If the associative properties have no effect, then representing them in a language processing model is not vital; on the other hand, if they do have effects, and can not be represented satisfactorily in a semantic network, then an area of inadequacy for the semantic model is revealed that can act as a focus for further research. However, if the properties are both effective and reasonably representable, then a contribution has been made to the understanding of memory by the combination of two different sets of theoretical concepts and data.

Method of Investigation

Sentences were composed using words that differed on three associative properties: imagery value, association value, and frequency. These sentences were used in a reading comprehension experiment. Subjects first performed an acquisition task consisting of reading a randomly ordered set of sentences describing a set of distinct events, as in the Bransford and Franks (1971) study, and then performing a verification task in which they answered a series of true-false questions about the described events. Certain characteristics of the questions and the time available during acquisition were also manipulated in ways that will be described later. Two sets of performance data were collected: the time spent in reading each acquisition sentence, and the speed and accuracy of answering each question.

Solving the thesis problem entailed providing a representation of the word properties in a semantic network and accounting for the effects of the word properties on performance in terms of these representations and the related processing. An important part of this effort was the development of a descriptive model that produced estimates of parameters governing subprocesses in the verification task; these estimates were used as measures of the effects produced by the experimental manipulations.

The remainder of this chapter is preparation for the next chapters which describe the actual experiment, results, and interpretation. The next section of this chapter discusses the network memory models and the use of word information for comprehension, followed by a description of memory representations for the associative word properties used in the experiment. The third section presents the descriptive model of verification performance, and the last section summarizes the material in the entire chapter.

The Role of Word Information in the Memory Network and in Language Processing

The Network Memory Models

The network memory models, reviewed by Frijda (1972) and Anderson and Bower (1973, Ch. 4), all assume that the knowledge stored in memory is organized in terms of a network of concepts interconnected by relations. A convenient representation of such a structure is a graph with the nodes standing for concepts and the labelled links standing for the relations between concepts. The nodes are usually of two kinds: A *type* node is the unique representative of a concept; the structure attached to a type node constitutes a definition of the concept. A *token* node represents a particular reference to a concept; by use of token nodes, knowledge involving a concept can be kept distinct from the concept itself, and different items of knowledge about the same concept can be distinguished from each other (cf. Anderson & Bower, 1973, pp. 19-21). The kinds of links vary widely in different models. The link types used in the Anderson and Bower (1973, Ch. 7) model, known as Human Associative Memory (HAM), are based on logical considerations and are few and simple, unlike other models which use a large variety of links based on linguistic distinctions.

The HAM Model

The theoretical discussion throughout this thesis is based on the HAM model, which despite certain serious inadequacies in connection with language comprehension, does have important advantages as a general theoretical framework. There are many competing network models; the problem is not really which model is best, but whether network models in general can account for the data in a natural and useful way. Since Anderson and Bower have progressed further than others in bringing a complex simulation-oriented memory model into contact with experimental data, the HAM model is a prime candidate for further work and is adopted here as the current best representative of the network model approach.

Representation and Retrieval in HAM

The HAM model is based on the notion of knowledge being expressible in terms of a few basic logical distinctions. In HAM, information is stored in the form of propositions; a proposition is a predication, a subject connected to a predicate, together with a context stating the location and time in which the predication is considered as a fact. Often it is convenient to use the term proposition in a broad sense, referring to any predication, rather than just restricted to a predication-context pair. A subject can be a token for any appropriate concept; a predicate can be either simply a token for a concept, as in the case of a simple modification, or a relation-object pair, indicating that the subject stands in the specified relation to the object. The relation is specified by means of a token of the relation concept; an object may be a token of a concept of another predication or proposition.

Figure 1 gives the propositional structure in node and link network form for a typical sentence used in this study. The links are labelled according to type; the links labelled ε are essentially token-to-type links that connect a token node for a concept to the type node elsewhere in memory. The propositional structure connected to the type node comprises a definition of the concept and all of the knowledge the system has about the concept. All links are two-way connections; the

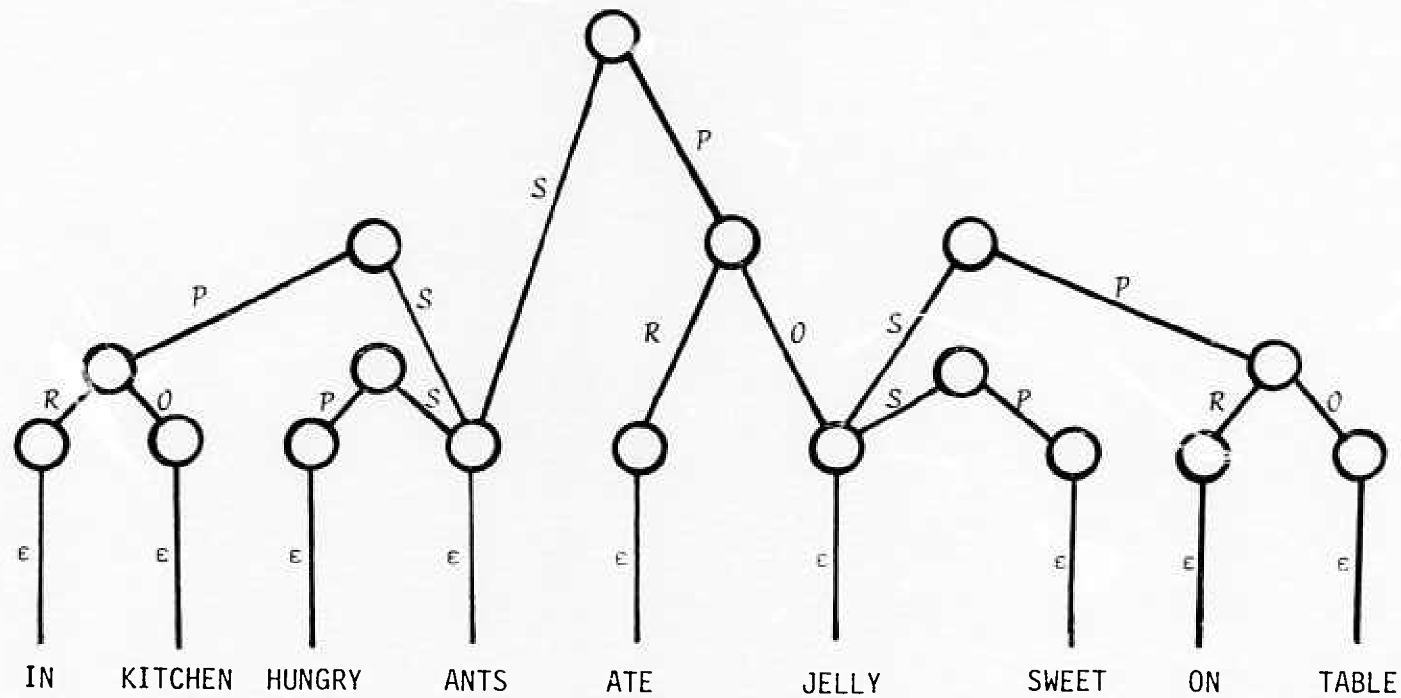

Figure 1. Example of a propositional representation in HAM. The abbreviations for the link types are: S = subject, P = predicate, R = relation, O = object, and ε = token-to-type link. The words at the bottom represent the type nodes for the corresponding concepts.

reverse ε connections, denoted by ε^{-1} links, permit a proposition to be accessed from the concept nodes referenced in the proposition. Thus a specific piece of knowledge can be obtained from a specification of what the knowledge is about.

The HAM model contains a general purpose memory search process, named the MATCH process (Anderson & Bower, 1973, p. 237 ff), that is used in both acquisition and retrieval operations. The process attempts to match a given <u>probe</u> proposition with a proposition in the total memory structure. Two propositional structures are said to <u>match</u> if they have the same configuration of links and nodes and corresponding terminal nodes that refer to the same concepts. Under this definition, it is possible that a proposition could match a subpart of a more complex proposition. Since searching all of memory one proposition at a time is impractical, certain assumptions, to be discussed later, are included in the model to improve the efficiency of the search and set limits on its extent.

The MATCH process works with either a retrieval or recognition probe. A retrieval probe proposition is incomplete, with some terminal nodes unspecified. The model discovers the best match to the probe structure and then fills in the empty nodes. In the recognition case, the probe is fully specified and the model attempts to find a match for it. Thus, in the retrieval case, the MATCH process returns memory information not in the probe, whereas in the recognition case it determines whether the probe information is already represented in memory.

An important feature of the HAM model is that the recognition match process is used during acquisition to determine if the input proposition has any overlap with information already in memory. If so, the new proposition can be added on to the old, with as much of the old structure used as possible; only the unduplicated nodes and links need be added to the memory structure. This use of the MATCH process during acquisition means that factors affecting retrieval will also affect the storage of new information that is related to preexisting memory information.

Comprehension in HAM

HAM as presently developed in Anderson and Bower (1973) was not intended to be a thorough model of the process of comprehending language. This limitation shows up in two areas: (a) The present HAM model has only a crude ability to deal with English syntax; only sentences conforming to a highly simplified grammar are acceptable as input. (b) The present HAM model makes no use of semantic or general knowledge information in the parsing of a sentence; on the basis of syntax alone, the simple grammatical component produces a complete propositional structure for the sentence that is then given to the storage processes.

Because of these limitations, applying the HAM notions to comprehension is treading on unexplored ground. The best extension would be to make use of the semantic processing devised by Quillian in his Teachable Language Comprehender program (Quillian, 1969). In this model, all the input words are looked up simultaneously in semantic memory using an intersection search technique (Quillian, 1968), which finds paths interconnecting the words in memory. These interconnecting paths specify possible semantic relationships between the words which can be used to construct the proposition underlying the input sentence. For example, in the phrase lawyer's client, semantic memory contains the propositions that a lawyer is a professional, and a client is someone who employs a professional. Hence the semantic information alone specifies the core meaning of the phrase as that of someone (the client) employing someone else (the lawyer). Note that a similar result would obtain even if the surface order of the phrase were reversed, as in client's lawyer. Unlike other language processing systems, processing in this system is not based on syntactic information; rather, syntax is only used where essential, to disambiguate the semantics, so to speak. Quillian's (1969) program operated only on short phrases; the idea has yet to be given a more thorough test.

The semantics-based approach to language processing is likely to be the most realistic conception of how people comprehend language. The fact that people normally handle language that is often seriously ungrammatical, and even deal easily with extremes such as understanding children's attempts at language and "Tarzan talk," is a strong case for the conclusion that constraints set by the semantic, contextual, and general knowledge in memory are the primary methods used to interpret sentences, rather than the use of surface syntactic structures (cf. Bransford & Johnson, 1972).

Thus the memory information playing the most important role in comprehension is not used in HAM's input processing. This is not necessarily a defect in the HAM conception, but a result of a research strategy decision by Anderson and Bower who felt it was wiser to study memory without taking on the difficulties of language comprehension. On the other hand, it is not clear that the HAM memory system can be used in a natural and efficient semantics-based language processor. The memory structures used by Rumelhart, Lindsey, and Norman (1972), and particularly by Schank (1972), based on the case grammar approach

(Fillmore, 1968), may be easier to use, since they appear to offer a more direct coupling between semantic knowledge and the structure of input sentences.

The most direct extension of HAM's language processing would be to supply a much more elaborate parser, for example, an augmented recursive transition network (Woods, 1970), which would handle a wider range of input sentences. The propositional structures produced by the parser would be matched to memory; if nothing related to the input was found, a good guess is that the input parsing was inappropriate; the system would then return to a previous choice point in the parse and try again. The trouble with this proposal is that the contextual memory information should be used in such a way as to ensure that the best parse will be produced first, rather than merely acting as a check on products blindly produced by the parser (cf. Quillian, 1968; Collins & Quillian, 1972).

It is not at all clear what sort of mechanisms would be required for such a semantics-based processor. The use of a process similar to the retrieval MATCH process is plausible. As the input is brought in, each content word could be looked up in memory to access any old and recent propositions making use of the concepts, as is done to some extent in the current version of HAM (Anderson & Bower, 1973, p. 146). For example, suppose a hippie who had kicked a sailor was under discussion, and the next input sentence were <u>The hippie touched a debutante</u>. The system would immediately identify the hippie with the sailor-kicking hippie already represented in memory. Looking deeper in the network would also produce the information that, at least as characterized by Anderson and Bower, hippies are always touching someone, usually debutantes. This item of information suggests candidate propositions that can be used to guide the parsing of the sentence by anticipating the content before the actual input words arrive. The HAM model has order-of-access assumptions that ensure that the most relevant propositions about a concept will be retrieved first. The identification of one hippie with the other in memory enables the system to integrate the new information with the old, by attaching the new pre...cation to the hippie token node already present in memory.

In conclusion, a realistic model of comprehension will draw very heavily on memory information which will be accessed on a word-by-word basis. That is, each word in the input will be looked up in memory and associated information retrieved for use in the sentence analysis. Hence the nature of the memory information connected to individual words will influence comprehension.

Memory Structure for Some Associative Word Properties

Word Properties and the Memory Network

If the comprehension process relies on pre-existing memory information, then a manipulation of the information in memory that is related to words in the material should produce effects on the processing of the material, both during comprehension and retrieval. One approach to assessing these effects would be to obtain or assume detailed knowledge about the memory structures for particular materials and then compare the processing of materials with structural differences. However, the approach used here is to use certain associative word properties that reasonably imply different memory structures. Sentence materials constructed of words with differences in these properties, but with identical propositional structure, should produce processing changes interpretable in terms of the differences in the memory structure.

The associative word properties manipulated in this study were Imagery (I) value, association value (M, after m, the traditional "meaningfulness" value), and word usage frequency (F). The experimental materials were sentences containing words that varied orthogonally in I, M, and F with high and low levels of each property represented. All content words in a sentence had the same combination of I, M, and F levels.

These particular associative properties have the advantage of the existence of a large set of ratings of nouns on these dimensions collected by Paivio, Yuille, and Madigan (1968). M and F are good candidates for manipulation in that they have reasonable interpretations in terms of the memory network, which will be discussed below, and hence offer an opportunity to extend the network memory models to cover the effects of these word properties.

Imagery value is an extremely important variable that has been a focus of much experimental and theoretical work in the associative tradition (reviewed by Paivio, 1971). Imagery is especially important since its theoretical nature is a very difficult problem. The main problem is the fact that although visual scenes and abstract (non-picturable) information can be coded using exactly the same format of concepts and relations (Minsky, 1963; Pylyshyn, 1973), the use of high-I materials in a memory or comprehension task usually produces much better performance compared to low-I materials. Thus the problem is not really the logical problem of how image information can be represented and processed, but why it can apparently be processed so much better than its formally identical abstract counterpart.

The approach to imagery developed here is based on Pylyshyn's (1973) contention that underlying the use of mental imagery are perceptual descriptions, which are abstracted representations, rather than exact copies, of visual scenes, and are similar to the descriptions of scenes produced as output by the processes of perception. Thus, imagery is based on representations that are like the results of perception, rather than copies of sensory input. These perceptual descriptions are expressible in the same concept-relation format as abstract information, although the concepts and relations may be different. The difference in processing between image and abstract information could be due to several reasons. For example, the image structure could be different in either contents or configuration in such a way that different cognitive processes, specialized for handling perceptual information, can operate on image information but can not be applied to abstract information. Another possibility is that the image information can not be handled any better than the abstract information, but if an image representation can be formed, two separate encodings of the information are available, one in abstract, the other in image form; this redundancy could permit more efficient processing.

The three factors, I, M, and F, only have meaning for an individual word in terms of the memory information available to the subject. For example, it is the structure of the memory network attached to the word doctor that makes it high in all three dimensions. Since the basic processes operating on the network are assumed to be constant, any effects of individual word properties have to be explained in terms of differences in the memory structure for the words. Thus, the task of this section is to produce an interpretation of I, M, and F in terms of the structure of the memory network connected to a word.

Figure 2 shows a hypothetical memory network for a single word, which has an acoustic representation W_a and a visual representation W_v. These representations, which are descriptions of the word as spoken and written, are linked to a node W_1 representing the word as a lexical item. This lexical node is tied by an idea link to a type node C_1 representing the concept referenced by the word. Note that other words may refer to the same concept, and the words may refer to more than one concept. The concept node C_1 is referenced by propositions (abbreviated by the circles labelled P_1, P_2, etc.) via ε^{-1} links. Other concepts are also used in the propositions and have connections by idea^{-1} links to lexical item nodes. Some concept nodes are connected via an appearance link to a perceptual description (shown as a picture frame) of the appearance of the object corresponding to the concept. It is this description that makes possible the perceptual recognition of an object, and underlies the process of forming a mental image of the object. Notice that not all concepts have perceptual descriptions attached to them. Now the manifestations of I, M, and F in the memory network can be discussed.

Imagery

Obtaining ratings of imagery value (Paivio, et al., 1968) involves presenting the subject with a word and asking him to rate how easy it is to form an image in response to the word. The actual process underlying the formation of an image from a sentence has not been considered in the imagery literature. As developed here, forming an image from a single word consists of finding the word node from the presented word, following an idea link to a concept node, following an appearance link to the

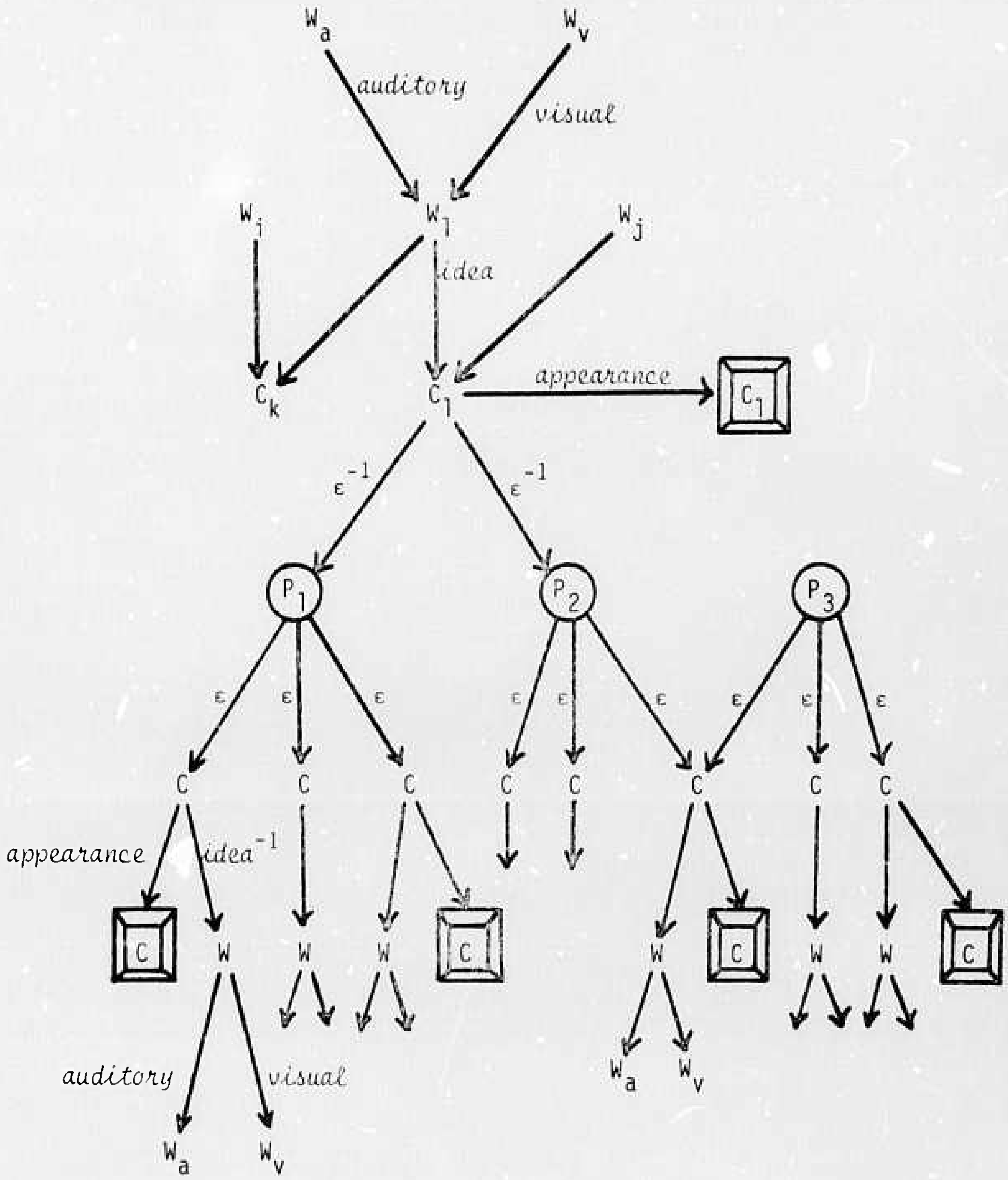

Figure 2. The memory network connected to a single word. See text for explanation.

perceptual description of the conceptualized object, and then performing a fairly mysterious final step resulting in conscious awareness of an image based on the perceptual description. The imagery rating would depend on the difficulty of finding a suitable perceptual description and using it in the last step. Although this last step is mysterious, it does make sense to consider imagery in terms of what the step must logically have available as input; it is assumed that the determinant of the ease of image formation lies in the perceptual description supplied to the mysterious last step.

If the concept corresponding to the stimulus word corresponds to some concrete object, the perceptual description of the object would be directly available. However, an abstract concept, such as "justice," is not connected to any particular perceptual description; forming an image in such a case requires constructing a usable perceptual description from other information in memory that is related to the concept, by bringing in other concepts with perceptual descriptions. For example, the subject might remember that justice is often dispensed in a courtroom of characteristic appearance, by a judge who is usually old, male, and wearing a black robe. Perceptual descriptions suitable for image formation are attached to many of these new concepts. It should be emphasized that in no sense does this collection of imagible tidbits constitute the <u>meaning</u> of the abstract word; despite the many eloquent statements of this point, dating back to the Wurzburg School (Humphrey, 1963, Ch. 2, 3, 4), many investigators of imagery still seem to slip into the logically inadequate meaning-as-image position (cf. Anderson & Bower, 1973, pp. 449 ff; Pylyshyn, 1973).

Once suitable propositional information related to an abstract concept has been collected, the subject faces a task that is similar for both abstract concepts and sentences: the formation of a perceptual description based on a proposition. The subject must construct perceptual descriptions based on relations, such as wear in the proposition judges wear robes, and combine all the description information together into one large perceptual description, which is the input to the mysterious last step. Given this rather complicated search and construction process, it is no surprise that image formation times for low-I material are so much longer than for high-I (Paivio, 1971, p. 442). The reason why low-I materials have their low imagery ratings is that a considerable amount of additional memory search and cognitive work is required to obtain the propositions and perceptual descriptions useful for an image.

M-value

The method under consideration for obtaining ratings of association value is to present the stimulus word and then have the subject write down free associates to the word. The average number of associates produced in a period of time, such as the 30 seconds used by Paivio et al. (1968), is the M-value of the word. In memory network terms, the number of associates depends on how many words in the network a subject can locate starting from the stimulus word. One strategy would be to produce the words immediately related as synonyms, antonyms, homonyms, and so on, but most associates appear to be propositionally related to the stimulus word (cf. the paradigmatic-syntagmatic distinction discussed in Cramer, 1968, p. 68 ff). Referring to Figure 2, the subject would go from the concept node for the stimulus word into some proposition, select some other concept in the proposition, and produce a word for it. Then he could return to the same proposition for another concept, or backtrack to the original concept and choose another proposition, or even return to the stimulus word and choose a fresh conceptual encoding. Another possibility is that he could start from one of the produced concepts and enter a new set of propositions related to that concept, but not directly connected to the stimulus concept. This possibility is shown by the chained propositions in the bottom right-hand corner of the network in Figure 2.

An obstacle to understanding M-ratings is that they are rather low, considering that a subject could just produce every word he came across in his path through the network, which would result in an incredible number of associates. Apparently some kind of selection process is at work, which might be revealed by a detailed study of the propositional relations between the stimulus word and the associates.

A second problem is that I and M are highly correlated (.72, Paivio et al., 1968) for reasons not at all clear. One idea is that if the subject can form a good image to the stimulus word, he can just name other objects in the image. Unfortunately, this does not explain where the additional objects came from, or why there would be more of them in the case of an image resulting from a high-I word than for an image resulting from the greater memory search for a low-I word. Another explanation for the correlation between I and M is that most high-I words correspond to concepts that people have a lot of direct and related knowledge about, such as doctor, and thus there are many accessible concepts to be considered for associates. This explanation attributes the correlation to the nature of the knowledge people have about the world, rather than any fundamental properties of the memory system itself. The fact that the correlation is not perfect provides some support for the position that associates are not a simple side product of forming an image.

Frequency

Word frequency here refers to the absolute frequency of usage of a word as listed by the Thorndike-Lorge count, or the more recent Kučera and Francis (1967) count. It is important to distinguish usage frequency from word conjoint frequency, which has a rather different role (cf. Anderson & Bower, 1973, p. 380).

Usage frequency could act in two general ways, only one of which has to do with memory propositions. The initial access from perceptual description to word node could be executed faster for frequent words; Forster and Chambers (1973) recently found that common words could be named faster than rare words. The other aspect of frequency is that high frequency words usually correspond to high frequency concepts; the more heavily used a concept is, the more propositions it will be referenced by. Hence high frequency words will connect to concept nodes that have many ε^{-1} links leaving them.

It is important to note the difference between M and F in terms of the memory network. The propositional manifestation of F is limited to the number of ε^{-1} links leaving a concept node, while the network structure relevant to M includes the propositions and their concepts, together with the indirectly connected concepts and propositions as well. Thus although there is some overlap, M and F reflect different properties of the memory network; this conclusion corresponds to the moderate correlation of .33 between M and F reported by Paivio et al. (1968).

A Descriptive Model of Verification Performance

Surface and Meaning Processing

As mentioned before, the experimental task consisted of an acquisition task, in which subjects read a randomly ordered series of sentences that described a set of events, followed by a verification task consisting of true-false questions about the events.

Many studies (reviewed by Anderson & Bower, 1973, pp. 222 ff) have shown that subjects can remember two kinds of information about sentence material: the exact surface form of the sentences encountered in the input, and the meaning, or gist, of the input; the extent to which information about exact sentence form or gist is stored depends on the task demands and the nature of the material. The experiment reported in this thesis compared surface and meaning memory by including two kinds of true question sentences that differed in the importance of surface memory in arriving at the correct answer. _Explicit_ questions were exact repetitions of acquisition sentences, and thereby were explicitly true. _Implicit_ questions were novel, not actually seen during acquisition, but were implicitly true, since they consisted of propositions that had appeared in several separate acquisition sentences which would be integrated during acquisition (Bransford & Franks, 1971). Explicit questions could be answered correctly if the subject recognized either the meaning or the surface form of the sentence; implicit questions could be answered only on the basis of the meaning of the acquisition material. Including these two question types made it possible to determine how the various conditions in the experiment affected both the surface and the meaning storage and subsequent recognition.

One approach to the problem of sorting out the effects of the experimental manipulations on surface and meaning processes would be in the form of detailed and closely reasoned arguments directly from the data to the hypothesized mechanisms. The approach that was used here was to devise a formal descriptive model of the verification task and obtain estimates of the parameters governing surface and meaning subprocesses in the model. Changes in these parameters resulting from the manipulations can then be interpreted directly as changes with theoretical significance.

An added advantage of using the descriptive model was that it assisted interpretation of the many joint changes in speed and accuracy in the verification data. As Pachella (1973) has pointed out, certain speed-accuracy changes are ambiguous, and often due to changes in a subject's speed and accuracy criteria for performance, rather than more profound changes in the interesting processing mechanisms. In many cases, the descriptive model decomposed a single ambiguous speed-accuracy shift into nonambiguous changes in each of the surface and the meaning-based subprocesses.

The Model

The model consists of four subprocesses: an input process that reads in the question sentence, a response execution process, and two comparison or matching processes. A surface-match process compares the surface form of the question sentence with memory information about the surface form of the acquisition sentences. A surface match success implies a true response, whereas a surface match failure implies nothing about the correct response. The other comparison process is a meaning match that derives the meaning of the question sentence and compares it to memory information about the meaning of the acquisition material, using a search mechanism similar to the HAM MATCH process discussed above. A successful meaning match implies a true response, a failure implies a false response.

In the data, the explicit questions were generally answered faster and more accurately than implicit questions. The model assumes that explicit questions have this advantage since they can be judged as true on the basis of either a meaning or surface match, whereas implicit questions could have only a successful meaning match. Thus, explicit questions are more accurate since there are two independent methods of achieving a correct answer. Explicit questions are answered faster because either the surface match is attempted first, and followed only if necessary by the meaning match (the _serial_ match model), or the surface and meaning matches are initiated simultaneously, but usually the surface match succeeds before the meaning match on explicit questions (the _parallel_ match model). Both models will be considered, since each has important advantages.

Figure 3 and Figure 4 show the flowcharts for the two models. In both the serial and parallel models, the first process consists of reading in the question sentence; the last process is response execution, which is assumed to take equal time for true and false responses. In the serial model, first the surface match (or S-match) is executed, followed by the meaning match (M-match) if necessary. A true response is made immediately if the S-match succeeds; if it fails, the M-match is attempted and leads to a true response upon succeeding and a false response if it fails. In the parallel model, both matches are started at the same time; the response execution produces a true response as

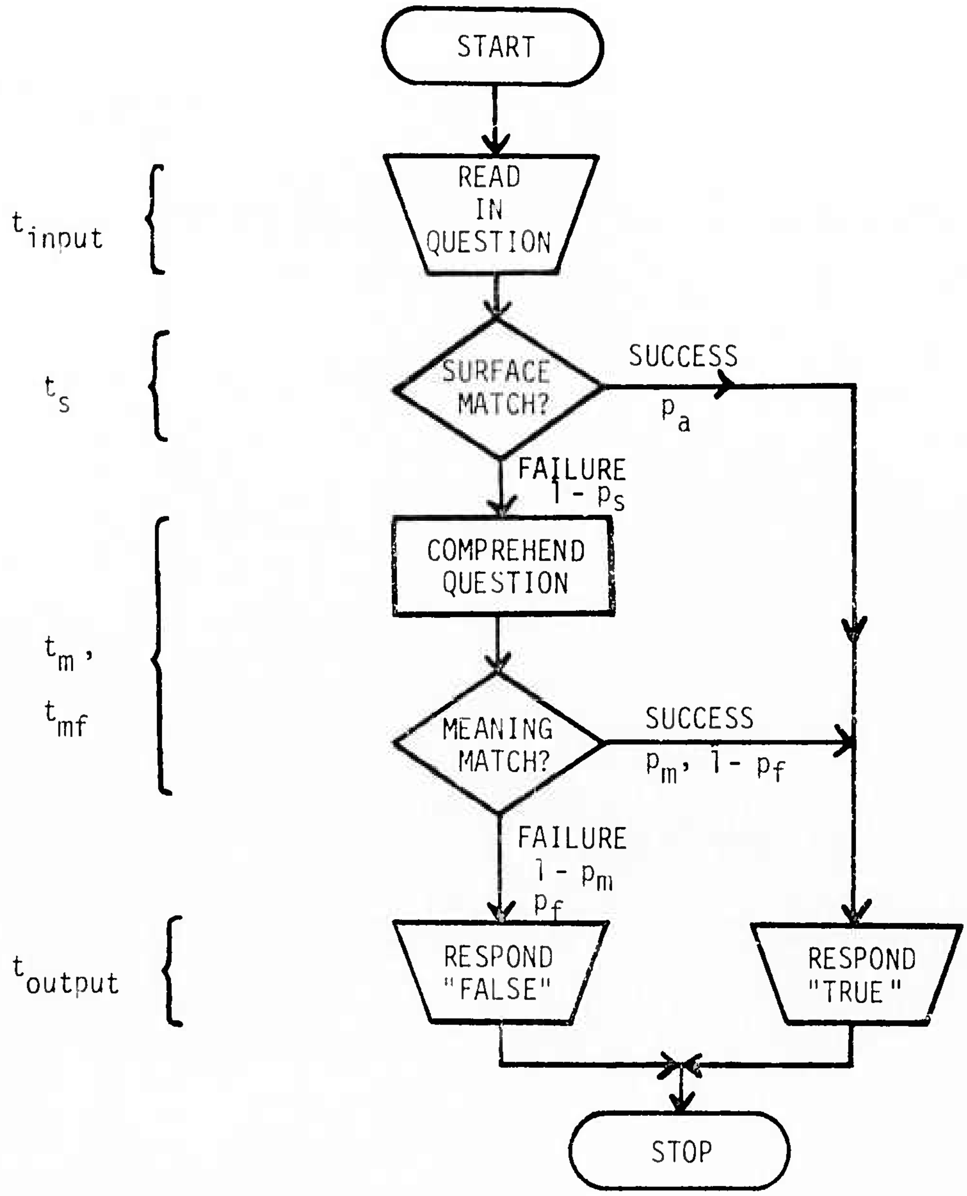

Figure 3. Flowchart for the serial match model of the verification task. The paths coming out of each match are labelled with the probability of that match outcome. Also shown are the durations used in the model for each stage.

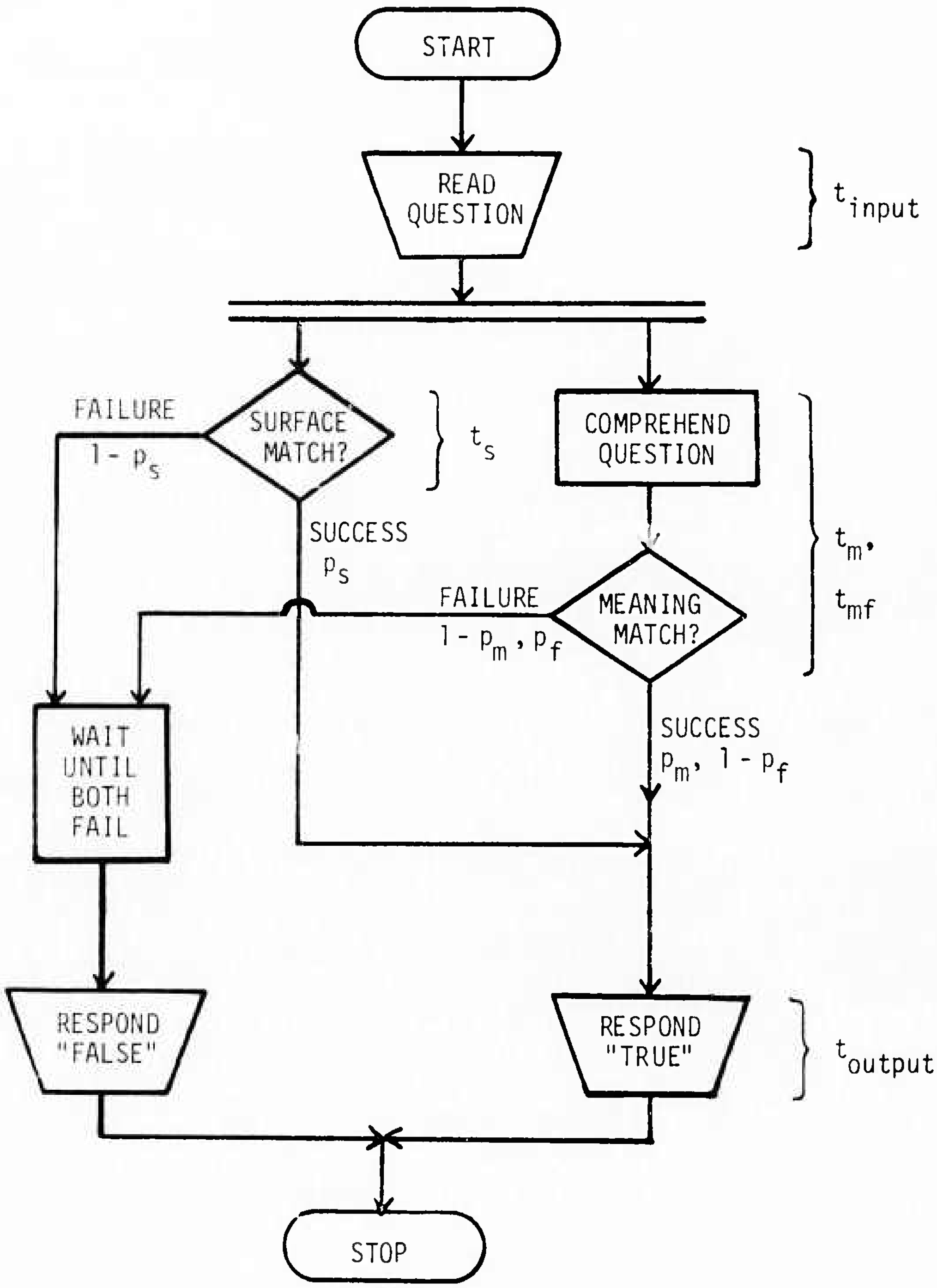

Figure 4. Flowchart for the parallel match model of the verification task. The parallel lines near the top symbolize the simultaneous initiation of the surface match and the question comprehension processes.

soon as either match succeeds, a false response as soon as both matches terminate in failure.

Many different characterizations of the S- and M-match processes could be made; the specific assumptions which follow seem to be the simplest and most useful in providing a worthwhile description of the data:

1. The S-match can succeed only on explicit questions; it will always fail on implicit and false questions. Symbolically,

$$P(\text{S-match success} \mid \text{explicit}) = p_s; \tag{1}$$

$$P(\text{S-match failure} \mid \text{explicit}) = 1 - p_s; \tag{2}$$

$$P(\text{S-match success} \mid \text{implicit or false}) = 0. \tag{3}$$

2. The probability that the M-match will succeed is the same for implicit and explicit questions. This means that the integration process operating during acquisition is no longer at work during verification; the meaning representation is not different for explicit and implicit sentences.

$$P(\text{M-match success} \mid \text{implicit}) = P(\text{M-match success} \mid \text{explicit}) = p_m; \tag{4}$$

$$P(\text{M-match failure} \mid \text{implicit or explicit}) = 1 - p_m. \tag{5}$$

3. An error on a false question can result only from a false alarm in the M-match process:

$$P(\text{error} \mid \text{false}) = P(\text{M-match success} \mid \text{false}) = p_f. \tag{6}$$

These assumptions lead to the following statements about the proportion of errors on the different question types:

$$P(\text{correct} \mid \text{implicit}) = P(\text{M-match success} \mid \text{implicit})$$

$$= p_m; \tag{7}$$

$$P(\text{correct} \mid \text{false}) = 1 - P(\text{M-match success} \mid \text{false})$$

$$= 1 - p_f; \tag{8}$$

$$P(\text{correct} \mid \text{explicit}) = p_m + p_s - p_m p_s. \tag{9}$$

The expression in Equation 9 above is just the total probability for the independent but nonexclusive events of either or both the S-match and M-match succeeding. An estimate of the accuracy of the S-match, $\hat{p}_s$, can be derived as follows:

$$p_s = \frac{(p_m + p_s - p_m p_s) - p_m}{1 - p_m} \tag{10}$$

$$\hat{p}_s = \frac{P(\text{correct} \mid \text{explicit}) - P(\text{correct} \mid \text{implicit})}{1 - P(\text{correct} \mid \text{implicit})} . \tag{11}$$

As shown in Equation 7 and Equation 8, p_m and p_f can be estimated by the proportion of correct answers on implicit and false questions.

The boxes in the flowcharts of Figures 3 and 4 each have a time associated with them; t_m and t_{mf} represent the M-match time for true and false questions, respectively, and t_s represents the S-match time. Let $T = t_{input} + t_{output}$ represent the remainder of the time which is occupied by the read-in and response boses.

These fixed values for the process durations represent the expected or mean times required by each process. The expected times to answer the three question types differ for the serial and parallel models. The serial model yields the following expected latencies (conditional on correct answers):

$$E(t \mid \text{implicit}) = T + t_s + t_m; \tag{12}$$

$$E(t \mid \text{explicit}) = T + t_s + (1 - p_s)t_m; \tag{13}$$

$$E(t \mid \text{false}) = T + t_s + t_{mf}. \tag{14}$$

The data may be used to obtain estimates of these times:

$$\hat{t}_m = \frac{t_{implicit} - t_{explicit}}{\hat{p}_s}; \tag{15}$$

$$\hat{T} + \hat{t}_s = t_{implicit} - \hat{t}_m; \tag{16}$$

$$\hat{t}_{mf} = t_{false} - (\hat{T} + \hat{t}_s). \tag{17}$$

The corresponding latency expressions in the parallel model differ somewhat from the serial. On implicit questions, only the M-match can succeed; thus the latency depends only on the M-match time (considering correct answers only):

$$E(t \mid \text{implicit}) = T + t_m. \tag{18}$$

Explicit questions can be answered by either the M- or S-match; usually the S-match is faster. Thus,

$$E(t \mid \text{explicit}) = T + p_s t_s + (1 - p_s)t_m. \tag{19}$$

As in implicit questions, false questions depend on the M-match process:

$$E(t \mid \text{false}) = T + t_{mf}. \tag{20}$$

The estimation formulas in Equations 15, 16, and 17 for the serial model can also be applied to the parallel, with different results:

$$\frac{t_{implicit} - t_{explicit}}{\hat{p}_s} = \hat{t}_m - \hat{t}_s. \tag{21}$$

Thus the serial model yields a direct estimate of t_m whereas the parallel yields an estimate of the duration by which the M-match exceeds the

S-match. However, subtracting the same estimation function in Equation 16 from $t_{explicit}$ yields the same estimate of $T + t_s$:

$$t_{implicit} - (\hat{t}_m - \hat{t}_s) = \hat{T} + \hat{t}_s . \qquad (22)$$

The estimate of t_{mf} corresponding to that of t_m can be obtained in the same way as in Equation 17:

$$t_{false} - (\hat{T} + \hat{t}_s) = \hat{t}_{mf} - \hat{t}_s . \qquad (23)$$

Variability of the Parameter Estimates

The repeated operations on the data involved in some of the estimates would be expected to lead to an accumulation of error. Particularly serious is the case of the estimate for p_s which enters into all the other estimates. There are two approaches to this problem: (a) devise the appropriate techniques for exact specification of confidence intervals and significance tests on the parameter values; (b) argue for the theoretical significance of changes in parameter values by referring to the appropriate statistically significant effects in the data. The second strategy will be followed here, because close correspondences between the estimates and the statistical analysis can be seen in most cases. The value of the descriptive model is that these effects are more easily understood with the model as an interpretive aid. Furthermore, the function for estimating p_s is extremely poorly behaved; performing the more exact techniques would be a long and complex process.

Serial versus Parallel Matching

The serial version of the descriptive model enjoys the important advantages that it is simpler than the parallel, and the M-match time estimate can be taken as an absolute time value. However, some problems arise in the detailed breakdown of the data. In some of the experimental conditions, $t_{explicit}$ turns out to be greater than, or very close to, $t_{implicit}$, which produces values of $\hat{t}_m$ ($\hat{t}_m$ - $\hat{t}_s$ for the parallel model) which are negative or very small. In the serial model, such values have to be the result of sampling error, since negative values of t_m are impossible, and very small values of t_m are implausible. However, that these are a result of sampling error seems unlikely in view of the fact that roughly 100 separate latency observations underlie each estimate.

The more flexible parallel version allows these very small or negative $\hat{t}_m$ - $\hat{t}_s$ estimates as indications that the M-match can operate as fast as or faster than the S-match under certain conditions. In fact, these small and negative estimates are not randomly distributed through the conditions, but all appear in the high-I conditions. A related aspect of the results is that false high-I questions have negative $\hat{t}_{mf}$ - $\hat{t}_s$ estimates even at the level of main effects. This could be explained as resulting from subjects being able to respond false very quickly by spotting a contradiction in the first few words of the sentence, without waiting for even the read-in process to complete. This suggests that evaluation of the question can proceed while it is being read and comprehended, a further use of parallel processing that the model does not attempt to capture.

The problem with the parallel version is that the M-match time estimate $\hat{t}_m$ - $\hat{t}_s$ is confused with the S-match duration t_s; the only way to get any interpretive use out of the parallel model is to assume that t_s is a constant and does not vary as a function of any of the experimental conditions. Once this assumption is adopted, changes in the serial and parallel M-match time estimates have identical implications, and changes in surface processing time, $T + t_s$, are all due to changes in T. Since the descriptive model is intended to be only an interpretive aid, neither of the two versions of the model will be rejected; as is clear from the above comments about the processing of high-I false questions, both models fail to capture some important properties of the actual processing being performed.

The M-Match Process

The emphasis will be on the most interesting process in the descriptive model, the M-match process. Overall, the M-match process consists of a comprehension process that derives the propositional structure of the question, followed by a HAM MATCH process using the question proposition as a probe, and ending with a decision process that chooses a response based on the results of the MATCH stage. It should be noted that the descriptive model does not separate the times required by these M-match subprocesses.

The HAM MATCH process. The detailed description of the MATCH process that follows is based on the Anderson and Bower (1973) model, but with some modifications.

The MATCH process operates in both serial and parallel. Referring to Figure 5, the probe proposition has token nodes which are connected to concept type nodes in permanent memory. Some of the propositions in memory also use the same concept type nodes. The MATCH process consists of a set of independent parallel searches which work from each of the concept nodes referenced by the probe; the searches do not interact or exchange information. Each search starts from the concept node and attempts to find in memory a propositional structure that matches the probe, that is, has the same configuration of links and nodes connected

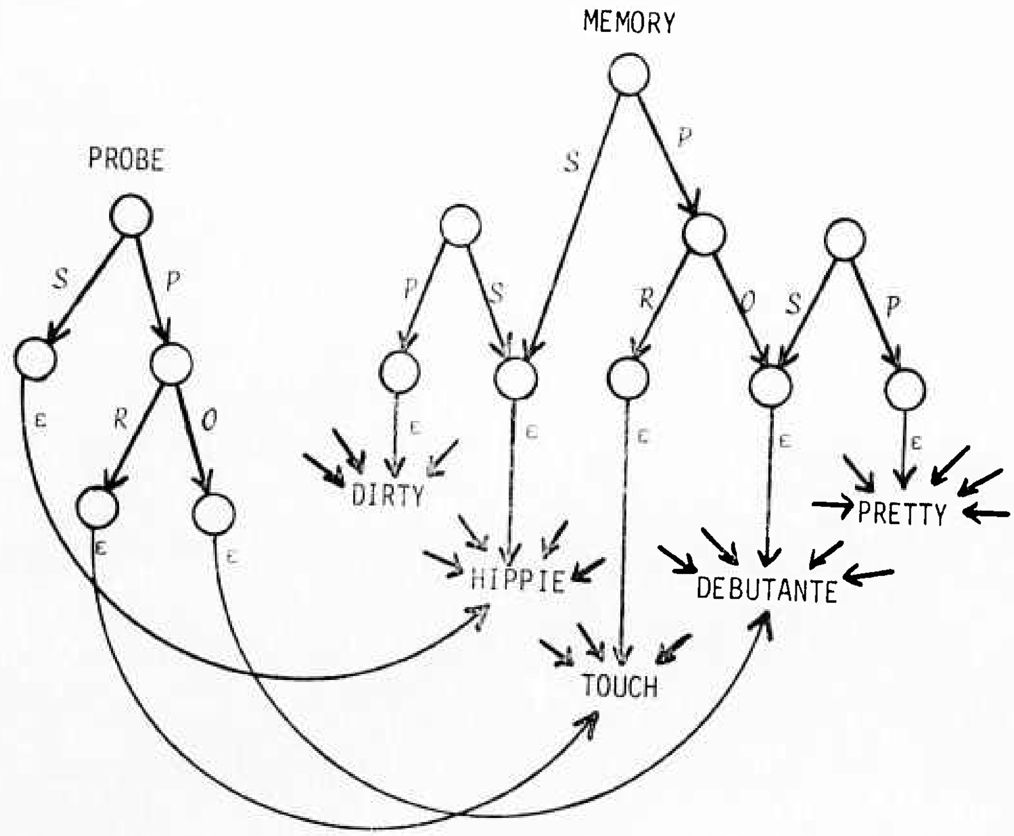

Figure 5. Illustration of the HAM MATCH process. Independent searches start from each of the concept type nodes and attempt to find a structure in memory matching the probe. Note the multiple ε links attached to each concept node.

to the same concepts. The search starts by finding matching links leaving the concept node, which are ε^{-1} links. Whenever there are multiple memory links of identical type leaving a node, each link and the structure connected to each link are examined one at a time by the search process; thus, at this level, the MATCH process is serial. Once the search has found matching links, it tentatively equates the probe and memory nodes at the end of the links, and then looks for another matching link leaving this new node. If the search comes to a concept node, that node is tested to see if it is the same node as the one referenced by the corresponding token node in the probe; if not, that part of the memory structure is not a correct match. The search backs up to the previous matching node and tries again with a different link to some other node.

If any one of the independent searches succeeds in finding a memory proposition that exactly matches the probe, the MATCH process terminates immediately with a complete match. Another possibility is that the process will terminate with a partial match, in which only part of the probe is found in memory; whether a partial match is accepted in lieu of a complete match depends on the decision strategy and can be a source of errors. In the present results, subjects made substantial numbers of errors on false questions, implying that they used partial match information in evaluating questions, rather than adhering to a strict complete match criterion for responding true. A final possibility is an exhaustive failure, which results when neither a complete match nor a satisfactory partial match is found after an exhaustive search of relevant memory structure. This could result from either the memory not containing the probe proposition, or the searches failing to examine all of the information actually present.

<u>The role of ε^{-1} links</u>. The chief source of difficulty for the MATCH process is the presence of multiple links that have to examined serially. A node that has a large number of links of identical type leaving it will be said to have a high <u>fan-out</u>. Usually the memory concept nodes have a high fan-out in ε^{-1} links, due to the many propositions known about each concept. The nodes in a propositional structure also may have some fan-out, for instance, if multiple predications are made of the same subject. Anderson and Bower (1973, p. 361 ff) report experiments in which the fan-out of such nodes was manipulated to obtain results consistent with the described MATCH process.

Since the ε fan-out of concept nodes is so large, some mechanism is necessary to ensure that important links will not be lost in the thicket of connections encoding all of the subject's knowledge. Another consideration is that the system has to be capable of forgetting, despite the all-or-none character of the links. The Anderson and Bower (1973, p. 238, p. 380) approach to these issues is to assume that:

1. The links leaving a node are ordered in a list that determines the order in which the links are examined in a search.

2. The list is searched to a probabilistically determined depth; links below this limit are not examined.

3. The link ordering is determined by recency and frequency; the more recently installed, or more frequently used, a link is, the higher it is on the search list.

Under these assumptions, interference and forgetting are due to the links being pushed further down the list by newly installed links. Furthermore, the system is able to distinguish old from new information by the list position. However, the degree of fan-out does not play a very direct role, since the most recent connections are always first on the list, and the depth to which the list is searched does not depend on the fan-out; hence the degree of fan-out would not affect the availability of recently installed links. This situation can be remedied by assuming that a random sample, or search set, of some fixed size is selected from the links leaving a node; the probability that a particular link is included in the search set is a function of the fan-out and search set size, as well as the recency and usage frequency of the link. This probability increases with the search set size, and decreases with the degree of fan-out.

Speed and accuracy of the M-match. One of the ways in which the probability of success of the MATCH process can be affected is that if certain links are missing in the propositional structure in memory, a complete match to the probe is not possible. Hence factors affecting whether memory links were installed during acquisition and remain until verification will have an effect. Another way is that anything determing the search set will also affect accuracy; essentially, this has to do with the retrievability of the material. For example, if the ε^{-1} link connecting a concept to the target proposition is not included in the search set for the search working out of that concept, the search is bound to fail. Of course, a search from one of the other concepts could succeed.

The speed with which the M-match terminates depends on several things:

1. Since there are independent searches from each concept, the more concepts available in the probe, the faster will a match be found.

However, in this experiment, more probe concepts always entailed more structure to be matched. Thus the relationship between match time and the number of probe concepts would be quite intricate and will not be discussed further.

2. The order in which the links in the search sets are examined will affect match time, since if the critical links are high in the search order, the match will be found after fewer links are examined.

3. In some tasks, special strategies may be possible that permit highly efficient use of relatively small amounts of retrieved information, thus leading to short match times.

Answering false questions. The most important special strategy that subjects probably used in this experiment is in answering false questions by the use of partial match information that contradicts the probe proposition. The details of this mechanism are not clear, but it can be sketched out.

In the experiment the materials have the property that a predicate is attached to only one subject. For example, the materials contained the statements The admiral was crusty and The prosecutor was shrewd. Crusty was applied only to the admiral, shrewd only to the prosecutor. Suppose that the probe is the false question sentence The admiral was shrewd. A match attempt would discover the two propositions mentioned above. The fact that the probe concepts are found in two disjoint memory propositions implies that the probe is false. If only one of the memory propositions were discovered, no such strong conclusion could be drawn, and an exhaustive search would be required to conclude that the probe was false.

A problem is that the subject may know of some shrewd admiral out of British naval history. How does he avoid confusing the pre-existing information with the experimental information? One way would be the link ordering; presumably the ε^{-1} links defining the historical admiral would be far down in the search order relative to the experimental admiral. A more powerful method is the context information that the HAM model carries attached to each proposition, which would indicate that the crusty admiral was in the experimental context, and the shrewd admiral was in the historical context. The full probe proposition would also carry an experiment context label that would be part of the structure to be matched and comprise another set of concepts to be searched frc . This scheme would result in only the experimental propositions being matched to the probe.

It would be expected that a contradiction could be unearthed fairly quickly since only part of the complete probe need be contradicted. Some of the false questions in the experiment are quite long and provide many opportunities for discovering contradictions on separate subject-predicate structures, any one of which is adequate for a false response. In contrast, true responses require a fairly thorough search for a complete match, which could take a long time for a complex question; exhaustive false responses also require an extensive memory search.

Another way to look at strategies for answering false questions is to consider the avoidance of false positives, rather than the discovery of falsehood. If a strong partial match is discovered, and no contradictions found, it is a good guess that the question is true and that only a slight amount of missing information precluded a complete match. Thus false question accuracy will depend jointly on the probability of strong partial matches and the probability of discovering contradictions. If neither of these events occur, the subject would then make an exhaustive memory search; his response choice after a complete search failure may not always be false; he might accept a hodgepodge of several weak partial matches as evidence for truth, saying, in effect, "Something like the question was mentioned during acquisition -- I guess it must be true." The conclusion is that even in terms of a highly specific model like HAM, the process of deciding falsehood is extremely complicated (cf. Collins & Quillian, 1969) and demands a comparison between more detailed data and an elaborated simulation model.

Summary

The semantic approach to memory has stressed the problem of arriving at the meaning of the sentence by using information about the semantic and logical relations of concepts denoted by individual words. The associative model has developed out of considerable research on the associative interconnections between words. The problem is to determine whether associative word properties can be included in the more general semantic network memory model and whether they have effects on the storage and retrieval of sentence information. To this end, three well-studied associative word properties, I, M, and F, were selected for study. The I level of sentence words was interpreted in network terms as governing the ease of formation of a complete perceptual description of a visual scene corresponding to the conceptual meaning of the sentence. M-value of sentence words affects the number of pre-existing memory propositions that can be used as parts of the new sentence proposition. F similarly affects the number of reusable memory propositions, but is restricted to governing the degree of fan-out from concept nodes, and also affects basic word access time. These properties will be investigated experimentally by composing sentence materials orthogonally varying in these properties and having subjects first read the sentences and then answer true-false questions about them.

The time subjects required to read the sentences will provide a direct, but gross, measure of comprehension difficulty. The same applies to some extent to the verification latency data, since the questions have to be comprehended before they can be answered, and the questions use the same words and sentences as the acquisition materials. However, the relation of acquisition and verification times to each other and to verification accuracy is not obvious. For example, it is possible that different materials can have very different study times, but little or no difference in verification performance, as in the case of Kintsch and Monk (1972), who manipulated the syntactic complexity of the acquisition material.

Study time reflects the time taken to read in the sentence, obtain relevant information from memory, and construct the new propositional structure in memory. Verification time includes the read-in time, time for comprehending the question sentence, and time for comparisons of the question with relevant information in memory to decide upon a response. Factors which influence study time could also influence verification time, if they affect the comprehension and memory access subprocesses. However, the tasks differ in that no storage is required in verification, and no comparison to determine truth value is needed in acquisition. Hence properties of the materials could affect study and verification times differently.

A descriptive model of the verification task was described; the model is descriptive in that its primary function is to assist interpretation of the data in terms of the more basic theoretical notions, rather than generate predictions to be tested against the data. The model is based on the two types of stored information, surface and meaning information, which are distinguished by the different levels of performance on explicit (exact repetition) and implicit (novel) true question sentences. The estimation functions generated by the model provide estimates of the speed and accuracy of the surface processing and the more interesting meaning recognition process, which consists of a HAM-like comparison process whose speed and accuracy are mainly determined by the degree of fan-out from the concept nodes appearing in the question. The parameters governing the meaning recognition process will be used to explain the effects of the word properties in terms of differences in the memory network and memory processes operating on the network.

CHAPTER II

METHOD

Overview of the Experiment

The experimental task comprised three subtasks: an acquisition task, followed by a distractor task, and concluded by a verification task. The acquisition task was modeled after the Bransford and Franks (1971) procedure. Sentences that fit together to describe a set of four distinct complete ideas, or events, were presented one at a time in a random order; subjects were instructed to assemble the complete ideas while reading the sentences. The time taken by the subject to study each sentence was recorded. After the distractor task, which consisted of arithmetic problems, came the verification task, which was a series of true-false questions about the events. The response accuracy and latency for each question were recorded. The questions were in the form of sentences that generally resembled the acquisition sentences in length and structure. The acquisition and verification sentences varied in size, or complexity, as gauged by the number of underlying propositions. True questions were either explicit (exact repetitions) or implicit (novel); false questions were sentences that contradicted the acquisition description of the events and used the same content words as the acquisition sentences.

A set of four complete ideas was composed for each combination of high and low imagery (I), associative value (M), and frequency (F) of the words in the descriptive sentences. Each of the resulting eight sets of complete ideas was tested in a separate repetition, or block, of the experiment task, so that the subject was dealing with materials with only one combination of word properties at a time.

A Stress Manipulation

A stress manipulation, consisting of a limitation on acquisition reading time, was included in the study and is introduced here since it is not directly related to the word property manipulations. Half of the subjects were allowed to read the acquisition sentences at their own rate (the self-paced group); the other half were forced to read the sentences at a fast rate (the force-paced condition). Note that no speed stress was imposed on the verification task.

The limitation of the time available for reading the acquisition sentences would be expected to lower the accuracy of subsequent verification. However, neither the effect of limited reading time on verification speed, nor its interactions with the word properties, is obvious. For one thing, since forming an image for a sentence takes longer than simply comprehending the sentence (Paivio, 1971, p. 442), the limitation of reading time may force subjects to adopt a different storage strategy with regard to imagery. The stress-induced strategy could diminish the natural superiority of high-I material, or it could alter the balance between surface and meaning information storage for high- and low-I sentences.

A second issue concerns the way in which the limitation of available reading time affects the stored form of the information. The results of many learning experiments show that longer study time improves the storage and retrievability of the information; this is another problem that the semantic memory models have not delved into. The patterns of verification speed and accuracy produced by the descriptive model for the self-paced and force-paced conditions should yield some insight into the way in which available acquisition time affects the memory structure for the material.

Materials

The complete ideas. The complete ideas, which were the events used to produce the acquisition and verification sentences, were sentences composed under several constraints. The manipulation of I, M, and F was achieved by controlling the properties of the nouns, using the list of ratings by Paivio et al. (1968). The categories of high and low values of I and M were obtained by dividing the list of nouns at the reported mean values. The Thorndike-Lorge frequencies in the list were replaced by the frequencies in the Kučera and Francis (1967) listing; high frequency was defined as a frequency greater than 49, low frequency as less than 21; these frequencies correspond roughly to the number of occurrences per million. Lists were compiled of the nouns in each cell of the 2 × 2 × 2 classification.

Sentences of the following form were composed: (adjective)(subject noun)(modifying phrase)(transitive active verb, past tense)(adjective)(object noun)(modifying phrase), abbreviated as $A_1N_1M_1V\ A_2N_2M_2$. Modifying phrases had various forms that included another noun, such as of the speech, who were filled with pep, or urging cooperation.

All nouns in each complete idea, including those in modifying phrases, had to appear in the same noun category; for example, all nouns in a high-I, low-M, low-F complete idea appeared in the high-I, low-M, low-F cell of the noun classification. Since the appropriate ratings for verbs and adjectives have not been obtained, these words could not be controlled on the I and M dimensions, but were controlled on F. An effort was made to ensure that putatively high-I complete ideas were in fact imagible and vice versa for low-I ideas; according to Paivio (1971, p.453), imagery ratings of sentences correspond to the ratings of the words; of course, nonsense sentences are not likely to be imagible. Other requirements were that the sentences had to be nonanomalous, natural, and fairly self-contextualizing, that is, not requiring an external context for fairly unambiguous and direct meaning (cf. Bransford & Johnson, 1972). An important contraint was that all sentences had to refer to events; this was required in order to avoid the natural tendency to compose maxim-like sentences with abstract words, such as The originator of a hypothesis should manifest prodigious flexibility.

For each cell of the I, M, F classification, four complete ideas were composed that were independent, not having any content words in common. An additional set of four ideas, two imagible and two abstract, were composed for use as a practice set, and were modified from the materials used in Bransford and Franks (1971) and Franks and Bransford (1972). A list of all the complete ideas appears in Appendix A; Table 1 gives an example complete idea for each I, M, F combination.

Since the nouns in the Paivio et al. (1968) norms were categorized as high or low I and M by splitting the distributions at the means, it is important to know the extent to which the generated sentences actually met the goal of orthogonally varying the properties over a reasonable range. The average characteristics of the complete idea nouns in each cell of the I, M, F classification were computed. Table 2 shows the overall means for high and low levels of each of the three properties; Table 3 shows the means for high and low levels of each of the properties in all eight cells. I and F had a large spread between the two levels (very high frequency words were tabulated as being either 50 or 100; cf.

TABLE 1

EXAMPLE COMPLETE IDEAS

Low I, Low M, Low F:	The novel hypothesis that elucidated antitoxins corrected antique misconceptions about immunization.
Low I, Low M, High F:	The extensive effort by the management improved the low quality of the product.
Low I, High M, Low F:	The chronic malady of rheumatism afflicted the respected theologian who received a blessing.
Low I, High M, High F:	The increased cost of welfare demonstrated the economic necessity for new legislation.
High I, Low M, Low F:	The squat keg buried in the rubble tripped the handcuffed captive who was weakened by fatigue.
High I, Low M, High F:	The radical speaker who shouted during the meeting demanded the violent destruction of military headquarters.
High I, High M, Low F:	The shrewd prosecutor who was chewing a cigar scrutinized the unidentified corpse which was in the morgue.
High I, High M, High F:	The drunk student who was playing the piano drank the red wine which was in a bottle.

TABLE 2

AVERAGE PROPERTY VALUES OF THE COMPLETE IDEAS

Property	Level of Property	
	Low	High
I	3.8	6.0
M	5.2	6.5
F	7	68

Paivio et al., 1968). M does not move over a very large range, since it is very hard to move M without affecting I. Reasonable orthogonality was achieved, since as can be seen in Table 3, the level of each property is fairly unaffected by a change in levels of the other properties. For example, in the first two cells, the I and M values remain constant while

TABLE 3

AVERAGE PROPERTY VALUES FOR EACH COMBINATION OF HIGH (H) AND LOW (L) PROPERTY LEVELS

I M F Combination	Property I	Property M	Property F
L L L	3.5	5.0	8
L L H	3.4	5.2	67
L H L	4.4	6.2	7
L H H	3.8	6.3	73
H L L	5.6	5.2	5
H L H	5.9	5.3	63
H H L	6.1	6.5	7
H H H	6.4	7.0	68

F swings from a value of 8 to a value of 67. The average high and low M-values are fairly constant across cells, except for the high-I, high-M cells which had a slightly higher M-value than the other high-M cells. This slight dependence of M on I is not important, since the results showed that the larger swing in M in the high-I cells produced a much weaker effect compared to the smaller swing of M in the low-I cells.

Subidea types. The complete idea sentence structure mentioned above was decomposed into a set of subidea sentence types and coded with a set of type numbers. Table 4 shows the type structures, along with the number of propositions (N) in each type of subidea sentence, and a corresponding example sentence. Types 1 through 4 present the subject and object nouns with each of their two modifiers; Types 5 through 8 contain each modifier in the context of the subject-verb-object relation. Types 9 and 10 contain two modifiers and are partially balanced with regard to the adjective-modifying phrase distinction.

TABLE 4

SUBIDEA TYPES

Type Number	Form	N	Example
1	A_1N_1	1	The ants were hungry
2	N_1M_1	1	The ants were in the kitchen
3	A_2N_2	1	The jelly was sweet
4	N_2M_2	1	The jelly was on the table
5	$A_1N_1VN_2$	2	The hungry ants ate the jelly
6	$N_1M_1VN_2$	2	The ants in the kitchen ate the jelly
7	$N_1VA_2N_2$	2	The ants ate the sweet jelly
8	$N_1VN_2M_2$	2	The ants ate the jelly which was on the table
9	$A_1N_1VN_2M_2$	3	The hungry ants ate the jelly which was on the table
10	$N_1M_1VA_2N_2$	3	The ants in the kitchen ate the sweet jelly
11	N_1VN_2	1	The ants ate the jelly

Type 11 contains only the subject-verb-object relation. The pool of sentences used in the experiment was obtained by constructing the subideas of all types for each of the complete ideas. Types 1 through 4 usually took the form of The (subject noun) was (modifier), but a few were expressed differently for the sake of naturalness.

Acquisition sentences. A subset of the subidea types was selected to make up the set of acquisition sentences for each complete idea. Types 1, 2, 3, 4, and 11 were always included in the acquisition sentence set for each complete idea, enabling the subject to acquire the complete idea from only these one-proposition sentences. One two-proposition and one three-proposition sentence were also included to give a total of seven subidea sentences per complete idea in each acquisition set. Details of the method of selection are given in the design section below. All four complete ideas were used in each I, M, F combination, yielding 28 sentences in the complete acquisition sentence set for a condition.

Question sentences. True questions were subidea sentences from the complete ideas presented in the same condition. The acqusition sentences were arranged so that some of the true questions were explicit, and other true questions were implicit; explicit questions were true since they were exact repetitions of acquisition sentences; implicit questions were also true, since they could be inferred from the acquisition sentences, but were novel, not having appeared in the acquisition sentences. Explicit questions with N equal to 1, 2, and 3 were included; the nature of the acquisition set precluded the existence of one-proposition implicit questions. Four questions, one from each complete idea, were used in each combination of explicit/implicit and N = 2 / N = 3; two subideas from each complete idea appeared as explicit N = 1 questions. There was a total of 24 true questions for each I, M, F condition.

False questions were produced by composing sentences using only the content words that appeared in the complete ideas for that condition; thus, the subject could not base his judgment on the presence of unfamiliar words. Further constraints were that the false sentences could not be semantically anomalous, had to be similar in structure to the subideas, and had to have roughly the same distribution of N as the true questions. The false questions contradicted the complete ideas in a manner similar to the "NONCASE" sentences used by Bransford and Franks (1971) in that they contained propositions incompatible with the complete ideas.

A detailed example of the materials appears in Tables 5 and 6

TABLE 5

EXAMPLE OF A COMPLETE ACQUISITION SENTENCE SERIES

The lobster was scarlet.

The prosecutor was shrewd.

The lice infesting the beggar shocked the inexperienced physician.

The prosecutor who was chewing a cigar scrutinized the unidentified corpse.

The lobster was smothered in sauce.

The prosecutor who was chewing a cigar scrutinized the corpse.

The lice shocked the physician.

The admiral was crusty.

The corpse was unidentified.

The physician was inexperienced.

The monarch was clad in luxurious garments.

The spray was stinging.

The wealthy monarch devoured a lobster which was smothered in sauce.

The lice were filthy.

The admiral braved the spray.

The corpse was in the morgue.

The crusty admiral braved the spray.

The monarch devoured a lobster.

The lice infested the beggar.

The prosecutor was chewing a cigar.

The monarch was wealthy.

The spray was blasted by the hurricane.

The lice shocked the inexperienced physician.

The monarch devoured a lobster which was smothered in sauce.

The prosecutor scrutinized the corpse.

The cursty admiral braved the spray blasted by the hurricane.

The physician administered the vaccination.

The admiral commanded the vessel.

TABLE 6

EXAMPLE QUESTIONS FOR THE ACQUISITION SET SHOWN IN TABLE 5

<u>Explicit (appear in the acquisition set)</u>:

The lobster was scarlet.

The lice shocked the inexperienced physician.

The prosecutor who was chewing a cigar scrutinized the unidentified corpse.

The spray was blasted by the hurricane.

The crusty admiral braved the spray blasted by the hurricane.

<u>Implicit (do not appear during acquisition, but are true)</u>:

The admiral braved the stinging spray.

The shrewd prosecutor scrutinized the corpse which was in the morgue.

The admiral who commanded the vessel braved the spray.

The filthy lice shocked the physician.

The monarch who was clad in luxurious garments devoured the lobster.

<u>False (contradict the acquisition sentences)</u>:

The lice infesting the monarch shocked the physician.

The beggar who was chewing a cigar scrutinized the vessel.

The admiral scrutinized the vessel.

The corpse was clad in luxurious garments.

The physician administered the vaccination to the wealthy monarch.

---

which show a complete set of acquisition sentences and part of a verification sentence set illustrating explicit, implicit, and false questions.

<u>Design</u>

The reading-time condition (RC) was a between-subject factor. The average study time for each level of N from the self-paced subjects was obtained; the force-paced subjects were then run, and were limited to one-half the average study time for sentences of corresponding N. These times were approximately 1.3, 1.6, and 2.3 sec for N = 1, 2, and 3, respectively.

The factors I, M, and F were within-subject; each subject performed the task in all eight combinations of I, M, and F, each combination comprising a separate block in the experiment. Two forms of acquisition set were used so that explicit questions for half of the subjects were implicit for the other half and vice versa. Furthermore, the subideas for acquisition and question sets were chosen to balance out the use of subideas containing adjectival and phrase modifiers. The balancing scheme is summarized by a selection table, Table 7, showing the subidea sentence types represented in each of the two acquisition set forms and the question set. In order to balance effects due to particular complete ideas being assigned to particular columns in the selection table, a 4 × 4 Latin square was used to determine which complete idea would be assigned to each idea number in the selection table.

TABLE 7

SUBIDEA TYPE NUMBERS USED IN ACQUISITION AND QUESTION SETS

		Complete Idea Number 1	2	3	4
Acquisition Set Form	A	9	10	9	10
		5	6	8	7
	B	10	9	10	9
		7	8	6	5
Questions*	N=1	2	1	1	2
		4	3	3	4
	N=2	5a	6a	6b	5b
		7b	8b	8a	7a
	N=3	9a	9b	9a	9b
		10b	10a	10b	10a

* All N=1 questions were explicit. For N=2 and N=3, the suffix (a or b) indicates the acquisition set form that makes the question explicit.

The order of blocks was balanced across subjects by means of an 8×8 Graeco-Latin square; the first term in each cell specified the I, M, F condition and the second (taken modulo 4) specified a row in the 4×4 Latin square used to assign complete ideas in the selection table. Together with the balancing scheme for acquisition set form, multiples of 16 subjects in each reading time condition would yield data balanced with respect to the nuisance variables involved with particular sentences and block order.

Procedure and Instructions

The experiment was implemented by means of the GEPS programming system (Kieras, 1973) on the IBM 1800 laboratory at the Human Performance Center, University of Michigan. The subject was seated in a booth containing a television monitor driven by the computer, and two response pushbuttons, labelled "true" and "false." Up to four subjects were run in each session.

After instructions, the subject began the practice block and continued with the eight blocks containing the different materials. Each block started with the subject viewing the 28 acquisition sentences one at a time. The sentence order was random, subject to the constraint that two sentences from the same complete idea could not appear in succession. Self-paced subjects pressed a button to remove each sentence and cause the next one to appear; the computer recorded the time in centisec each sentence was left on the screen. In the force-paced condition, each sentence was displayed by the computer for the appropriate amount of time determined by N and was automatically replaced by the next sentence.

After the last acquisition sentence, the subject performed a distractor task consisting of making true/false judgments of simple addition problems that appeared on the screen. Twenty problems were given, resulting in a minimum of 40 sec delay between the acquisition and verification phases of the block. The 48 question sentences of the verification task immediately followed completion of the distractor task. The subject viewed a question sentence and judged it true or false by pressing the appropriate response button, whereupon the computer recorded the response and latency in centisec. No feedback was given. The next question appeared after a one sec delay. After answering all questions, the subject rested for one minute before beginning the next block, and was given a long break at the midpoint of the session, which required a total of about two hours.

The instructions specified that in the acquisition phase of each block, the subject should look for and try to put together the four situations or events described by the sentences and be ready for some blocks being harder than others. When making true-false judgments, the subject was told to judge as false a sentence that could be possibly true, but not known to be true of the described situations. For example, if the acquisition sentences described as distinct events ants in the kitchen and a rock crushing a hut, the statement <u>The rock crushed the ants</u> should be judged as false, even though it was possible that the kitchen was in the hut, because this event was not actually described by the acquisition sentences. It was emphasized that accuracy was more important than speed, but the subject should try to take no more time than actually necessary to ensure accuracy.

Subjects

Subjects were University of Michigan students of both sexes who were paid $4.00 for their participation in the experiment.

Since some subjects made many errors in the verification task, steps had to be taken to ensure that the latency data would not be rendered useless by large amounts of missing data. Hence subjects were replaced who produced no correct answer in some cell of the design of the verification task. Each cell was represented by four presented questions; if all four were missed, no estimate of the latency in that cell could be made. Of the 38 self-paced subjects run, four produced at least one empty cell and were eliminated; one other was dropped for failure to understand the instructions. One surplus subject was eliminated to preserve the counterbalancing, bringing the total to 32 subjects in the self-paced group. Of the 37 force-paced subjects run, three were dropped for empty cells, and two eliminated to achieve the final set of 32. Surplus subjects were eliminated on the basis of maintaining highest overall accuracy consistent with preserving the counterbalancing.

CHAPTER III

RESULTS

Study Time Results

The recorded study times for the acquisition sentences were averaged together within each combination of imagery (I), M value (M), frequency (F), and sentence complexity (N), giving 24 data points for each of the 32 self-paced subjects in a completely crossed design. A log transformation, followed by an analysis of variance, was performed on these average times.

Higher levels of I, M, and F produced faster study times than lower; the means for these main effects are shown in Table 8. Significance levels for the properties I, M, and F were $\underline{F}(1,31) = 128$, $\underline{p} < .001$, $\underline{F}(1,31) = 4.74$, $\underline{p} < .05$, and $\underline{F}(1,31) = 13.82$, $\underline{p} < .01$, respectively. The property M had a greater effect at low I, as shown in Table 9; $\underline{F}(1,31) = 7.25$, $\underline{p} < .01$. F and I interacted in the opposite manner shown in Table 10, with F producing a greater effect at high I, $\underline{F}(1,31) = 6.57$, $\underline{p} < .05$.

Longer sentences produced greater study times: Mean times were 2.65, 3.35, and 4.67 seconds for N= 1, 2, and 3, respectively; $\underline{F}(2,62) = 219$, $\underline{p} < .001$. As shown in Table 11, F produced a greater effect

TABLE 8

STUDY TIMES, IN SECONDS, UNDER HIGH AND LOW LEVELS OF I, M, AND F

Property	Level of Property	
	Low	High
I	4.10	3.01
M	3.77	3.34
F	3.82	3.29

TABLE 9

I × M INTERACTION ON STUDY TIMES IN SECONDS

Level of M	Level of I: Low	High
Low	4.45	3.10
High	3.76	2.92

TABLE .0

I × F INTERACTION ON STUDY TIMES IN SECONDS

Level of F	Level of I: Low	High
Low	4.27	3.37
High	3.94	2.65

TABLE 11

F × N INTERACTION ON STUDY TIMES IN SECONDS

Level of F	Level of N: 1	2	3
Low	2.74	3.60	5.11
High	2.55	3.09	4.24

on longer questions; $\underline{F}(2,62) = 4.05$, $\underline{p} < .05$. Table 12 shows that if I was low, M also had a greater effect at larger N; $\underline{F}(2,62) = 6.13$, $\underline{p} < .01$. However, the I × N interaction failed to approach significance.

In summary, higher levels on all three word properties facilitated comprehension in that they reduced study time. I was the strongest

TABLE 12

I × M × N INTERACTION ON STUDY TIMES IN SECONDS

Level of I	Level of M	Level of N 1	2	3
Low	Low	3.30	4.05	5.99
Low	High	2.83	3.66	4.78
High	Low	2.21	3.11	3.99
High	High	2.25	2.58	3.92

facilitator, reducing study time by about a full second. M and F were weaker, producing about a half second decrease. Higher I considerably diminished the effect of M on study time, but substantially increased the effect of F. Both M and F had greater effects on longer sentences, whereas I did not interact with N.

Verification Task Data

The approach. The conventional way to analyze and interpret data of the kind obtained in the verification task would be to consider either the accuracy, or the latency, of question-answering, but not both. This strategy is not feasible since the errors, especially in the force-paced condition, were too numerous to ignore. To take advantage of the fact that valuable information is present in both the latency and accuracy measures, the data will be considered as bivariate, consisting of (latency, proportion correct) pairs.

Analyses of variance were performed on both the latency and proportion correct dimensions. Two separate breakdowns of the verification data were analyzed; the first treated the questions by truth value (V) as being in only two categories, true and false; the second considered only true questions, classified by explicitness (E) as being either explicit or implicit, with N constrained to 2 and 3. To facilitate reading of this section, $\underline{F}$ ratios will be labelled according to whether they are from the true-false, or true-only, analyses of latency or accuracy data where necessary to avoid ambiguity. As a further convenience, since all $\underline{F}$ ratios had degrees of freedom of 1 and 62, the degrees of freedom will not be listed with each $\underline{F}$ ratio.

True-false analysis description. The design for the true-false analysis consisted of six factors; subjects were nested within reading condition (RC); RC, I, M, F, and V were completely crossed with two levels each, giving 16 combinations of conditions for each of the two groups of 32 subjects. A total of 24 questions were represented in each cell. Average question latencies were obtained by averaging the latencies of correct answers in each of the 16 cells for each subject. An analysis of variance was performed on the log-transformed average times. Accuracy scores were obtained by tabulating the proportion correct in each cell for each subject; an arcsine transformation (Winer, 1962) was applied, followed by an analysis of variance. Appendix B contains the mean latencies and accuracies for each cell; Table 13 is a summary of the analyses of variance, showing only those effects achieving at least a 0.1 level of significance.

True-only analysis description. The data from true questions were analyzed similarly to the true-false analysis. The design had seven factors; subjects were nested within RC; RC, I, M, F, N, and explicitness (E) were completely crossed with two levels each, giving 32 conditions for each of the two groups of 32 subjects. Four questions were represented in each cell of the design. Average latencies were computed by averaging the latencies to correct responses in each of the

TABLE 13

SUMMARY OF TRUE-FALSE ANALYSIS OF VERIFICATION DATA[1]

EFFECT[2]	LATENCY EFFECT MS	LATENCY ERROR MS	LATENCY F[3]	ACCURACY EFFECT MS	ACCURACY ERROR MS	ACCURACY F
RC	-	-	-	8.266	.271	30.50**
I	25.197	.069	366.45**	29.268	.109	268.53**
M	1.003	.061	16.44**	3.499	.070	50.08**
F	.345	.065	5.28*	4.731	.068	6.91*
RC×I	.277	.069	4.02*	.838	.109	7.69**
I×M	-	-	-	.511	.056	9.11**
I×F	.196	.043	4.55*	.981	.057	17.19**
M×F	.161	.044	3.67	-	-	-
RC×V	.209	.053	3.91	-	-	-
I×V	.921	.022	41.86**	.642	.082	7.80**
M×V	.687	.016	42.67**	.410	.033	12.42**
F×V	.179	.022	8.24**	-	-	-
I×M×F	-	-	-	.154	.052	2.96
I×M×V	-	-	-	.856	.032	26.75**
I×F×V	.202	.021	9.72**	-	-	-
M×F×V	.157	.015	10.18**	-	-	-

[1] Latencies were log-transformed; accuracies were arcsine-transformed.

[2] Only effects having $p < .1$ are shown.

[3] All _F_ ratios had 1 and 62 degrees of freedom.

* $p < .05$

** $p < .01$

32 cells for each subject. A log transformation was applied to the averages and the results subjected to an analysis of variance. Accuracy was analyzed by computing the proportion correct in each cell, applying an arcsine transformation, followed by an analysis of variance. Appendix C contains the mean latencies and accuracies for each cell;

TABLE 14

SUMMARY OF TRUE-ONLY ANALYSIS OF VERIFICATION DATA[1]

EFFECT[2]	LATENCY EFFECT MS	LATENCY ERROR MS	LATENCY F[3]	ACCURACY EFFECT MS	ACCURACY ERROR MS	ACCURACY F
RC	10.548	2.049	5.17**	18.869	1.294	14.58**
I	28.700	.170	168.82**	68.482	.417	164.22**
M	-	-	-	1.965	.360	5.45*
F	1.994	.178	11.20**	17.832	.421	42.33**
E	8.163	.074	109.57**	49.030	.377	130.05**
N	18.064	.096	181.24**	7.912	.239	33.10**
RC×I	.703	.170	4.127*	2.591	.417	6.21*
RC×M	.386	.122	3.164	-	-	-
I×F	-	-	-	6.525	.382	17.09**
I×E	.737	.105	7.03*	5.825	.267	21.84**
M×E	-	-	-	.928	.309	3.00
F×E	-	-	-	2.572	.268	9.63*
RC×N	.269	.096	2.79	-	-	-
I×N	.834	.091	9.22**	.834	.164	5.08*
E×N	.230	.075	3.04	1.259	.128	6.99*
RC×I×E	-	-	-	1.536	.267	5.76*
M×F×N	.374	.086	4.34*	-	-	-
RC×I×M×E	.275	.090	3.06	-	-	-
RC×I×F×N	-	-	-	.601	.196	3.06
RC×I×M×F×E×N	-	-	-	.755	.230	3.28

[1] Latencies were log-transformed; accuracies were arcsine-transformed.

[2] Only effects having $p < .1$ are shown.

[3] All F ratios had 1 and 62 degrees of freedom.

* $p < .05$

** $p < .01$

Table 14 contains a summary of the analyses of variance, showing only those effects achieving at least a 0.1 significance level.

N produced some main effects and interactions which are shown in Table 14. However, the effects of N will not be considered in detail here, but will be discussed briefly later; most of the results to be described were obtained by collapsing over N as well as other factors.

Method of display of the results. As mentioned above, the data consist of (latency, proportion correct) points for the different conditions. The interpretation of the data will make use of the descriptive model parameters (Chapter I) for different conditions which take a similar (time, probability) form. The surface processing is described by estimates of the S-process time and S-match accuracy in $(\hat{T} + \hat{t}_s, \hat{p}_s)$ points; the meaning match process is described by M-match time and accuracy estimates for true and false questions. For the serial version of the descriptive model, these estimates are $(\hat{t}_m, \hat{p}_m)$ for true questions, and $(\hat{t}_{mf}, \hat{p}_f)$ for false questions. The corresponding points for the parallel version are $(\hat{t}_m - \hat{t}_s, \hat{p}_m)$ and $(\hat{t}_{mf} - \hat{t}_s, \hat{p}_f)$.

The most effective way to display these bivariate results is in the form of a two-dimensional performance plot which shows the (time, probability) points for the data and the model parameters for different conditions. This method of display is illustrated by Figures 6 and 7 which present the data and corresponding model parameters for the three question types after collapsing across all factors in the experiment. Referring to Figure 6, explicit questions were answered with about a 2.7 sec latency and 92% accuracy, false questions with 2.7 sec latency and 87% accuracy, and implicit questions with 3.1 sec latency and 84% accuracy. Applying the estimation formulas in Chapter I yields the

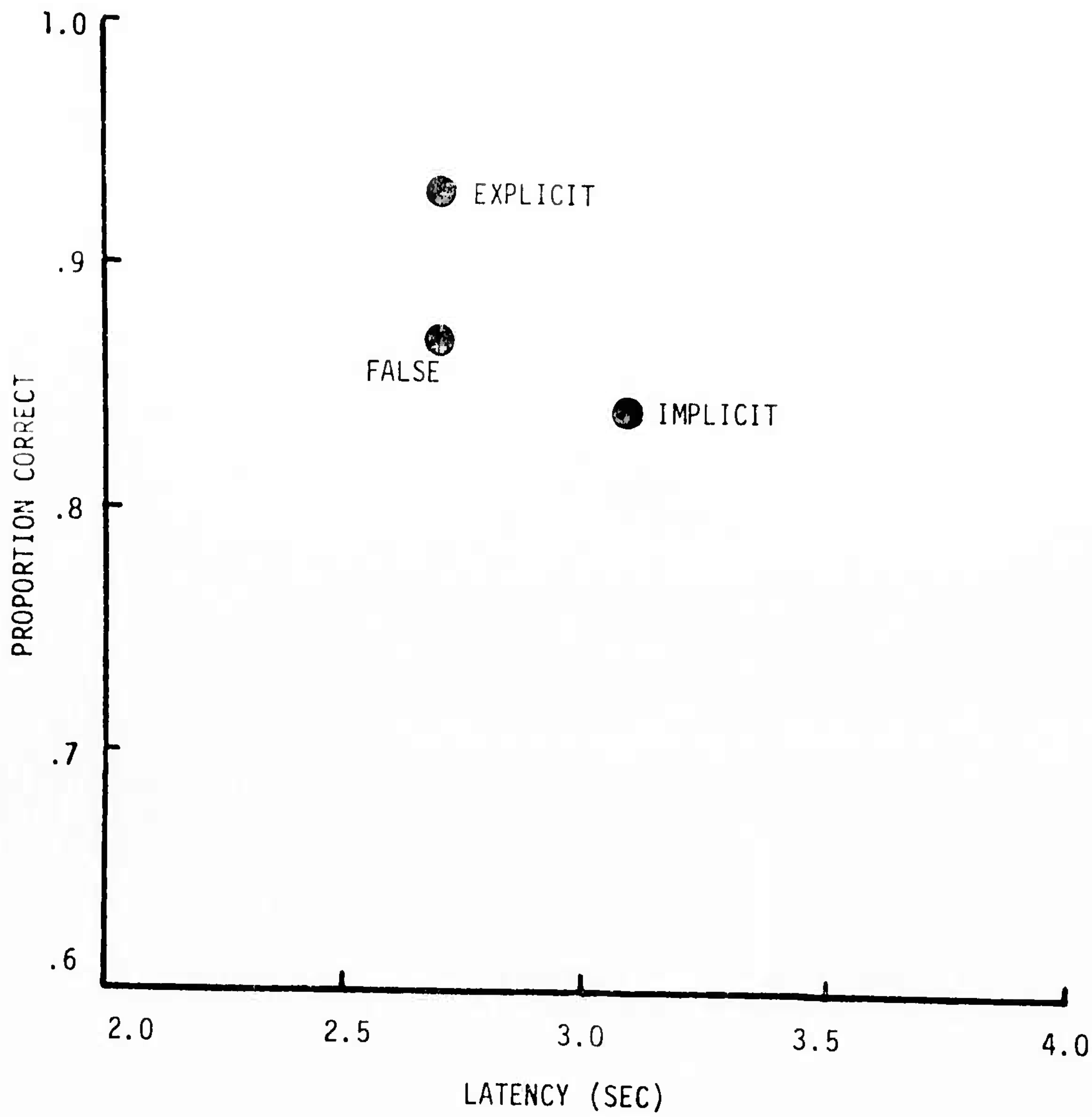

Figure 6. Speed and accuracy coordinates for explicit, implicit, and false questions. Notice that there is a true performance difference between each pair of question types.

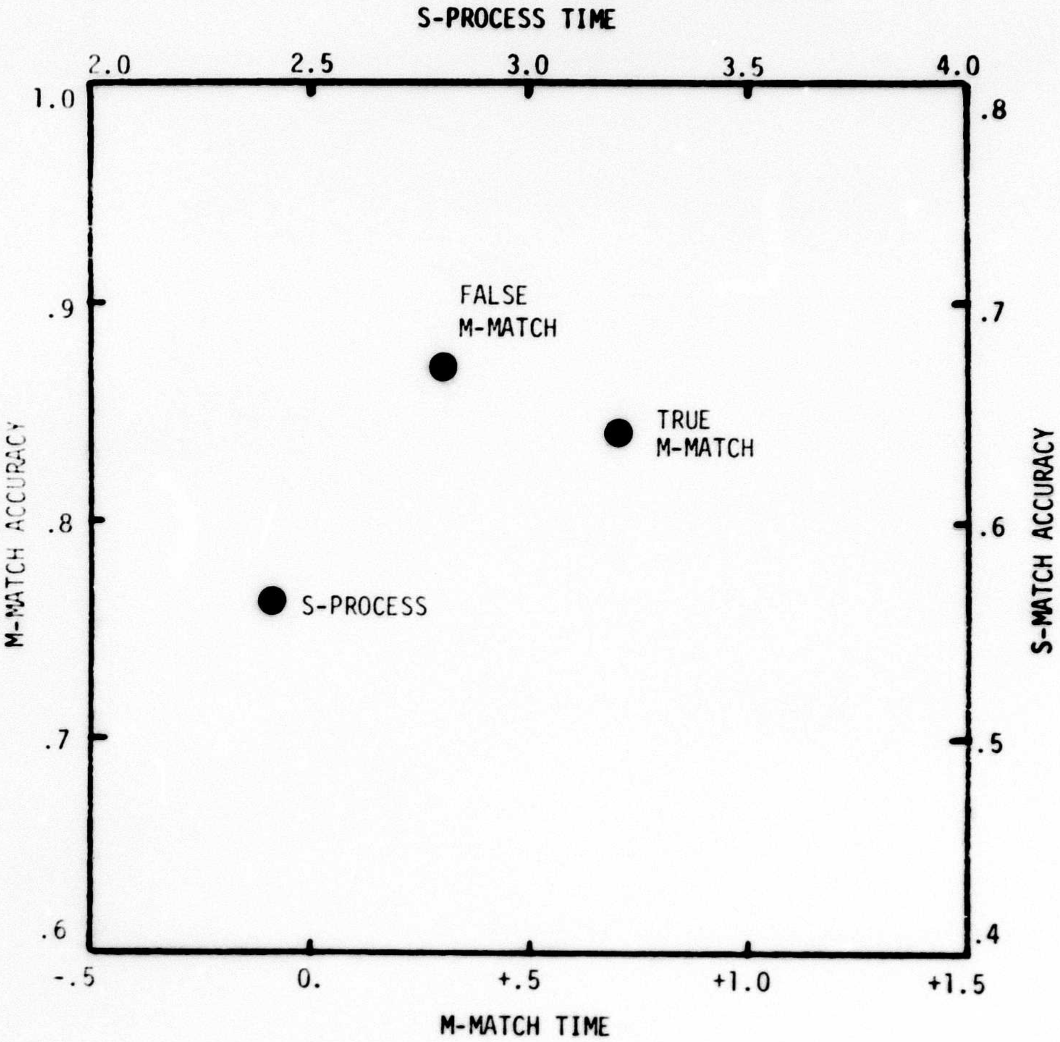

Figure 7. Descriptive model parameters for the three question types. The S-process point is plotted against the upper abscissa and right-hand ordinate. The M-match parameters for true and false questions are plotted against the lower abscissa and left-hand ordinate. Notice that higher performance appeared on the false M-match than on the true M-match. The S-process point can not be compared with the M-match points, since they represent different processes in the descriptive model.

parameter values shown in Figure 7. Since both M- and S-processing parameters are shown on the same set of axes, the scales require explanation. The M-match parameters use the lower abscissa and left-hand ordinate scales; the S-process parameters are plotted with the upper abscissa and right-hand ordinate scales. Hence the M-match time for true questions is .71 sec and the accuracy estimate is .89, while for false questions, the time and accuracy estimates are .31 sec and .87. The S-processing parameters are 2.39 sec for the time and .56 for the S-match accuracy.

The value of the performance plot lies in the ease with which differences in the locations of points from different conditions can be interpreted in terms of performance. Points falling on a negative diagonal show a genuine performance difference, since a point in the upper left-hand corner of the plot indicates high performance, with both high speed and high accuracy, whereas a point in the lower right-hand corner shows poor performance due to both low speed and low accuracy. Examples of performance differences are the points for implicit and explicit questions in Figure 6 and the M-process points for true and false questions in Figure 7. Differences in location of points falling on a positive diagonal are ambiguous, being a result of either a genuine performance improvement in one dimension accompanied by a performance degradation in the other, or a shift in speed-accuracy tradeoff reflecting a change in the subjects' criteria for performance (cf. Pachella, 1973).

The effects of a factor on verification performance will be displayed by showing the data points for the three question types at both levels of the factor, with the two points for each question type connected by an arrow. This somewhat elaborate plot shows graphically how the factor pushes performance on each question type from one location in the speed-accuracy space to another. A similar display of arrows will be used to show the effects of a factor on the model parameters.

Reading condition. The reading time manipulation produced an ambiguous speed-accuracy shift in the verification data, shown in Figure 8, that the descriptive model helps to clarify. Notice that on all question types force-pacing produced lower accuracy and higher speed. It is not surprising that force-paced subjects were less accurate than the self-paced subjects; true-false $F = 30.50$, $p < .01$. However, it is not clear why force-paced subjects answered questions faster by about a third of a second than the self-paced subjects, true-only $F = 5.17$, $p < .05$, even though the speed stress was only during acquisition, not during verification.

The speed increases for explicit and implicit (explicitness, E) questions were identical, as shown by the nonsignificant RC × E: $F < 1$, which suggests that the speed increase rests in a process equally involved in both types of questions. Correspondingly, the S-process parameters in Figure 9 show a speed increase, which is also accompanied by an accuracy loss. Furthermore, the fact that less of a speed increase is made on false questions than on true (truth value, V), RC × V: $F = 3.91$, $p < .1$, suggests that force-pacing causes some process to take more time in the case of false questions than in true. The M-match parameters bear this out, showing a performance degradation that is greater for false questions than true.

In summary, force-pacing produced a mixed effect of a speed

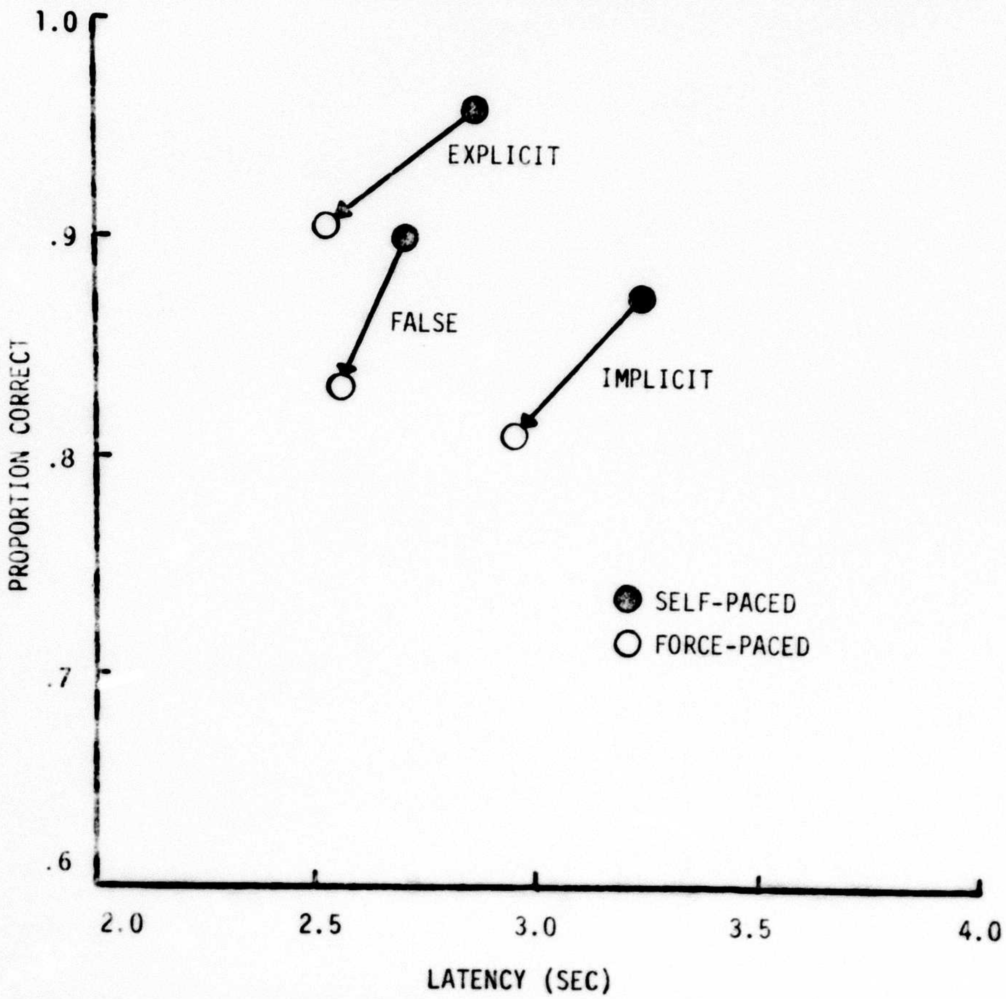

Figure 8. Data for the effects of RC. Notice the overall increase in speed and decrease in accuracy produced by force-pacing.

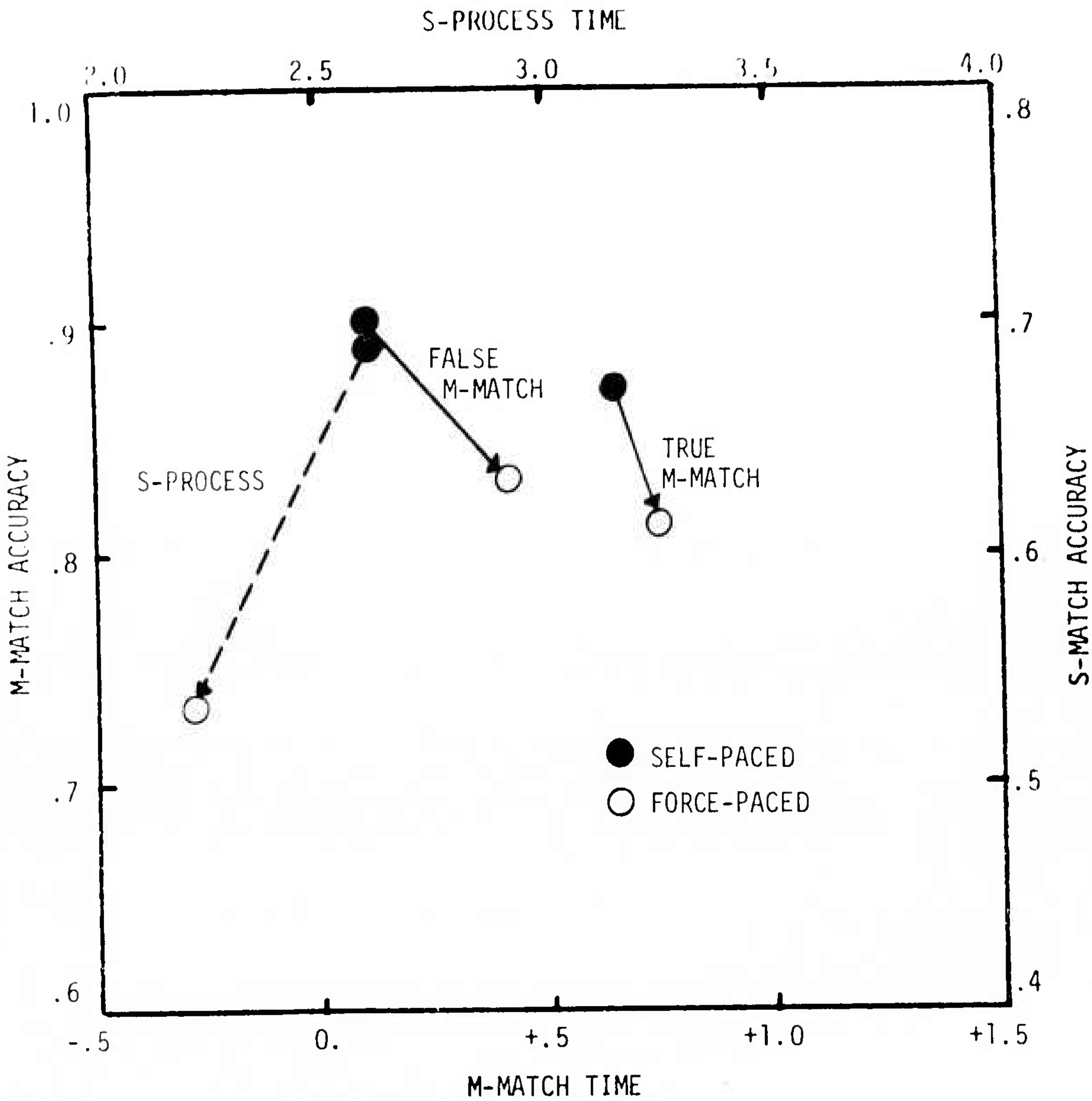

Figure 9. Model parameters for the effects of RC. M-match performance was degraded (slower speed and less accuracy); S-process speed was increased, but S-match accuracy was decreased.

increase and an accuracy loss on surface processing, and a meaning match degradation which was more severe on false questions than on true.

Frequency. The descriptive model also helps in the interpretation of the ambiguous effects of F shown in Figures 10 and 11. Overall, higher F produced lower question accuracy; true-false $F = 6.91$, $p < .05$, true-only $F = 42.33$, $p < .01$. Higher F also produced higher speed; true-false $F = 5.28$, $p < .05$, true-only $F = 11.20$, $p < .01$. As in the reading condition effects, the similar speed increases indicate that the overall increase resides in a process common to both explicit and implicit questions; F × E: $F < 1$. Again, the corresponding S-process parameters show a speed increase. The lack of a speed increase due to higher F on false questions, supported by F × V: $F = 8.24$, $p < .01$, indicates that some process takes longer in false questions than in true. There was only a slight change in the M-match time for true questions, but a substantial increase in time for false questions, as shown in Figure 11.

Reading condition (RC) and frequency (F) demand comparison, since both produced faster but less accurate verification performance. Force-pacing and higher F both increased S-process speed and slowed down the M-match. However, force-pacing lowered S-match accuracy, but higher F did not appreciably affect it. Force-pacing degraded false M-matches more than true; higher F produced a speed loss on false M-matches but little or no speed change in true M-matches, but a greater accuracy loss on true M-matches than false. This pattern of differences in the effects of RC and F is more distinct in the low-I conditions which will be discussed later. So, even though RC and F had superficially similar effects, the two factors are actually quite distinct in their effects on

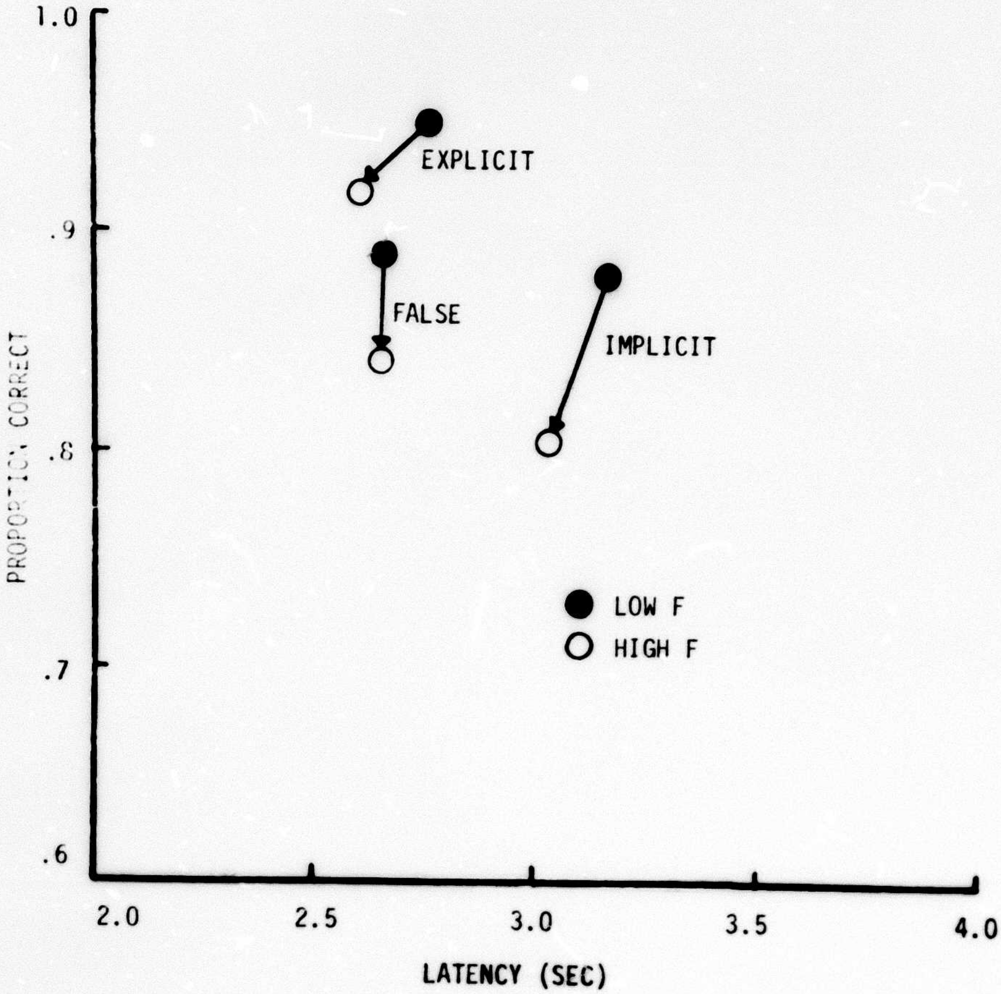

Figure 10. Data for the effects of F. Notice the overall speed increase and accuracy decrease, with similar time differences on implicit and explicit, and little time difference on false questions.

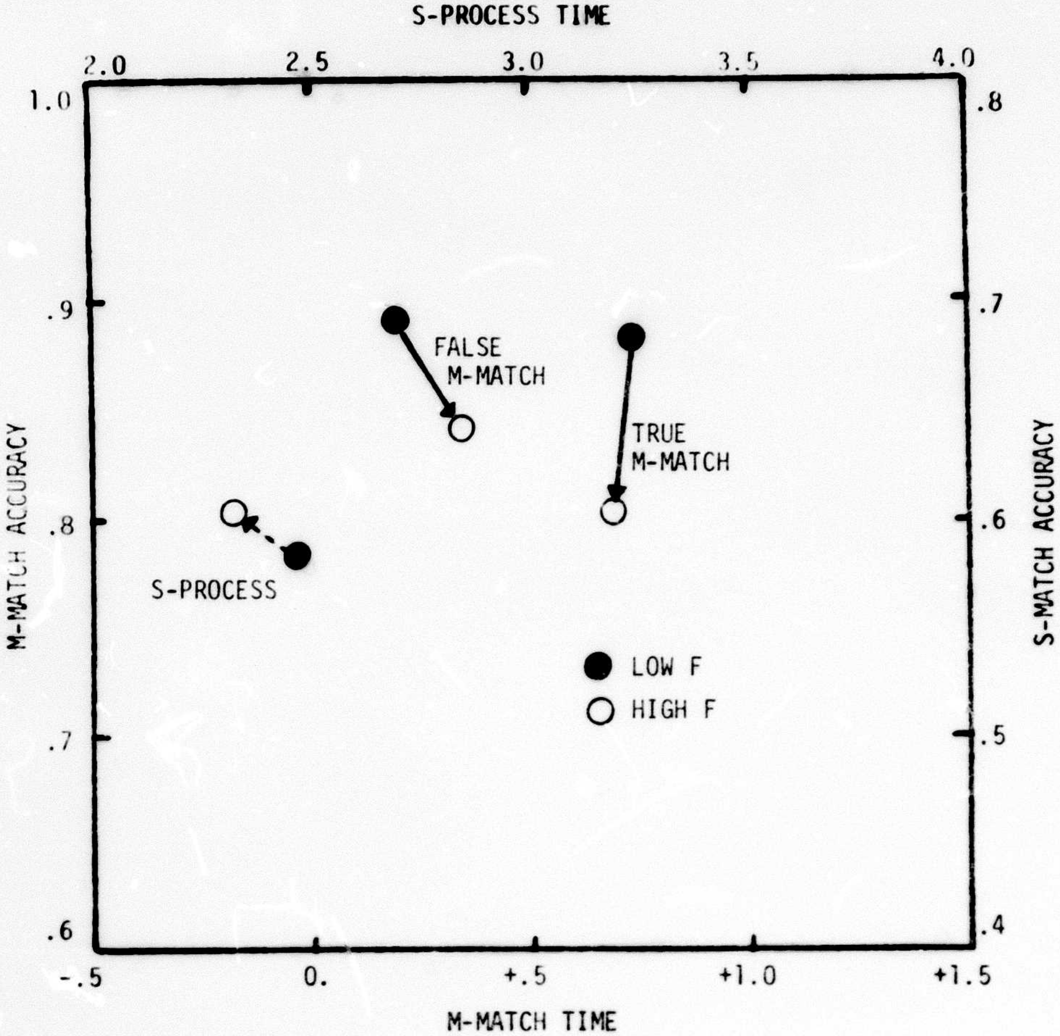

Figure 11. Model parameters for the effects of F. The S-process speed improved slightly, and M-match accuracy went down. Note the differences in M-match time effects for true and false questions.

the verification subprocesses.

M-value. In contrast to the study time effects, the effect of M on verification performance, as shown in Figure 12, was fairly weak. M had essentially no effect whatever on explicit questions, which means that M had no influence on surface processing, implying that the model parameters look almost exactly like the data, as can be seen in Figure 13. Despite the weakness of M, it did produce significant main effects; true-false analysis: latency $F = 5.28$, $p < .05$, accuracy $F = 6.91$, $p < .05$. However, the effect of M differed for true and false questions; latency $M \times V$: $F = 42.67$, $p < .01$, accuracy: $F = 12.42$, $p < .01$. Thus the major effect of M on verification performance was to improve M-match processing on false questions. No effect on surface processing appeared, and the effect on true M-matches was very small.

Imagery. I-value produced the strongest effects on verification performance, as it did with study times. The relevant main effect statistics for the results in Figure 14 and Figure 15 are: true-false analysis: latency $F = 366$, $p < .001$, accuracy $F = 30.50$, $p < .01$; true-only analysis: latency $F = 168$, $p < .001$, accuracy $F = 164$, $p < .001$. The lesser effect of I on explicit questions, $I \times E$: latency: $F = 7.03$, $p < .05$, accuracy: $F = 21.84$, $p < .01$, is simply a matter of the S-match being successful often enough to compensate partially for the poorer M-match performance on low-I material. However, the relation between the surface processing and I is complicated by an interaction with RC to be discussed below. The $I \times$ truth value (V) interactions were also significant, but this is due to the fact that the true-false analysis collapsed over explicit and implicit for the true questions; hence the $I \times V$ interaction is misleading. The clearly equal speed and accuracy

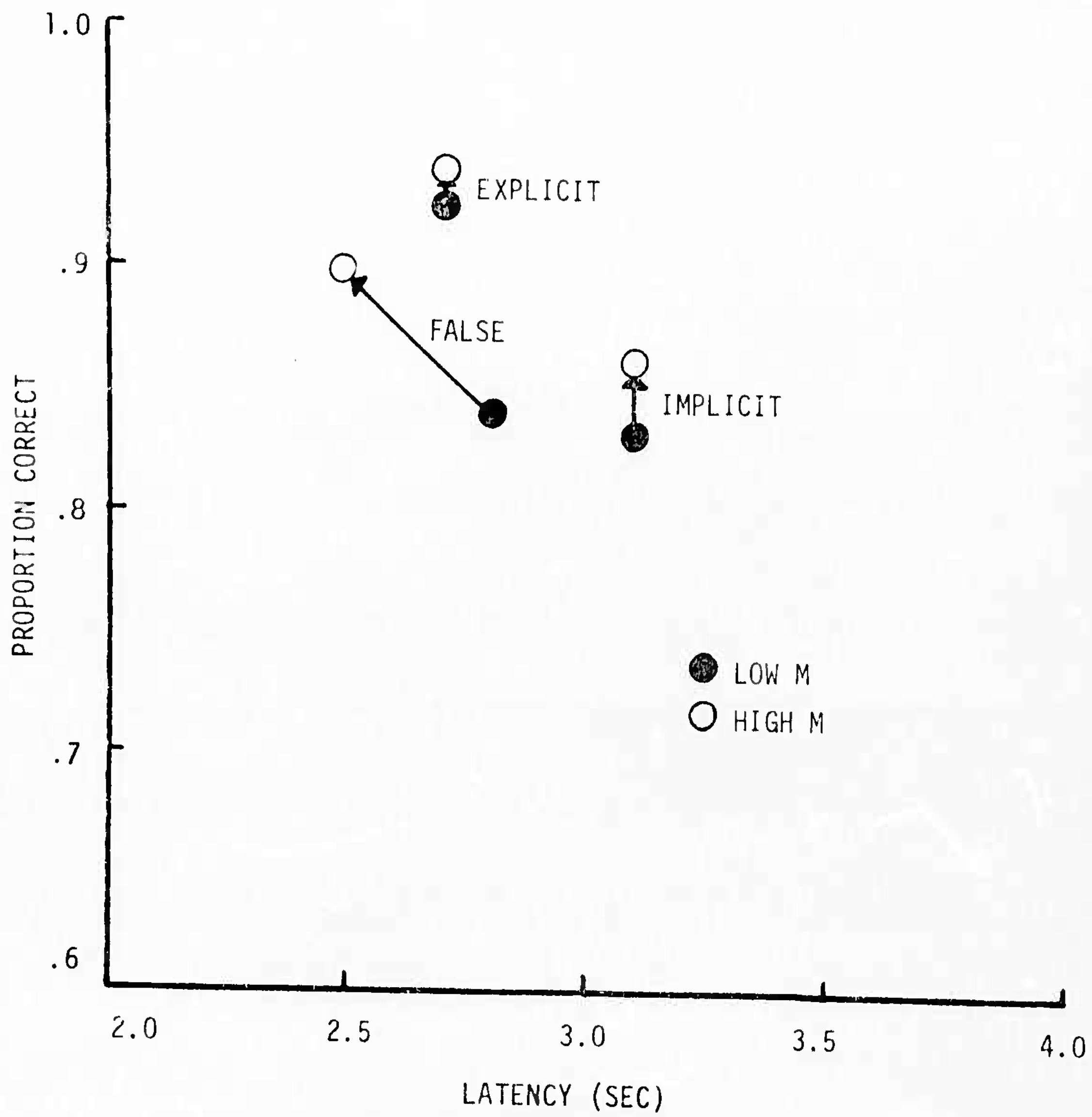

Figure 12. Data for the effects of M. There was no effect on explicit questions, a slight accuracy effect on implicit questions, and a large speed-accuracy improvement on false questions.

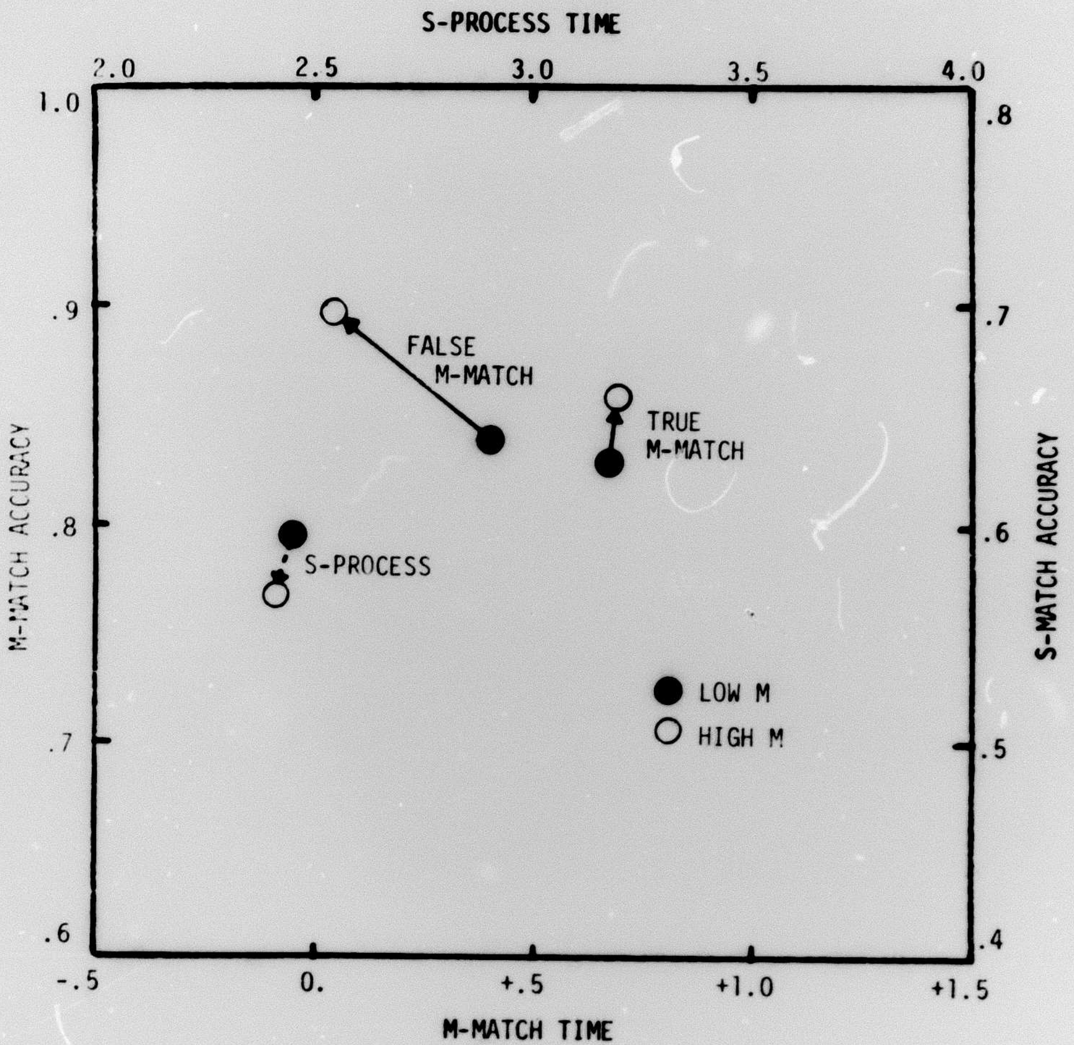

Figure 13. Model parameters for the effects of M. Nothing happens to the S-process, and the M-match parameters correspond directly to the data.

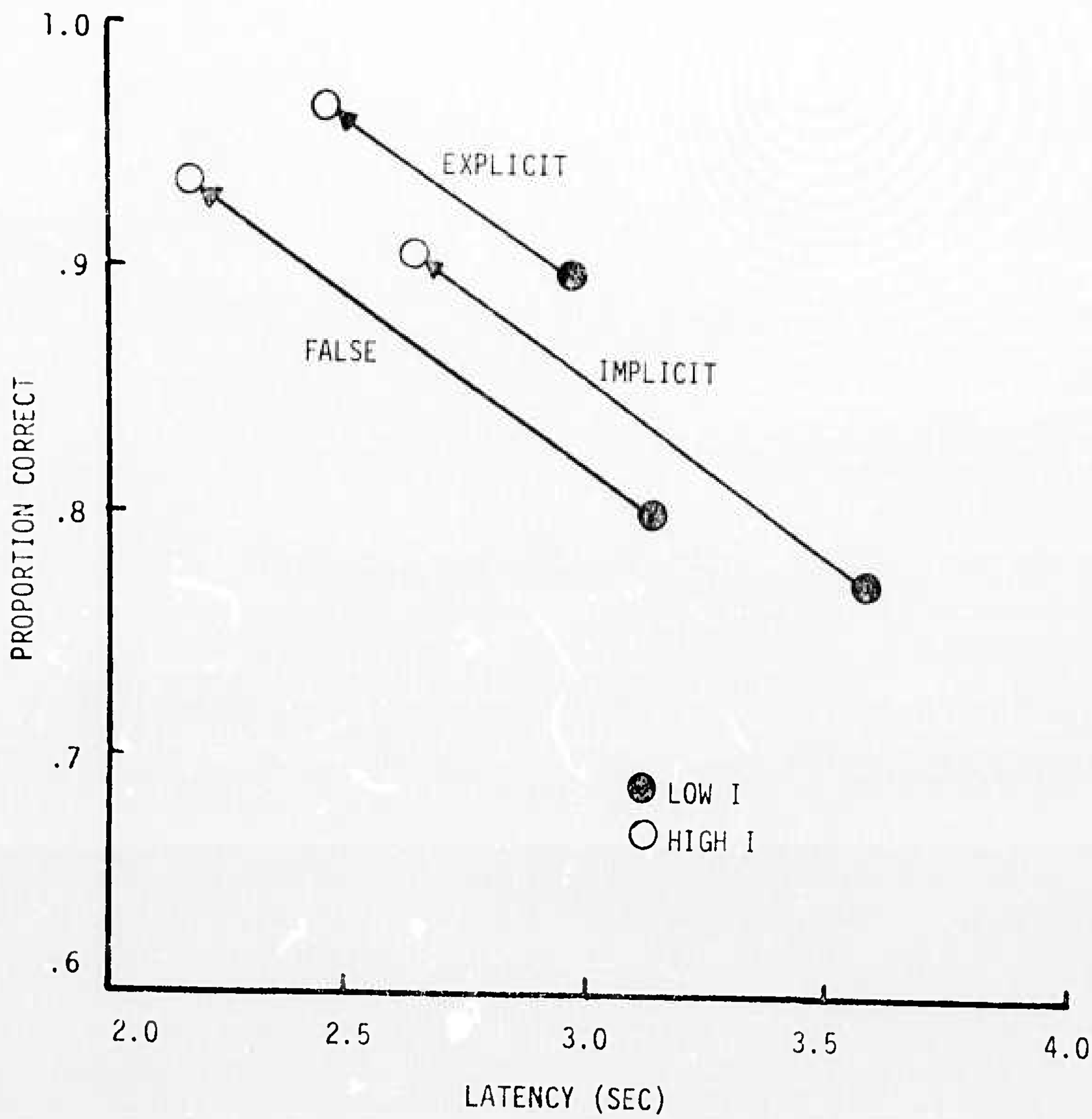

Figure 14. Data for the effects of I. Notice the large and equal speed-accuracy improvements for false and implicit questions.

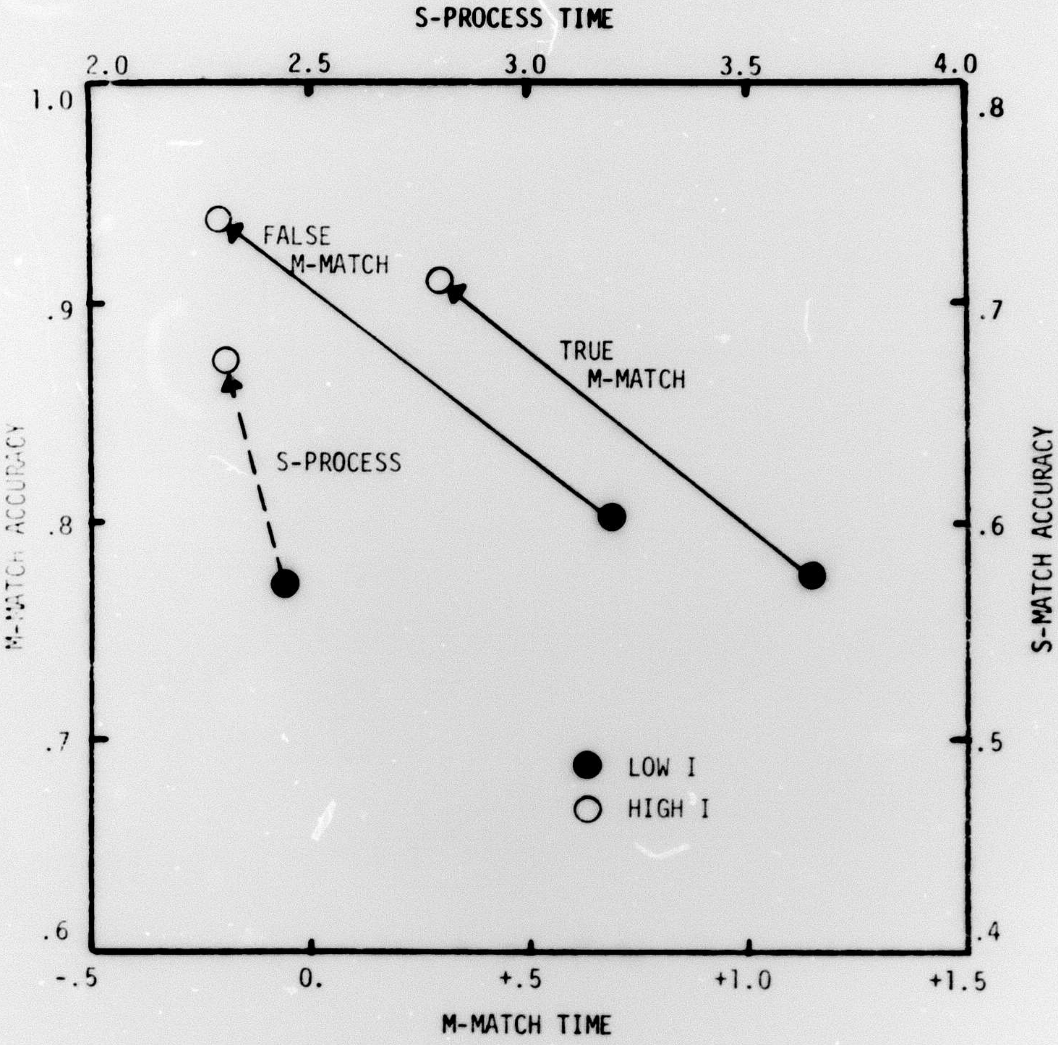

Figure 15. Model parameters for the effects of I. The M-match effects parallel the data; the S-process shows an improvement under higher I.

increases produced by I on implicit and false questions indicate that the way in which I influenced the M-match process was the same for true and false questions. This equal effect on true and false M-matches stands in contrast to the effects of the other three factors, which produced different effects on true and false questions.

Interactions of imagery with other factors. To some extent, the interactions involving I force modification of some of the above conclusions, but overall, with low-I materials, the effects of RC, M, and F described above appear more definitely; with high-I materials, RC and M are ineffective, but F has a qualitatively different effect.

I and RC interacted in both analyses; true-false analysis: latency $F = 4.02$, $p < .05$, accuracy $F = 7.69$, $p < .01$; true-only analysis: latency $F = 4.13$, $p < .05$, accuracy $F = 6.21$, $p < .05$. The I × RC interaction was affected by the true question type; accuracy I × RC × E: $F = 5.76$, $p < .05$; but no I × RC × V interaction approached significance, p .1. The effects of RC under low I, shown in Figure 16, look like an exaggerated version of those in Figure 8 with an overall lower level of performance. The model parameters in Figure 17 show the previous greater effect on false M-matches than true, although there is a stronger effect of force-pacing increasing true M-match time as well as false. The surface parameters show a drastic decline in accuracy, along with a large increase in speed, produced by force-pacing. Under high I, the effects of RC are much reduced. At these high levels of p_m, the estimation function for p_s is quite unstable, so it is not clear whether the increase in $\hat{p}_s$ under force-pacing is meaningful.

The plots for the I × M interaction shown in Figures 18 and 19 need little comment. Only interactions involving false question

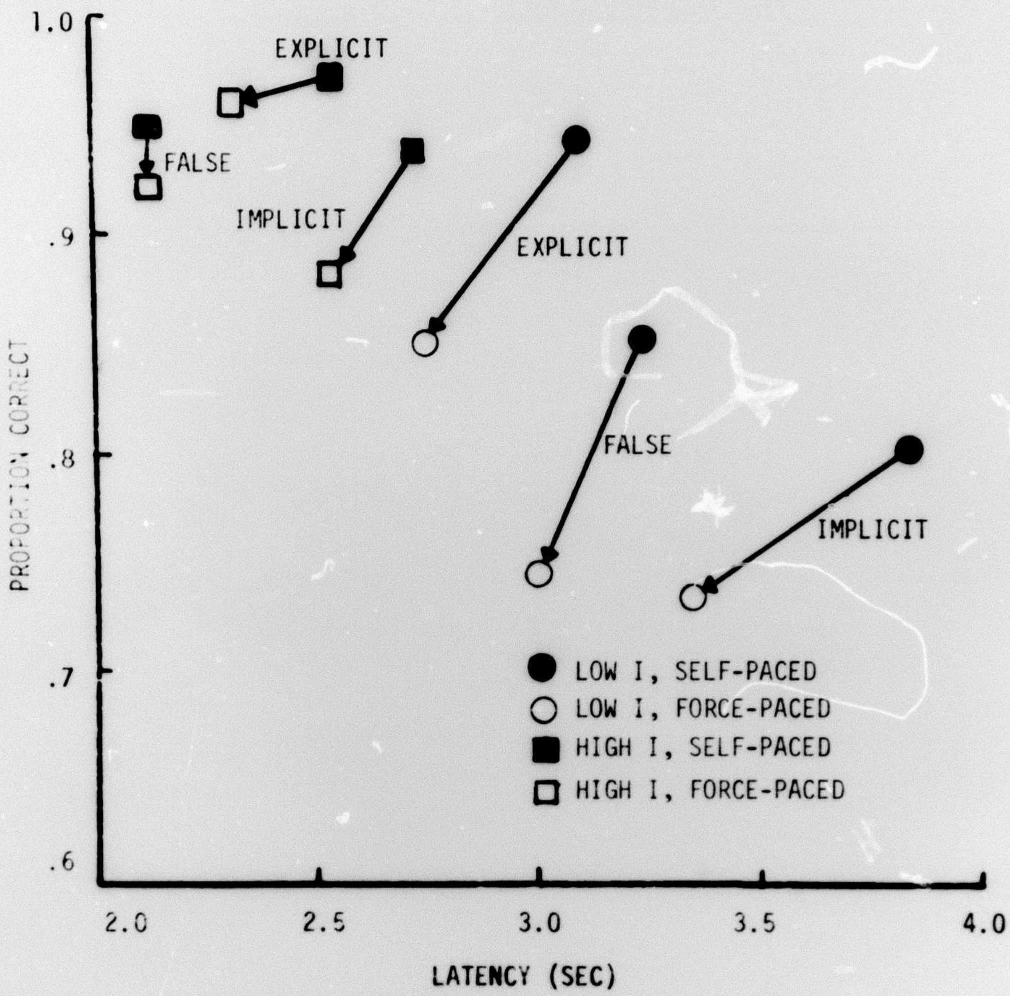

Figure 16. Data for the I × RC interaction. The high-I points are the squares in the upper corner of the diagram; the low-I points are the circles in the lower part of the diagram. Notice the diminished effects of force-pacing when I is high.

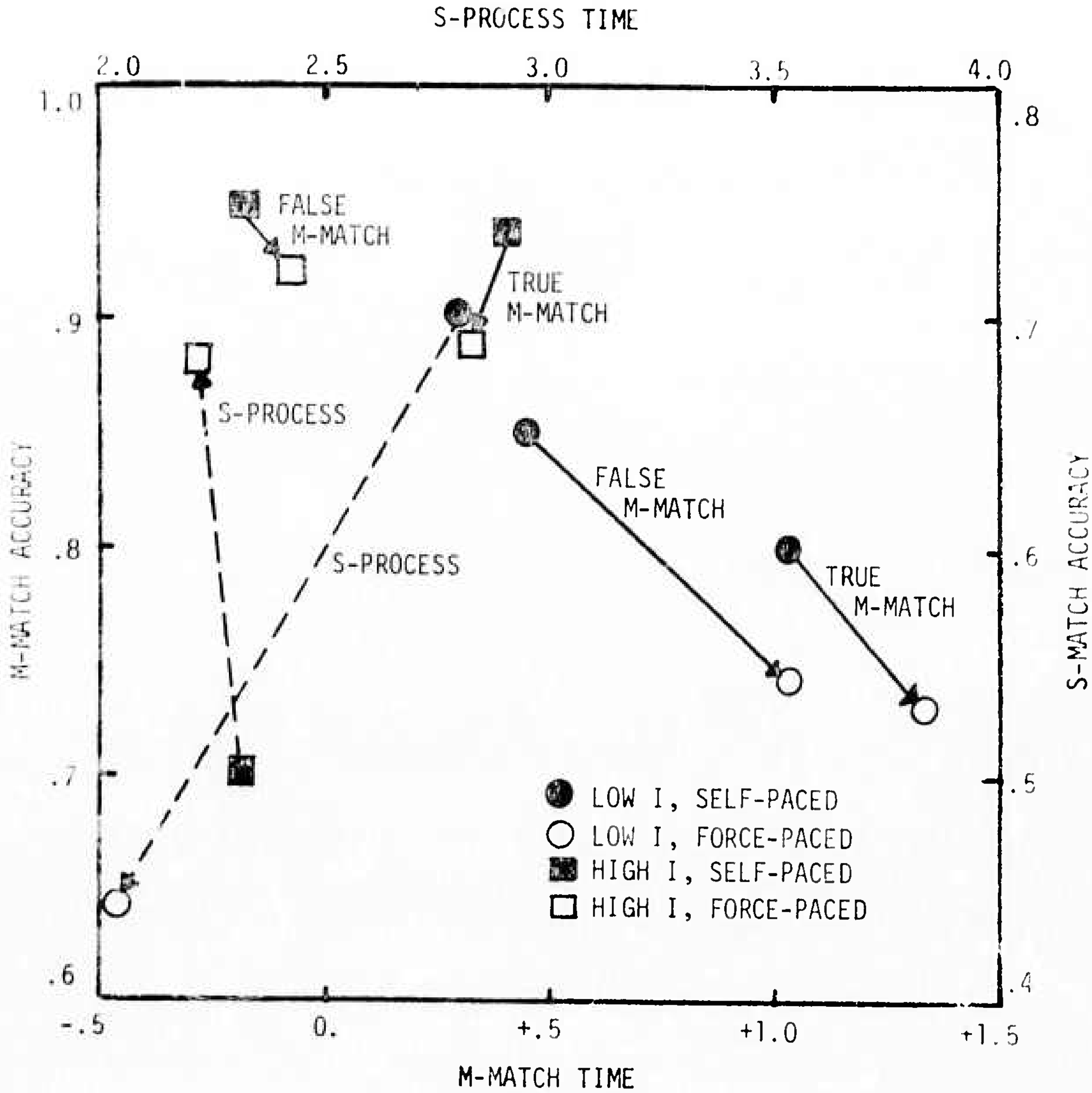

Figure 17. Model parameters for the I × RC interaction. Again, the square points denote the high-I conditions, the round ones the low-I conditions. The low-I M-match parameters resemble those in Figure 9. The S-process shows a speed gain and accuracy loss under low-I and an accuracy gain under high-I.

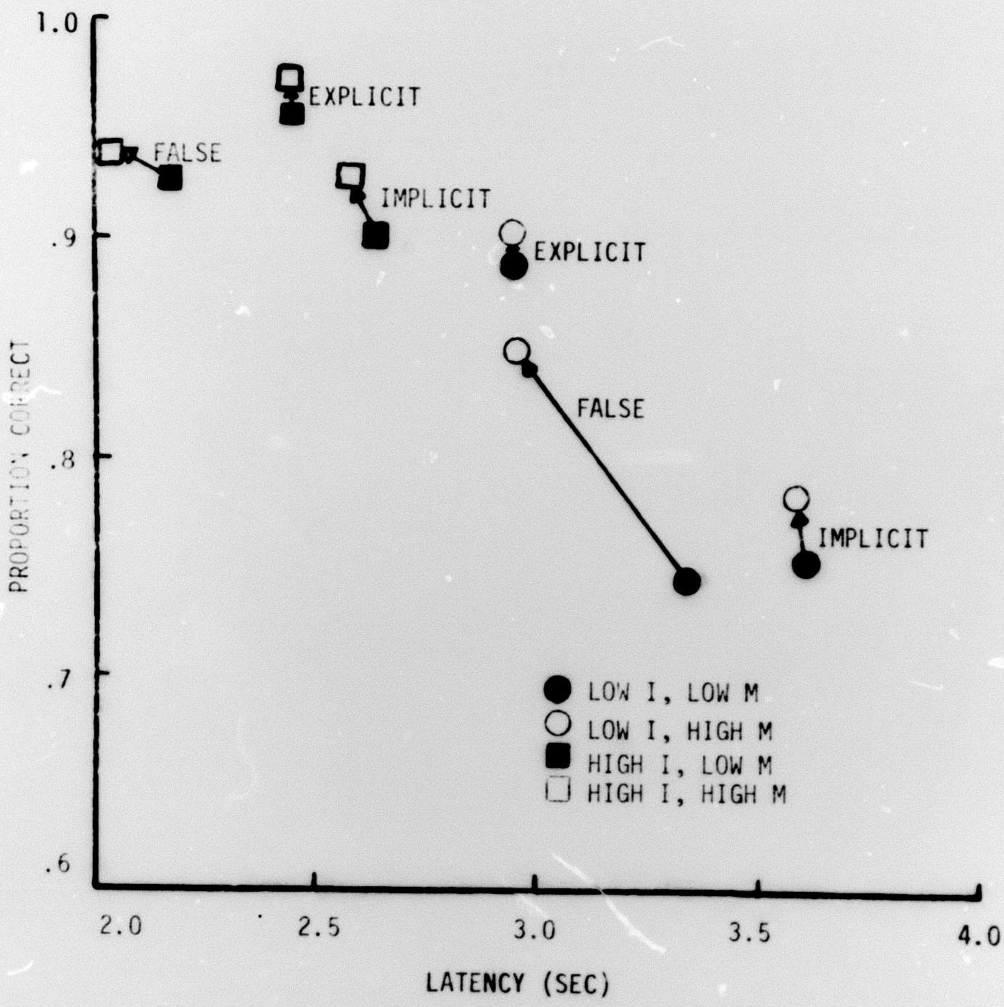

Figure 18. Data for the I × M interaction. Under high-I, M makes little difference. The pattern of Figure 12 appears more strongly under low-I.

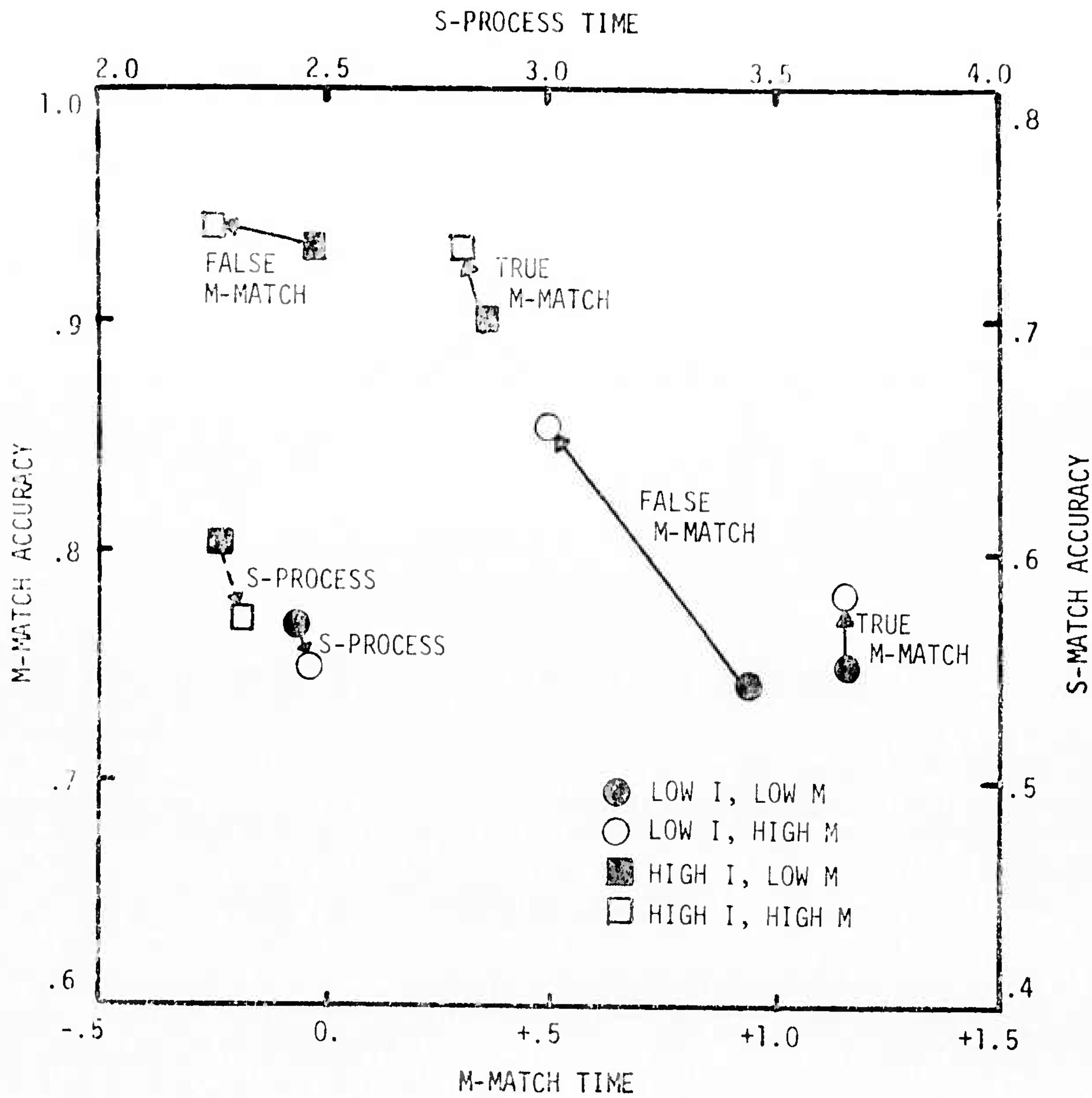

Figure 19. Parameters for the I × M interaction. As in the data, there is no effect of M under high-I. Compare the low-I parameters with Figure 13.

accuracy data were significant: true-false I × M: $F = 9.11$, $p < .01$, I × M × V: $F = 26.75$, $p < .01$. Under low I, the effect of M is exactly as described before, with the effect on false questions much larger than that on true. Under high I, the effect of M on false questions is considerably weakened. It is likely that the I × M interaction on verification accuracy is a ceiling effect in that high-I materials could produce as much accuracy as subjects are capable of.

Under low I, the effects of F, as shown in Figures 20 and 21, are again an exaggeration of the overall effects described earlier. However, under high I, F produced a radically different effect than at low I. Statistics for the I × F interaction are: true-false analysis: latency $F = 4.55$, $p < .05$, accuracy $F = 17.19$, $p < .01$; true-only analysis: latency $F < 1$, accuracy $F = 17.09$, $p < .01$. With I low, F produced a large accuracy loss in the M-matching and a slight speed increase in the surface processing. On false questions, the M-match took longer for high F than for low F. On the other hand, with I high, there was only a small accuracy loss and a drastic speed increase in the M-matching on false questions, corresponding to the true-false latency I×F×V: $F = 9.72$, $p < .01$. Thus, rather than another ceiling effect, F appears to operate qualitatively differently in high- and low-I materials.

Summary of Results

All three word properties produced decreased study times, but produced distinct effects on verification performance, as did the reading time manipulation.

I was very strong, decreasing study time substantially and drastically improving verification speed and accuracy equally for true and false questions. M decreased study time and improved verification

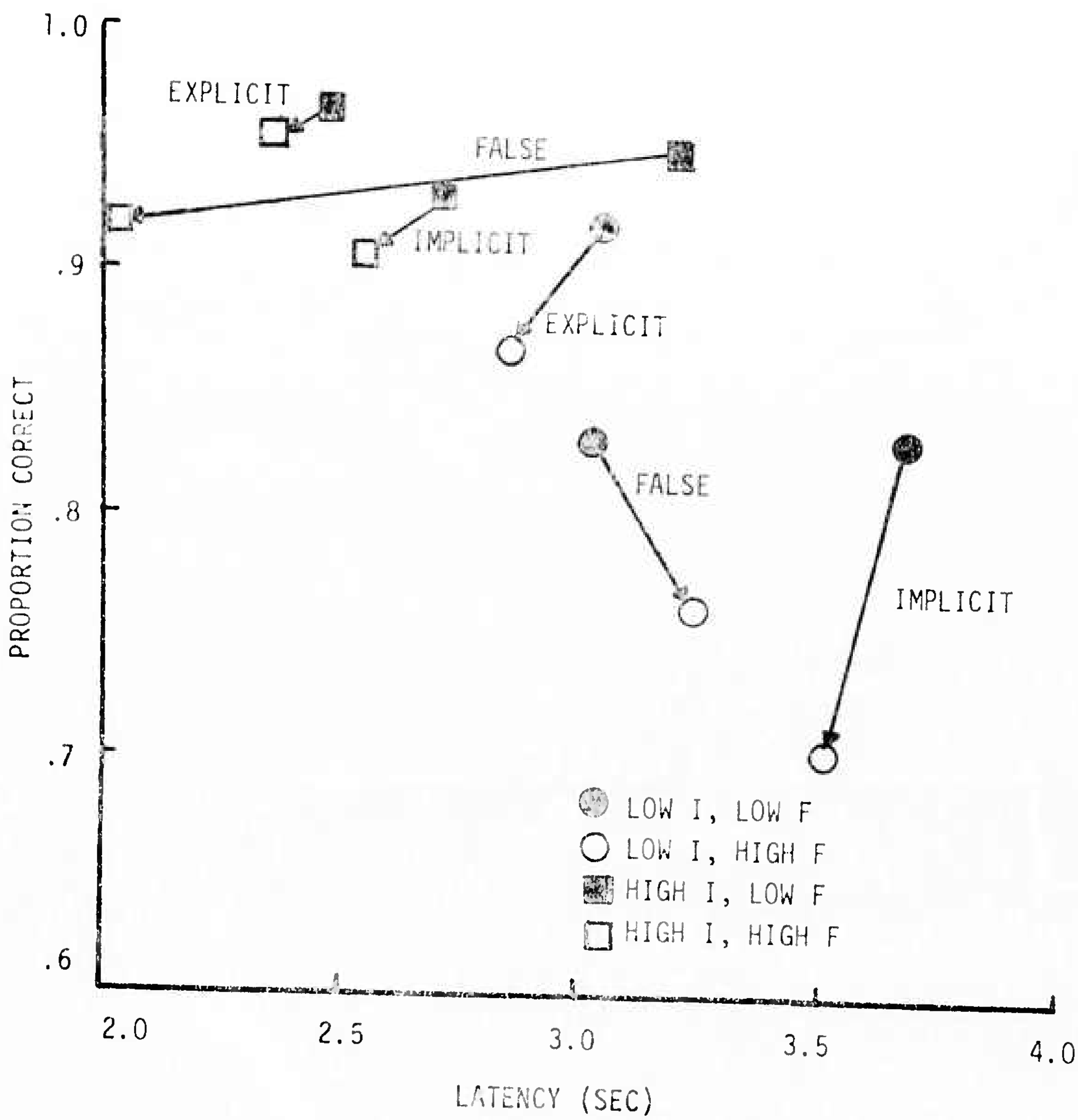

Figure 20. Data for the I × F interaction. Notice the similarity of the low-I pattern to the overall effects in Figure 10, but with a definite speed loss on false questions. Under high-I, false questions are extremely fast if F is high.

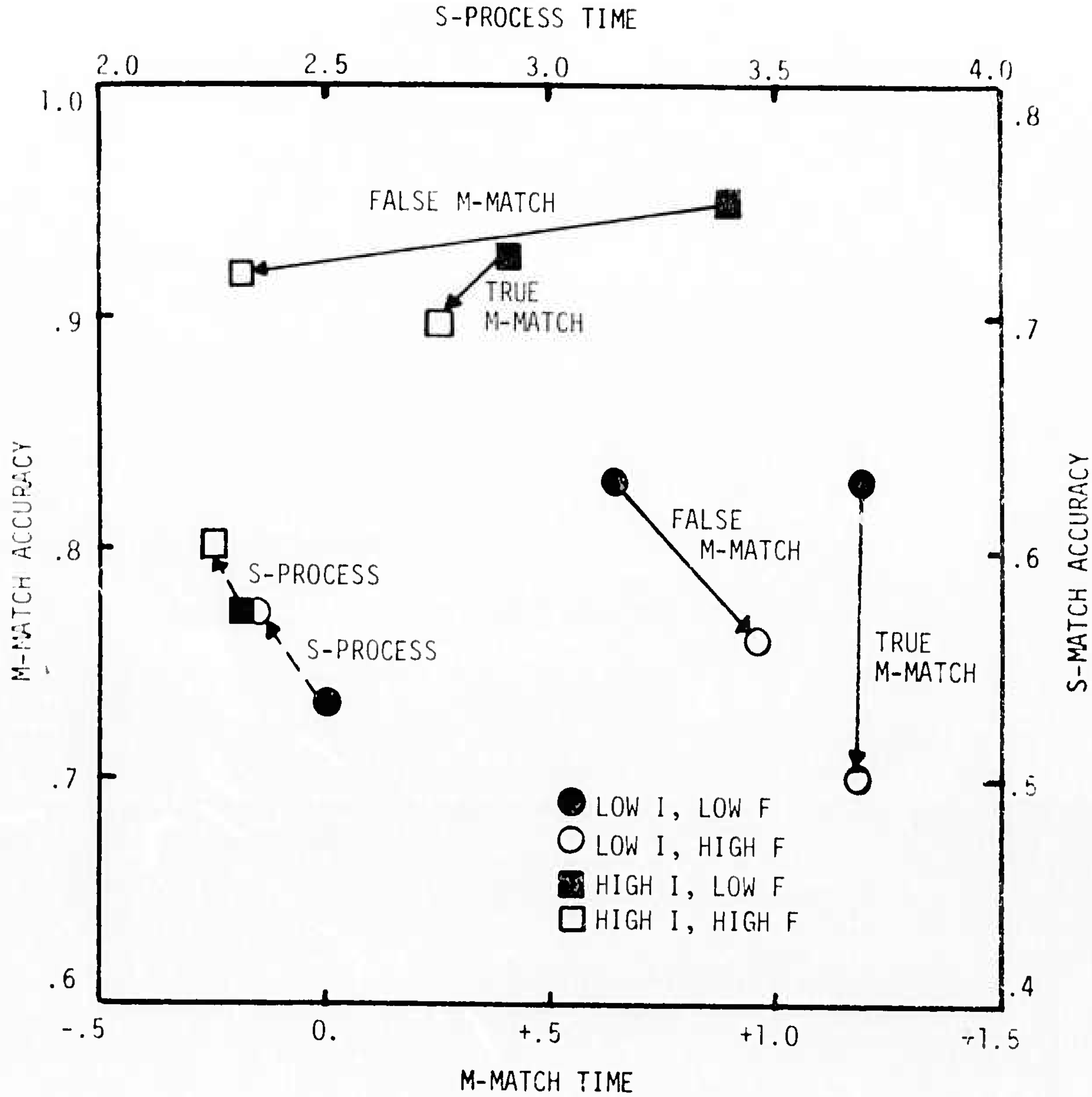

Figure 21. Model parameters for the I × F interaction. There is a definite difference between the effect of F on true and false low-I questions. The M-match process is extremely fast, in fact, the time estimate is negative, under high I. Notice the small effect of F on high-I true M-matches.

performance somewhat, mostly on false questions; the effects of M were much stronger when I was low. F increased study time, but like RC, produced a mixture of effects in verification performance. Both force-pacing and higher F increased surface processing speed, but force-pacing lowered S-match accuracy, unlike higher F. M-match speed and accuracy was degraded by both factors, but the pattern of effects on true and false questions differed for the two factors. Force-pacing degraded both true and false M-matching, but false was hurt more than true. Higher F left true M-match time untouched, injured true M-match accuracy more than false, and degraded false M-matches in both speed and accuracy.

The effects of RC, M, and F showed up more distinctly in the low-I conditions. Changes produced by RC in surface processing depended on I: the speed increase produced by RC was confined mainly to low-I materials; force-pacing produced a severe S-match accuracy loss in low-I conditions, but an improvement in accuracy in high-I conditions.

An important result was the I × F interaction that appeared in both the study time and verification data. In low-I materials, the effect of higher F is to hurt verification accuracy with only the faster surface processing responsible for the small improvement in study and verification time. However, in high-I materials, higher F produces very little loss in verification accuracy, but much faster study and false verification times.

CHAPTER IV

DISCUSSION OF THE RESULTS

The goal of performing the experiment was to determine if the associative word properties and the reading-time manipulation had effects on language processing, and if so, whether the effects are consistent with the hypothesized memory representation of the properties. It is clear from the results that imagery (I), association value (M), frequency (F), and reading condition (RC) all had effects on acquisition and verification subprocesses. The task of this chapter is to account for these effects in terms of the memory network. The effects of RC, M, and F can be reasonably accounted for with the extended HAM model and network representations described in Chapter I. In contrast, the very strong effect of I is not so easily accounted for, although plausible steps toward understanding the effect of I can be made. The remainder of this chapter comprises a discussion of each factor.

A Comment about N

The effects involving sentence complexity (N) are all reasonable; the approximately linear function relating N to study time corresponds to the results obtained by Kintsch and Keenan (1973). However, any detailed theoretical use of this function is not justified since the number-of-propositions measure N is only crudely related to the complexity of the actual HAM structure. The structures of the sentences seem to be disproportionately more complex on the average at higher N; thus the greater effects on study time of I, M, and F at greater N are not very interesting.

In the verification data, the effect of N was unsurprising:

increased question N led to poorer verification performance. This result stands in contrast to the interestingly counterintuitive result of higher recognition confidence at greater N that was obtained by Bransford and Franks (1971) in a verbatim sentence recognition procedure. Since nothing so unusual or interesting appeared in this experiment with regard to N, the verification data were collapsed over N, and N was subsequently ignored.

Reading Condition

The reading time stress manipulation produced a complicated mixture of effects. Force-pacing the reading increased overall verification speed, degraded meaning match performance, and produced changes in surface match accuracy that depended on the level of I.

The most natural explanation for the verification speed increase is that force-paced subjects acquired a set for speed-reading, thereby reducing the time required to read in the question sentences; this time is included in the parameter T, estimated by $\hat{T} + \hat{t}_s$. The fact that the surface-processing speed increase was confined to low-I material as shown in Figure 17 is rather puzzling, pointing to a possible inadequacy of the model, in that supposed surface processing parameters are influenced by things which would have their natural effect on the question comprehension process.

The HAM model assumes that the probability of installing all of the links for a proposition depends on the time available for study (Anderson & Bower, 1973, p. 285). Hence it would be expected that the reduced study time in the force-paced condition would lower accuracy by reducing the likelihood of an intact complete proposition being available at verification time. However, the increase in the M-match times shown in Figure 9 and Figure 17 with I low means that the link ordering is also affected by force-pacing. Thus adequate study time ensures not only that links are installed successfully, but also that they are placed higher up in the search order.

An interesting aspect of the effects of the reading condition (RC) is the larger effect on false M-match times than true. If all false questions were answered by means of an exhaustive search, no effect of RC on false M-match times would be expected at all. However, as discussed in Chapter I, the false questions could also be answered correctly by discovering contradictions. Hence, false M-match times are determined not only by the details of the search processing, but also by the relative proportions of very fast contradiction outcomes of the search, and very slow exhaustive terminations. Thus any factor that influences the proportion of contradiction outcomes will have considerable "leverage" on the false M-match times. Since force-pacing reduced the amount of intact information available during verification, fewer contradictions would be discovered by the force-paced subjects, who would have to rely more often on exhaustive searches to reach a conclusion of false, thereby lengthening the false M-match time.

The way in which S-match accuracy depended on both I and RC as shown in Figure 17 indicates that subjects in the two reading conditions used different storage strategies. The effect of force-pacing was straightforward in the low-I conditions: The difficulty of low-I material means that subjects would have to give priority to meaning storage; thus, the limitation of study time sharply reduced the amount of surface information that subjects were able to store. On the other hand, high-I material is so well comprehended that self-paced subjects did not need to worry about storing the supplementary surface information. However, the force-paced subjects were at a disadvantage with meaning storage, since a long time is required to form an image, and so they devoted more attention to the storage of surface information and produced higher S-match accuracy on high-M material than the self-paced subjects.

M-value

The reason for the high-M materials requiring less study time may be that they are more stereotypical in some sense; inspection of the high-M sentences suggests that they have a certain "hanging-together" quality. In terms of the memory network, this means that some of the propositions connected to high-M words in memory were the same as some of the propositions in the input sentence; thus the new propositional structure could simply reference these old propositions, rather than duplicate them in the new structure. The greater use of pre-existing memory information in high-M materials reduces the amount of new structure that has to be built, thus reducing the required study time. Since more effort can be expended on the new parts of the structure, the acquisition information is more likely to be successfully stored, thereby increasing verification accuracy.

The lack of a change in true M-match time shown in Figure 13 means that there is no effect of M on link order, or the manipulation of M was too weak to affect true verification performance, despite its effect on study times and false verification performance. The solid change in false M-match time is due to a change in the contradiction-exhaustion search proportion, as in the case of RC. It is possible that false high-M questions may be unusually susceptible to contradiction due to being non-stereotypcial, but inspection of the false questions does not support the idea that the false high-M questions are bizarre or tainted with anomaly.

The conclusion to be drawn from the reasonably-sized effect of M on study time and its feeble verification accuracy effect is that M affects a subprocess used only in acquisition. This subprocess can not be the basic sentence comprehension process, since the input sentence has to be comprehended in verification, meaning that M would affect true M-match times comparably to its effect on study times. Hence M must be affecting the key acquisition process, that of constructing the new propositional structures in memory, which is not performed in verification. This conclusion agrees with the way in which high- and low-M materials differ, namely, in the number of pre-existing memory propositions that can be used during the construction as building blocks.

It could be argued that since the observed effects of M appear only in the low-I conditions, the effects of M are due only to slight residual differences in the I-value of the high- and low-M materials. As discussed in the materials section, this residual I-value difference is very small. However, comparison of Figures 14 and 12, showing that true and false questions are affected equally by I, but differently by M, shows that I and M operate in different ways and are therefore distinct variables. As a final argument that the effects of M in this experiment are not simply artifacts produced by I, M has been shown to act independently of I in certain memory studies reviewed by Paivio (1971, Chap. 7).

Frequency

The reduced study time produced by higher F is due partly to the increase in basic access speed, which also underlies the faster verification times, together with an effect similar to that of M. High-F words are connected to concepts that are likely to be connected to many memory propositions; thus it is more likely that there are propositions in memory encoding parts of the input propositions, which would reduce study time as in the case of M. However, the study time benefit is not accompanied by greater verification accuracy, since the concept nodes for high-F words have a greater fan-out of ε^{-1} links connecting the concepts to propositions and thus there is a lower probability of the critical links being included in the search sets. F has no effect on link ordering, since there is little change in M-match time for true questions; a link search order effect would not be expected, since the HAM frequency assumptions apply to conjoint frequency, that is, link usage frequency, and not absolute frequency which was manipulated in this experiment.

I and F interacted in an important way. In the study time results, F facilitated study more if I was high; in verification of high-I questions, the accuracy impairment produced by F was almost eliminated, and a large increase in speed on false questions was produced by high-F. This result is important to understanding the mechanism underlying the effects of I, and will be discussed later.

<u>Imagery</u>

The effect of I on study time resembles results of an experiment by Paivio and Begg (described by Paivio, 1971, p. 442 ff). Subjects read a sentence and then pressed a button when they had either comprehended the sentence, or formed an image, depending on the task set. Comprehension latencies were shorter than image formation latencies, and the imagery value of the material had no effect on comprehension latency, but made a substantial difference in image formation latency. That comprehension of a sentence is faster than image formation is plausible, since comprehension of a sentence would seem to be necessary before an image about its meaning could be formed. These results suggest that in the present experiment, the difference in study times due to I is a result of subjects' attempting to form images of the material, rather than simply comprehending it. Note that subjects were not given any particular instructions regarding the use of imagery.

The effects of the other factors, RC, M, and F, have fairly straightforward network model interpretations. However, even though I was by far the strongest word property manipulated in the experiment, its interpretation in network terms is hardly straightforward. The current network models make little effort to handle imagery effects in language processing, and do not explicitly include the perceptual descriptions discussed in Chapter I as part of the memory information. In fact, most network modelers, such as Anderson and Bower (1973, p. 449 ff), see the whole problem as that of trying to explain language imagery effects strictly in terms of the general, or abstract, conceptual information making up the semantic network. Three notions of how imagery effects could so operate will now be discussed. These hypotheses seek to account for imagery effects without the use of a different type of memory information by assuming that imagery effects are produced by changes in the storage or processing of the same abstract propositional structures used to represent low-I material.

Simple encoding rate change. This hypothesis holds that high-I materials are characterized by a simple increase in the rate of encoding and storage and has received some support (Anderson & Bower, 1973, p.

319 ff). The increase in encoding rate is responsible for the faster study times for high-I material as well as the faster verification times, since the high-I questions are more quickly encoded into probe structures. The improvement in accuracy is a simple consequence of the faster encoding during acquisition, which permits more information to be stored in the available time.

One general problem with the simple rate hypothesis is that it has little explanatory content; preferably, a theory of imagery would explain the faster processing shown on high-I materials.

A more specific difficulty is that although the hypothesis says that the higher accuracy is due to the faster acquisition storage rate, in the self-paced reading condition, shown in Figure 16, high-I accuracy was still superior to low-I accuracy, even though the self-paced subjects spent more time studying the low-I material, which would seem to make up for the slower encoding and storage rate.

The worst problem for the rate hypothesis lies in the interactions of I with other factors in verification latency. If I produces a simple increase in rate, then the patterns of effects of the other factors should be simply translated from one region of the performance plot to another, without any alteration in the patterns. In other words, if I produces a simple processing rate increase, I and the other factors should be additive in their effects on verification latency. The RC × I and I × M interactions in Figures 16 and 18 argue against additivity; but the strongest case is the I × F interaction on verification latency and study time, shown in Figure 20 and Table 10. The fact that the effects of F depend on I in both pattern and degree means that F plays a different role at the two imagery levels. This is inconsistent with the simple rate change hypothesis.

Propositional elaboration. This idea, discussed in another context in Anderson and Bower (1972) and with regard to imagery in Anderson and Bower (1973, p. 460), is that high-I materials allow the subject to elaborate the input by generating extra propositions that improve accuracy by providing redundancy in the memory representation. In recall, the redundant propositions could provide extra retrieval routes, although exactly how they would be used is not clear. However, in order to assist recognition of meaning, the extra propositions would have to contain redundant information about the exact same proposition that is being judged. For example, if the subject elaborated the acquisition sentence The ants ate the jelly with many details about the appearance and other aspects of the ants and the jelly, but did not in some way duplicate the representation of the relation between the ants and the jelly, then the elaborative propositions could not help the recognition of a statement about this relation. The problem with the elaboration hypothesis is that it is not clear that adding details to a representation of a proposition will duplicate the basic propositional information.

It is possible that the elaborative details could, in effect, provide redundancy by supporting inferences that can compensate for missing information. For example, suppose the subject had elaborated the above sentence with a proposition that the ants had six legs, and when presented with the question statement The ants ate the jelly could remember only that something with six legs ate the jelly. By searching memory, the subject would discover that ants have six legs and could thereby conclude that the statement is true of the acquisition events. However, he could also recall that cockroaches have six legs and are at least as likely as ants to be eating jelly; in the absence of any other information, the subject could readily conclude that the question statement is false. The point is that the use of elaborative information to support inferences is an extremely complicated matter (cf. "semantic triangulation" in Anderson & Bower, 1972).

A further problem is that low-I material can also be elaborated if it is reasonably meaningful, suggesting that elaboration per se is not the whole story to imagery. Finally, since the complex elaboration hypothesis has not be given a rigorous statement or tested in a simulation, it is hard to say whether it would indeed work as advertised.

Link-ordering. This hypothesis admits the use of imagery, but holds to the storage of only abstract propositions; the use of mental images simply ensures the superior storage of the propositions. At study time, the subject forms an image which acts as an additional and more vivid presentation of the material. This re-presentation improves the storage of the propositions by causing the ε^{-1} links to be placed at the top of the search lists for each concept used in the propositions. This higher position of the links leads to better accuracy and faster verification times since the links are examined earlier.

However, it would appear that this change in search order would produce different effects on true and false verification times, as did the other factors that influenced the search set; however, the effect of I on true and false performance was essentially equal, as shown in Figure 14. Furthermore, the strong link-ordering effect of I would override other search set influences and render the ε^{-1} fan-out governed by F unimportant. This is hard to reconcile with the I × F latency interaction, although it does fit with the weakened effects of RC and M under high I.

A Dual-representation Model

The above three hypotheses about imagery attempt to retain parsimony by assuming the storage of only abstract propositions. However, this parsimony is purchased at the cost of an unseemly struggle with the data. Furthermore, the best of the lot, the elaboration hypothesis, has neither simulation results nor compelling arguments in support. Under these circumstances, the dual-representation model developed here can not be considered unnecessarily complex, although its operation in network terms can only be outlined.

The model assumes that the meaning of verbal material can be represented by both abstract propositions, and perceptual descriptions of corresponding scenes. High-I materials possess very good representations of both types, whereas low-I materials have only the abstract representation. The model differs from the dual-code model advanced by Paivio (1971, pp. 178-181) in that the abstract representations are propositions, rather than verbal strings, and the images are constructed perceptual descriptions, rather than associated sensory reactions. The I × F interaction observed in both acquisition and verification results is strong support for the dual-representation model in that it strongly suggests that some different mechanisms are at work in the processing of high- and low-I materials.

An implicit true question can be answered more accurately if both representations are present. For a rough calculation, let p_a be the probability of a match success based on the abstract representation along, and p_i the probability of a match based on the perceptual representation alone. Then the probability of a correct answer on an implicit question is

$$P(\text{correct} \mid \text{implicit}) = p_a + p_i - p_a p_i \ .$$

Assume that $p_i = 0$ for low-I material, and p_a is the same for both high- and low-I material. Then $\hat{p}_a = P(\text{correct} \mid \text{implicit low-I})$. From the data for the effect of I on verification accuracy in Figure 14, $\hat{p}_a = .77$ and $\hat{p}_i = .61$. Notice that the value of $\hat{p}_i$ is fairly close to and lower than the value of $\hat{p}_a$, indicating that the superior retention of high-I material is due to the redundancy of the image representation, rather than being due to a vastly superior accuracy of image representations compared to abstract representations.

The M-match time estimated by the descriptive model includes both the time to comprehend the question and the time to carry out the search and decision stages. The comprehension of the question consists of deriving the propositional representation of the sentence meaning, and then forming the perceptual description for a corresponding image. Then both representations are matched to memory in the search stage. The reduced M-match time for high-I material could be the result of faster image formation times in the comprehension stage. This stage should be the same subprocess as that used in acquisition; hence the difference between high- and low-I verification times should be similar to the difference in study times; this similarity does, in fact, appear in the data shown in Table 8 and Figure 14.

An alternative locus for the speed increase produced by higher I is in the search process. If both representations are searched for simultaneously, the low-I questions would be answered more slowly because only the search for the abstract representation could succeed. The problem is that there is about a full second difference out of about 3.5 seconds average latency to be accounted for; it seems likely that two searches with identical completion time distributions could produce such a large difference. Perhaps the completion time distributions are not identical; that is, perhaps perceptual descriptions can be searched for and compared much faster that abstract representations. This explanation of the latency difference seems weak in view of the ease of accounting for the effect by changes in image-formation time during question comprehension.

Since the I × F interaction was used against the abstract-representation models, it should also be applied to the dual-representation model in some detail. In study times, the greater effect of F in high-I material could be due to a facilitation of the image formation process: High-F concrete words correspond to familiar objects; such objects have perceptual descriptions in memory that are more available and have higher quality, making them more suitable for use in image formation than the descriptions attached to low-F concrete words. Thus, high-F high-I material will have shorter image formation times than low-F high-I materials.

The image facilitation account of the I × F interaction in study times is fairly satisfactory; however a problem for the question comprehension interpretation of the I effect arises. If the action of I is on the question comprehension process, then the image facilitation produced by high F would appear equally on true and false verification latencies, since the comprehension process, as it appears in the descriptive model, must be fully performed for both types of question. As is obvious from Figure 20, the speed increase produced by higher F on high-I questions was confined mainly to the false questions, with little effect on true questions. It is possible that this disparity is simply another instance of the leverage mechanism at work on false questions, as in the case of RC and M. The problem is that now the theoretical significance of the I × F latency interaction is impugned; the situation is aggravated by the fact that the I × F accuracy interaction considered by itself could be just another ceiling effect. A more detailed model of how the false questions are answered, together with data from well-specified false questions, is necessary for a more satisfactory interpretation of the I × F interaction. Meanwhile, the difficulties of the question comprehension explanation of I suggests that it is more plausible that I could operate by influencing the search process speed.

In any event, the fact that the verification speed increase produced by F is dependent on I means that F is governing some property of the memory structure that distinguishes high- from low-I material, such as the availability of perceptual descriptions. Further experimental investigation of the relations of I and F to image formation times and verification performance is highly desirable, since interactions of this sort offer good opportunities for confirming the existence of the two types of memory representations.

Summary of the Discussion

The network model interpretations developed above for each factor will now be briefly summarized; a final statement summarizing the entire project is reserved for the next chapter.

Force-pacing produces a speed-reading effect in which the question read-in process operated faster. The lesser study time under force-pacing results in the stored propositional structures being incomplete and with the ε^{-1} links between concepts and propositions being further down in the MATCH process search order, producing slower and less accurate meaning recognition. The larger effect of reading condition (RC) on false questions is due to a leverage effect in which small changes in the accuracy of the stored information influence the relative proportion of the fast contradiction and slow exhaustion match outcomes, producing relatively large changes in response speed and accuracy. RC also produces strategy differences with regard to the storage priority of surface and meaning information.

The word property M (association value) has an acquisition process effect in that words with high M-value permit a greater use of preexisting memory propositions as parts of the new structure, resulting in faster study times and a slight accuracy improvement on true questions. M has a larger effect on false questions because of the leverage mechanism discussed in connection with RC.

The use of high-F (frequency) words enables surface processing to be performed more quickly, but impairs accuracy because of the larger number of competing ε^{-1} links that have to be searched to find the matching propositions. In the case of high-I (imagery) material, higher-F words facilitate image formation since they correspond to more familiar objects with better perceptual representations.

After consideration of some single-representation alternatives, the effects of I were explained in terms of a dual-representation model in which the high-I materials are stored in the form of both propositional and perceptual representations. The verification comparison process consists of match searches for both types of representation. The redundancy of the representations provides for the high accuracy in high-I materials. The greater verification speed either is due to faster image formation during comprehension of the question, as in acquisition, or it is due to perceptual descriptions allowing faster search processing than propositions. The similarity of the study time and verification latency effects argues for the question comprehension locus of the speed increase; the failure of the I × F interaction to appear on true question latency argues for the search time locus. Further work is necessary to resolve the issue.

CHAPTER V

CONCLUDING DISCUSSION

Some Comments on the Experiment

The experiment described in this thesis was a large and complex effort whose value is mainly exploratory and suggestive. A simpler experiment might have missed the interesting and theoretically important interactions; a more elaborate experiment might have given more conclusive results, but only at the cost of even greater cumbersomeness, time, and expense. Hence some of the experimental effort was wasted, while some of the results suggest avenues for further exploration, which will be discussed later.

A particular problem with the experiment is that many of the useful results involve comparing performance on true and false questions. The false questions were not well-controlled, which weakens some of the results. It is generally difficult to devise theoretically informative false questions, since to do so presumes prior knowledge of the processes of answering them. Controlling the number and types of contradictory predications in false questions would reveal whether the contradition/exhaustion mechanism discussed in connection with reading condition (RC), association value (M), and frequency (F) is viable, and more about the processes underlying the I (imagery) × F interaction in verification performance.

Another specific problem is that the amount of materials representing each I, M, and F combination is rather scanty; hence replication with new material would be needed to rule out the possibility that the results are due to idiosyncratic materials. The statistical methods advocated by Clark (1973) would be appropriate, except that treating materials (within conditions) as a random factor commits one to using a large set of materials in each condition, for the same reason that a large number of subjects is required. Furthermore, the counterbalancing schemes would have to be arranged to avoid intractable incomplete or unbalanced designs with regard to the materials factor, which can interfere with other design considerations. Furthermore, composing even the four complete ideas for each condition was extremely difficult; it would be very hard to produce the many more required to support a Clark-style analysis. However, the results are reasonably credible since most of the results appear over a set of materials that vary widely in at least one other factor. For example, the main effect of F appeared in a set of 16 complete ideas for each level of F which varied strongly in I and M. Replication of parts of the experiment would be simpler since M can be ignored without much loss, which would greatly facilitate composing sentences of varying I and F.

Suggestions for Further Work

The strength of the effect of I, and its modification or attenuation of the effects of other factors, suggest that a manipulation of I should be routinely included in any study of language processing. Under low-I certain properties of the material may have noticeable and important effects that simply get swamped under high-I. Furthermore, if different representations or processes are involved in high- and low-I material, the failure to include such a manipulation renders the results ambiguous with regard to what the representations or processes are. Alternatively, if a single-representation model is correct, only imagery manipulations will reveal the common mechanism in processing both types of material. Since imagery has often been shown to be confounded with other manipulations, it would be good insurance to routinely manipulate imagery value.

A final comment on I: The other factors had moderate effects which were explainable in a straightforward manner. On the other hand, I-value had extremely powerful effects, and does not yet have a decent explanation. If the I-value of the material is usually the main determinant of performance in language processing tasks, it should have a correspondingly central place in any theory of memory. Hence, an explanation of imagery effects deserves first priority in the development of memory models.

The verification accuracy decrement produced by higher F parallels the verbal-learning results that previously presented low-F words are more accurately recognized than high-F words. A possible explanation for this result could emerge from developing a propositional interpretation of the word recognition task, along the lines of Anderson and Bower (1973, Chap. 14), incorporating the assumption about high-F words connecting to concepts with higher fan-out.

If the assumption of higher fan-out from high-F concepts is true, it should be possible to ask subjects to describe all the facts they know about concepts, or involving the concepts, and demonstrate a correlation between the amount of such productions and the frequency of the word. A similar task could be used to arrive at a better measure than M-value of the knowledge associated with a word. Since some words are high in F, but low in M, using words differentiated by this technique would allow separate testing of "pure" frequency and meaningfulness effects.

Another alternative to M-value would be the stereotypicality of the material. If the availability of pre-existing memory propositions is really effective in language comprehension, then manipulating the extent to which the material makes contact with common knowledge should produce effects on comprehension. For instance, at this time, material such as *The committee debated impeachment* may be more comprehensible than similar material with higher-frequency words: *The committee talked about new legislation*. The relation of such possible effects to Bransford and Johnson's (1972) context-generation notions remains to be explored.

The failure of the reading time limitation to do anything really dramatic was a disappointment. To some extent, it revealed the ability of subjects to adopt storage strategies depending on the time limitation. The main problem was that the experiment cut the reading time to the same value for both high- and low-imagery material; halving the time separately for the two imagery levels might make it so difficult to form images that the low-imagery material could actually gain some kind of advantage over the high-imagery material; manipulations of this sort are related to the selective interference experiments such as Atwood (1971).

Summary

This thesis undertook the study of the role of three associative word properties, imagery, association value, and frequency, in the processing of sentences in a task involving integration of sentence information followed by a verification task consisting of true-false questions. The questions, together with a descriptive model, provided information about the storage and recognition of sentence information of two types: exact wording (surface) information, and content (meaning) information. The three word properties, and a stress manipulation, had effects on sentence processing which were explained in terms of a network memory model. Each manipulation and its effects can be briefly summarized:

Reading condition. This stress manipulation consisted of allowing subjects to read at their own rate, or forcing them to read at a fast rate. No speed stress was applied during the verification task. Force-paced subjects acquired some speed-reading ability as a result of the fast reading, but were less free to store the supplementary surface information. Limitation of the reading time interfered with both the storage and later verification of meaning information in that force-read propositions were not stored as completely, and were more slowly searched later. Hence adequate study time ensures both the completeness and accessibility of stored propositions.

Frequency. The materials were composed of words of either high- or low-frequency of occurrence in the language. Higher frequency permitted faster reading, but impaired verification accuracy. The faster processing is a result of high-frequency words being more quickly looked-up in memory; the lower accuracy is due to the target proposition being lost in the greater number of propositions connected to high-frequency words.

Association value. The materials used nouns that were either high or low in association value, which is the average number of free associates produced in response to the word. High association-value sentences were studied for less time than low because the larger number of associates means that there is a greater chance that some of the sentence propositions are already represented in memory and do not have to be built from scratch. The slight improvement in verification accuracy is due to the more reliable storage afforded by the re-use of old propositions. Consistent with this analysis, association value had no effect on sentence surface-level processing.

Imagery. The materials contained nouns that were high or low in imagery value, as measured by ratings of the ease of image formation of the referent of the noun. Higher imagery materials were studied for much less time and were much more quickly and accurately verified. Imagery produced differences in surface-form recognition which were attributed to changes in subject strategy. Tentative theoretical conclusions about the nature of imagery effects were drawn. Apparently most consistent with the results is a dual-representation model in which both abstract and image representations are stored and later searched for. The accuracy improvement is due to the redundancy advantage enjoyed by high-imagery material, which possesses both representations, compared to low-imagery materials with only one. The study time decrease is due simply to the faster image representation formation in high-imagery materials. The verification speed improvement is due either to faster image formation prior to search, or to an intrinisically faster processing of image representations during the search. Alternative explanations of the imagery effects by single-representation models were also discussed.

General Conclusion

The effects of word association value and frequency on language processing appear to be representable in a semantic network memory model. Imagery, however, demands further study. The single representation memory models require much more theoretical work in order to explain how imagery could produce the patterns of increased verification speed and accuracy observed in this experiment. The dual-representation model requires additional experimental evidence to support the assertion that two information types, possibly differing in the speed of processing, are represented in memory. In both cases, however, it appears that imagery is explicable in terms of memory network structures.

Thus the associative word properties have effects on language processing and can be represented in the terms of the semantic memory approach, although further work is necessary to clarify the memory structures and processes involved with these word properties.

APPENDIX A

THE COMPLETE IDEAS USED IN THE EXPERIMENT

Practice set:

The gigantic boulder which fell off the cliff crushed the tiny hut at the edge of the woods.

The arrogant appeal for support aroused severe criticism of the leader.

The unrealistic goals in the program caused frequent disillusionment in the government.

The hungry ants in the kitchen ate the sweet jelly which was on the table.

Low I, low M, low F:

The unanticipated flexibility of the arbiter astounded the ambitious originator of the franchise.

The novel hypothesis that elucidated antitoxins corrected antique misconceptions about immunization.

The wicked demon in the allegory symbolized the deadly hatred stemming from jealousy.

The melodious encore executed by the ensemble emphasized the prodigious mastery manifested during the recital.

Low I, low M, high F:

The emotional impact of the speech started the great interest in the enterprise.

The extensive effort by the management improved the low quality of the product.

The detailed investigation of the incident revealed the complete facts about the crises.

The calm advice urging cooperation produced a temporary attitude of agreement.

Low I, high M, low F:

The creeping onslaught of opium weakened the corrupt dynasty which was near chaos.

The chronic malady of rheumatism afflicted the respected theologian who received a blessing.

The deadly infection of malaria slaughtered the starving inhabitants who were numbed by grief.

The loyal retailers who were filled with pep salvaged the mismanaged centennial commemorating the truce.

Low I, high M, high F:

The increased cost of welfare demonstrated the economic necessity for new legislation.

The widespread tragedy of disease created strong pressure for medical research.

The fresh air from the atmosphere contained a high amount of oxygen.

The psychological theory supported by the evidence explained the emotional development leading to crime.

High I, low M, low F:

The obese baron who was a glutton drained the collosal goblet filled by the wench.

The inappropriate joviality of the attendant ruined the solemn procession opening the ceremony.

The squat keg buried in the rubble tripped the handcuffed captive who was weakened by fatigue.

The pallid phantom who haunted the passageway terrified the frail maiden who dropped her bouquet.

High I, low M, high F:

The old professor who taught at the village drew a large circle which was inside the square.

The attractive girl wearing a pink dress conducted a long interview with the poet.

The radical speaker who shouted in the meeting demanded the violent destruction of military headquarters.

The thin person standing in the shadows watched the dark building on the corner.

High I, high M, low F:

The crusty admiral who commanded the vessel braved the stinging spray blasted by the hurricane.

The shrewd prosecutor who was chewing a cigar scrutinized the unidentified corpse which was in the morgue.

The wealthy monarch who was clad in luxurious garments devoured a scarlet lobster which was smothered in sauce.

The filthy lice infesting the beggar shocked the inexperienced physician who administered the vaccination.

High I, high M, high F:

The tired troops who were covered with dust approached the thick forest which was in the valley.

The drunk student who was playing the piano drank the red wine which was in a bottle.

The proud chief who was at the camp refused the thin cattle offered by the army.

The alert doctor who was using his instruments examined the brown blood taken from the sick boy.

APPENDIX B

MEAN LATENCY IN SECONDS (RT) AND PROPORTION CORRECT (PC) FOR EACH COMBINATION OF RC, I, M, F, AND TRUTH VALUE

RC	I M F Combination	FALSE RT	FALSE PC	TRUE RT	TRUE PC
Self-paced	L L L	3.56	.83	3.18	.90
	L L H	3.53	.77	2.92	.82
	L H L	2.71	.94	3.13	.93
	L H H	3.34	.85	2.98	.83
	H L L	2.26	.96	2.30	.95
	H L H	2.15	.93	2.18	.93
	H H L	2.16	.97	2.36	.97
	H H H	2.01	.96	2.27	.96
Force-paced	L L L	3.31	.72	2.77	.82
	L L H	3.10	.66	2.68	.76
	L H L	2.68	.85	2.97	.85
	L H H	2.95	.75	2.59	.78
	H L L	2.38	.92	2.34	.92
	H L H	2.14	.90	2.13	.90
	H H L	2.03	.94	2.17	.94
	H H H	1.93	.91	1.98	.91

Approximate standard errors for these cell means were obtained from error mean squares from an analysis of variance of the untransformed data, using the method of Cochran and Cox (1957, Ch. 7). For the latency means, the standard error is 0.08 seconds for within-subject comparisons, and 0.34 seconds for between-subject comparisons. For the mean proportion correct, the standard error is 0.03 for within-subject comparisons, and 0.03 for between-subject comparisons.

APPENDIX C

MEAN LATENCY (RT) IN SECONDS AND PROPORTION CORRECT (PC) FOR EACH COMBINATION OF RC, I, M, F, N, AND EXPLICITNESS

RC	I M F Combination	N = 2				N = 3			
		Implicit		Explicit		Implicit		Explicit	
		RT	PC	RT	PC	RT	PC	RT	PC
Self-paced	L L L	3.64	.89	2.95	.99	4.47	.83	3.55	.96
	L L H	3.26	.74	2.86	.92	4.45	.65	3.00	.91
	L H L	3.53	.95	3.26	.98	4.14	.86	3.52	.95
	L H H	3.36	.77	2.80	.95	4.02	.69	3.20	.87
	H L L	2.25	.95	2.28	.99	3.07	.92	2.81	.97
	H L H	2.31	.93	2.11	.98	2.86	.90	2.66	.95
	H H L	2.48	.96	2.28	.99	3.28	.95	2.85	.98
	H H H	2.33	.96	2.28	.97	3.10	.94	2.87	.98
Force-paced	L L L	2.77	.82	2.64	.90	3.91	.75	3.06	.84
	L L H	3.14	.69	2.70	.84	3.25	.67	2.85	.80
	L H L	3.52	.82	2.56	.91	3.49	.76	2.88	.87
	L H H	3.22	.73	2.62	.84	3.48	.65	2.72	.86
	H L L	2.48	.88	2.15	.96	3.10	.88	2.67	.97
	H L H	2.30	.88	1.94	.96	2.61	.83	2.64	.95
	H H L	2.27	.95	2.12	.96	2.70	.92	2.59	.96
	H H H	2.00	.91	1.92	.99	2.59	.84	2.43	.93

Approximate standard errors for these cell means were obtained from error mean squares from an analysis of variance of the untransformed data, using the method described by Cochran and Cox (1957, Ch. 7). For the latency data, the standard error is 0.14 seconds for within-subject comparisons and 0.57 seconds for between-subject comparisons. For the proportion correct data, the standard error is 0.02 for within-subject comparisons, and 0.05 for between-subject comparisons.

REFERENCES

Anderson, J.R. FRAN: A simulation model of free recall. In G.H. Bower (Ed.), The psychology of learning and motivati n. Vol. 5. New York: Academic Press, 1972.

Anderson, J.R., & Bower, G.H. On an associative trace for sentence memory. Journal of Verbal Learning and Verbal Behavior, 1971, 10, 673-680.

Anderson, J.R., & Bower, G.H. Configural properties in sentence memory. Journal of Verbal Learning and Verbal Behavior, 1972, 11, 594-605.

Anderson, J.R., & Bower, G.H. Human associative memory. Washington, D.C.: Winston, 1973.

Atwood, G. An experimental study of visual imagination and memory. Cognitive Psychology, 1971, 2, 290-299.

Bransford, J.D., & Franks, J.J. The abstraction of linguistic ideas. Cognitive Psychology, 1971, 2, 331-350.

Bransford, J.D., & Johnson, M.K. Contextual prerequisites for understanding: Some investigations of comprehension and recall. Journal of Verbal Learning and Verbal Behavior, 1972, 11, 717-726.

Bobrow, S.A. Memory for words in sentences. Journal of Verbal Learning and Verbal Behavior, 1970, 9, 363-372.

Clark, H.H. The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. Journal of Verbal Learning and Verbal Behavior, 1973, 12, 335-359.

Cochran, W.G., & Cox, G.M. Experimental designs (2nd ed.). New York: Wiley, 1957.

Coll ns, A.M., & Quillian, M.R. Retrieval time from semantic memory. Journal of Verbal Learning and Verbal Behavior, 1969, 8, 240-247.

Collins, A.M., & Quillian, M.R. How to make a language user. In E. Tulving & W. Donaldson (Eds.), Organization of memory. New York: Academic Press, 1972. Pp. 310-351.

Conrad, C. Cognitive economy in semantic memory. Journal of Experimental Psychology, 1972, 92, 149-154.

Cramer, P. Word a sociation. New York: Academic Press, 1968.

Deese, J. On the structure of associative meaning. Psychological Review, 1962, 69, 161-175.

Fillmore, C.J. The case for case. In E. Bach and R.T. Harms (Eds.), Universals in linguistic theory. New York: Holt, Rinehart, & Winston, 1968. Pp. 1-10.

Forster, K.I., & Chambers, S.M. Lexical access and naming time. Journal of Verbal Learning and Verbal Behavior, 1973, 12, 627-635.

Franks, J.J., & Bransford, J.D. The acquisition of abstract ideas. Journal of Verbal Learning and Verbal Behavior, 1972, 11, 311-315.

Frijda, N.H. Simulation of human long-term memory. Psychological Bulletin, 1972, 77, 1-31.

Humphrey, G. Thinking. New York: Wiley, 1963.

Katz, J.J., & Fodor, J.A. The structure of a semantic theory. Language, 1963, 39, 170-210.

Kieras, D. A general experiment programming system for the IBM 1800. Behavior Research Methods and Instrumentation, 1973, Vol. 5(2), 235-239.

Kintsch, W. Models for free recall and recognition. In D.A. Norman (Ed.), Models of human memory. New York: Academic Press, 1970. Pp. 333-374.

Kintsch, W. Notes on the structure of semantic memory. In E. Tulving & W. Donaldson (Eds.), Organization of memory. New York: Academic Press, 1972. Pp. 249-308.

Kintsch, W., & Keenan, J. Reading rate and retention as a function of the number of propositions in the base structure of sentences. Cognitive Psychology, 1973, 5, 257-274.

Kintsch, W., & Monk, D. Storage of complex information in memory: Some implications of the speed with which inferences can be made. Journal of Experimental Psychology, 1972, 94, 25-32.

Kučera, H., & Francis, W.N. Computational analysis of present-day American English. Providence, Rhode Island: Brown University Press, 1967.

Minsky, M.L. Steps toward artificial intelligence. In E.A. Feigenbaum & J. Feldman (Eds.), Computers and thought. New York: McGraw-Hill, 1963. Pp. 406-456.

Pachella, R.G. The interpretation of reaction time in information processing research. Human Performance Center Technical Report No. 45, University of Michigan, November, 1973.

Paivio, A. Imagery and verbal processes. New York: Holt, 1971.

Paivio, A., Yuille, J.C., & Madigan, S. Concreteness, imagery and meaningfulness values for 925 nouns. Journal of Experimental Psychology Monograph Supplement, 1968, 76 (1, Pt. 2).

Pylyshyn, Z.W. What the mind's eye tells the mind's brain: A critique of mental imagery. Psychological Bulletin, 1973, 80, 1-23.

Quillian, M.R. Semantic memory. In M. Minsky (Ed.), Semantic information processing. Cambridge, Mass.: M.I.T. Press, 1968. Pp. 216-270.

Quillian, M.R. The teachable language comprehender. Communications of the ACM, 1969, 12, 459-476.

Rumelhart, D.E., Lindsay, P.H., & Norman, D.A. A process model for long-term memory. In E. Tulving & W. Donaldson (Eds.), Organization of memory. New York: Academic Press, 1972.

Schank, R.C. Conceptual dependency: A theory of natural language understanding. Cognitive Psychology, 1972, 3, 552-631.

Wilkins, A.J. Conjoint frequency, category size, and categorization time. Journal of Verbal Learning and Verbal Behavior, 1971, 10, 382-385.

Winer, B.J. Statistical principles in experimental design. New York: McGraw-Hill, 1962.

Woods, W.A. Transition network grammars for natural language analysis. Communications of the ACM, 1970, 13, 591-606.

RECENT TECHNICAL REPORTS

40. Malin, J. T. An analysis of strategies for solving certain substitution problems. April 1973.
41. Robinson, J. K. Storage and retrieval processes under conditions of proactive interference. April 1973.
42. Mayer, R. E. Acquisition and resilience under test stress of structurally different problem solving procedures. May 1973.
43. Jagacinski, R. J. Describing multidimensional stimuli by the method of adjustment. June 1973.
44. Egan, D. & Greeno, J. G. Theory of rule induction: Knowledge acquired in concept learning, serial pattern learning, and problem solving. August 1973.
45. Pachella, R. G. The interpretation of reaction time in information processing research. November 1973.
46. Rose, A. M. Human information processing: An assessment and research battery. January 1974.
47. Polzella, D. J. The effect of sleep-deprivation on short-term memory. April 1974.
48. Pew, R. W. Human perceptual-motor performance. August 1974.
49. Glenberg, A. Retrieval factors and the lag effect. August 1974.
50. Eft, D. R. The buildup and release of proactive interference. August 1974.
51. Wickens, C. D. The effects of time sharing on the performance of information processing tasks: A feedback control analysis. August 1974.
52. Kieras, D. Analysis of the effects of word properties and limited reading time in a sentence comprehension and verification task. August 1974.
53. Whitten, W. B. Retrieval depth and retrieval component processes: A levels-of-processing interpretation of learning during retrieval. August 1974.

RECENT MEMORANDUM REPORTS

14. Halff, H. M. The differential effects of stimulus presentation during error- and success-feedback intervals in concept identification. August 1971.
15. Mayer, R. E. Dimensions of learning to solve problems. February 1972.
16. Martin, E. Serial learning theory. February 1973.
17. Pachella, R. G. The effect of set on the tachistoscopic recognition of pictures. October 1973.