ADULT AGE DIFFERENCES IN MEMORY FOR METAPHORIC SENTENCES: THE ROLE OF IMAGERY AND SEMANTIC FEATURE MODELS

by

DIANE PATRICIA FOX
B.Sc., University of Calgary, 1986

A THESIS SUBMITTED IN PARTIAL FULFILLMENT

OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF ARTS

in the Department

of

Psychology

We accept this thesis as conforming
to the required standard

[redacted]

Dr. Roger A. Dixon

[redacted]

Dr. Michael A. Hunter

[redacted]

Dr. Paul M. Baker

[redacted]

Dr. Joseph F. Kess

Supervisor: Professor Roger A. Dixon

ABSTRACT

On the basis of theories of metaphor processing, two experiments were conducted on the retention of metaphoric sentences by young and old adults. Experiment 1 manipulated encoding by requiring 40 young and 40 old adults to process metaphors in one of four conditions: (a) rating them on an imageability dimension, (b) rating them on a semantic relatedness dimension, (c) rating them in terms of number of alternative interpretations, or (d) learning them intentionally. Retention was then measured at three levels of retrieval support: (a) free recall, (b) cued recall (topic and vehicle cues), and (c) recognition, in a within-subjects design. The results indicated that, in the free recall task, the intentional group recalled more metaphors than the interpretations and imagery conditions, and recall in the three incidental conditions did not differ. Recall improved significantly with increasing levels of retrieval support. Vehicle cues facilitated memory performance more than topic cues. The young remembered more than the old adults at all levels of retrieval support, and the age by recall task interaction indicated that the difference in their performance was greatest with cued recall, less for free recall, and was minimal with the recognition test.

Further, the old improved proportionally more than the young adults across the three retention tasks.

The second experiment comprised two tasks to test the utility of semantic feature, and imagery, models, respectively, with a sample of 20 young and 20 old adults. In the first task, metaphoric noun pairs were presented in or out of sentence context. Participants rated the similarity of the topic and vehicle terms in each noun pair and generated adjectives (features) the two nouns in each pair shared. Free recall of the noun pairs was then tested. The results revealed support for semantic feature theory in the: (a) significant and positive correlation between the rating and feature generation data, and (b) higher ratings and more features produced for high semantic relatedness noun pairs as compared to low semantic relatedness noun pairs. Furthermore, although similarity ratings were not affected by age, the young and old adults generated different shared features for the noun pairs, suggesting that the two age groups may perceive the same degree of similarity between two nouns, but the actual basis of this similarity may be different. The young recalled more noun pairs than the old adults, and the effect of context was not significant in any analysis.

In the second task, high and low imagery metaphors were auditorily presented with or without instructions to image to facilitate memory performance. Free recall, cued recall

(topic and vehicle cues), and recognition tests were administered in a within subjects design. The findings demonstrated no effect of imagery instructions on recall. Support for imagery was obtained in the superior recall of high imagery sentences when compared to low imagery sentences, and in the interaction between sentence type and instructions. The recall results indicated analogous effects of age, cue type, and recall task to those obtained in the first experiment.

The empirical outcomes of these experiments were discussed with reference to considerations of context, individual differences, the better operational definition of constructs, and the implications for cognitive theories of information processing.

Examiners:

[redacted]

Dr. Roger A. Dixon

[redacted]

Dr. Michael A. Hunter

[redacted]

Dr. Paul M. Baker

[redacted]

Dr. Joseph F. Kess

TABLE OF CONTENTS

CHAPTER Page

CHAPTER Page

LIST OF TABLES

LIST OF FIGURES

Figure

ACKNOWLEDGEMENTS

I would like to thank my supervisor, Dr. Roger Dixon, for his support and invaluable advice during all stages of preparation of this manuscript, and Dr. Michael Hunter and Dr. Paul Baker for their assistance as committee members. I am also grateful for the financial support provided me over the past two years by the Natural Sciences and Engineering Research Council of Canada. Thanks are also owed to Dr. Albert Katz, at the University of Western Ontario, who supplied the normative data upon which the stimuli for these experiments were selected. In addition, this research was supported in part by a grant from the Natural Sciences and Engineering Research Council of Canada to Dr. Dixon.

Thanks go also to the participants here in Victoria without whom these studies would not be possible.

To Wayne, Odette, Lori, Diana, Jack, and the "DT crowd", I want to say a heartfelt "thank you" for all the emotional support, and for the much needed diversions: the walks, talks, and generally crazy comraderie!

Finally, I would like to express special thanks to my family for their continual encouragement, support, and love.

CHAPTER 1

INTRODUCTION

Once thought of as a "mere" stylistic device designed to enhance the aesthetic value of speech and literature, metaphor is now recognized by psychologists and those interested in language and cognition as worthy of investigation (for reviews, see Billow, 1977; Ortony, Reynolds, & Arter, 1978; Tourangeau & Sternberg, 1982). In fact, some scholars have proposed that analogic and metaphoric reasoning form the basis of all cognition (Cassirer, 1953; Edie, 1963). Others have argued that the study of natural language lacks ecological validity if it does not encompass both literal and nonliteral aspects (Ortony et al., 1978).

This latter assertion is reflected by the frequency of figurative language use in everyday discourse. For instance, Pollio, Barlow, Fine, and Pollio (1977) suggest that, on the average, approximately four figures of speech are uttered per speaking minute in free discourse. If this includes both novel and common (frozen) forms of figurative expression, about 21 million figures of speech per lifetime are implicated (Hoffman & Honeck, 1980). More specifically, it has been estimated that, on average, approximately five metaphors are produced for every 100 words of ordinary speech (Johnson, 1975).

Given the ubiquity of metaphorical speech, interest in understanding how figurative language is understood has led

to the formulation and empirical investigation of several relevant theories. Each of these views postulates a different relationship between the elements comprising a metaphor. Employing the now-conventional terminology developed by Richards (1936), the subject term is named the topic, and the predicate term- that which is being used metaphorically- is referred to as the vehicle. The relationship that the two terms share is designated as the ground. Given these distinctions, three general classes of metaphor theories- the comparison, anomaly, and interaction views- will now briefly be explicated.

Throughout history, probably the most widespread view of the nature of metaphor has premised a comparison between objects that are literally incompatible. That is, the topic is assumed to bear a similarity to the vehicle, with the degree of similarity conceptualized as a continuum from literal to figurative to anomalous utterances. High levels of similarity characterize literal constructions, intermediate levels constitute figurative comparisons, and low levels are descriptive of anomalous sentences.

This comparison view originated with Aristotle (1927) (cited in Tourangeau & Sternberg, 1982), who proposed that the ground of a metaphor was constructed with an analogy involving both terms, or was comprised of a category to which both the topic and vehicle belonged. Today, the latter supposition is embodied in the semantic feature theory,

which states that the basis of a metaphor is expressed by a set of features shared by both the topic and vehicle terms (Johnson & Malgady, 1979; Malgady & Johnson, 1976). This theory will be elaborated in Chapter 2, but suffice it here to establish that it originates in a comparison view of metaphor.

Unlike the comparison view, the anomaly model (Campbell, 1975) emphasizes the dissimilarities between the topic and the vehicle. It predicts that metaphoricity is inversely related to topic-vehicle similarity. Dissimilarity is believed to create complexity, incongruity, and novelty. Accounting for figurative language thus requires that the "normal" processes of interpretation be augmented.

Lastly, the interactionist view (Black, 1962) recognizes both similarity and dissimilarity between the topic and vehicle, and posits that a curvilinear function relates similarity and metaphoricity. In this model, the topic and vehicle interact to produce a resultant meaning that is new and transcends both. This view has been elaborated more recently, and is referred to as the "domains-interaction" view (Tourangeau & Sternberg, 1981, 1982). It espouses that metaphors involve perceiving something in one domain in terms of something in a second domain, with a consequent change in the view of both domains.

These theoretical models were recently tested by

Marschark, Katz, and Paivio (1983) in two rating studies that examined the efficacy of 10 hypothesized dimensions of metaphor comprehension. Each dimension represented a variable intimated as being relevant to metaphor processing by at least one class of psychological model. Metaphors of the form "An A is a B" were utilized in the ratings. The results revealed that the 10 dimensions were highly intercorrelated, and support was obtained for each major metaphor model, but none received unequivocal confirmation. These findings were replicated by Katz, Paivio, & Marschark (1985) employing more ecologically valid poetic metaphors.

The study of these attribute dimensions was extended to the memory domain in a series of three experiments conducted by Marschark and Hunt (1985). They found that the dimensions of rated number of interpretations and imageability of the subject were the best two predictors of recall, and concluded that variables previously identified as important to metaphor comprehension were probably not analogous to those underlying memory for metaphors.

The investigation of memory and metaphor is also of particular interest in the cognitive aging domain. This is because it is congruent with the current emphasis on identifying areas where the performance of the elderly does not decline, but is maintained at a relatively high level (e.g., studies of expertise and compensation). In this regard, it has been speculated by some cognitive aging

theorists that metaphor comprehension may represent a type of processing that does not deteriorate with age, but is a specialized skill that develops over time. Interest has been directed at discovering what mechanisms are responsible for this continued competence, and the extent to which they will generalize across other activities. For example, Boswell (1979) asked young and old adults to provide explanations of four metaphors, and found that the older adults offered more poetic or synthesizing explanations compared with the literal or analytic interpretations of the young. Boswell concluded from these results that the qualitative development of metaphoric comprehension continues throughout the lifespan.

In addition, Kausler (1982) has used the term "functional reasoning", or reasoning through the use of analogy, to describe the mental processes involved in interpreting the meaning of proverbs and metaphors. He asserted that because metaphors are a common form of expression, facility in thinking by analogy has direct implications for our communicative skills. After a brief review of studies on metaphor, Kausler concluded that "functional reasoning is maintained at a high level of proficiency in late adulthood" (p. 552).

The age variable was examined by Backman and Dixon (1987), who manipulated the number of alternative interpretations dimension in a study of memory for

metaphoric and nonmetaphoric sentences in young and old adults. Retention was tested with both free and cued recall tasks. Half of the sentences had been categorized as having few interpretations, and half as having many interpretations. An orienting task required participants to rate each sentence for the number of alternative interpretations possible. Free recall of the sentences was followed by cued recall with both topic and vehicle cues. The results indicated that, overall, the young adults recalled more sentences than the old adults, and females remembered more than males. In the free recall situation, age differences were more apparent for information containing many interpretations than for information suggesting few interpretations. In the cued recall task with a strict scoring criterion, the old adults improved more than the young adults with information having many interpretations than with information connoting few interpretations.

Given this background, the present research represented an extension of this work by examining age differences in memory for metaphors with a subset of three of the 10 dimensions- number of interpretations (INTS), imagery (IMG), and semantic relatedness (SRL). The INTS dimension was included in an attempt to replicate earlier results (Backman & Dixon, 1987; Marschark & Hunt, 1985). The IMG and SRL dimensions were incorporated to investigate the roles of

imagery and semantic feature models, respectively, in memory for metaphors.

Specifically, two experiments were conducted. The first experiment examined age differences in memory for metaphors by utilizing traditional memory tasks to manipulate different stages of information processing. In this regard, orienting condition, cuing, and retrieval support were varied. The second study utilized data from the first to experimentally manipulate the dimensions of imagery and semantic relatedness to provide a more robust test of the two theories with respect to memory for metaphor. Thus, the former experiment was perceived as an extension of the memory domain to figurative language, and the latter as a validation of two current theoretical models of metaphor processing.

In summary, the purposes of this investigation were fourfold: (1) to provide evidence for the replicability of previous research with dimensions of metaphor processing, and to extend their utility to older adults (Backman & Dixon, 1987; Katz et al., 1985; Marschark & Hunt, 1985; Marschark et al., 1983), (2) to expand, through various manipulations of encoding and retrieval conditions, the methodology of previous memory research in this area (e.g., Marschark & Hunt, 1985), (3) to examine the roles of imagery and semantic feature theories in memory for metaphor by manipulating the imagery and semantic relatedness

dimensions, respectively, and (4) to compare the memory performance of young and old adults utilizing metaphoric sentences (Backman & Dixon, 1987).

The literature pertinent to this investigation is presented in the following chapter. Chapters 3 and 4 then describe the methodology and results of Experiments 1 and 2, respectively. The final chapter elaborates on the findings and discusses related issues.

CHAPTER 2

LITERATURE REVIEW

This chapter is organized into four major sections. The first two sections discuss the literature relevant to the first experiment conducted here, and the last two are applicable to the second experiment. Beginning with the first experiment, a brief review of the stages of information processing will now be addressed.

The investigation of memory performance is often conceptualized with respect to three fundamental stages of information processing (Klatzky, 1980). The first - encoding - involves placing information into a system, and may include the modification of this information. Generally, the faster the input and organization of information, the more efficient the thought process at this initial stage. Information is then retained (i.e., stored) in memory following encoding, and is potentially available for further processing. However, this stored information may be affected by the input of subsequent information, or it may become lost. Retrieval represents the last stage of processing, and refers to how efficiently and flexibly the stored information can be accessed. These three components of retention are interdependent, although they are often manipulated individually in experimental contexts. As will be elaborated in the following sections, various encoding processes can operate on the input information to prepare it

for storage in long-term memory. Further, the organizational demands of the recall test can be manipulated to examine the process of retrieval. These two stages - encoding and retrieval - will now be elaborated in greater detail.

Encoding

The levels-of-processing framework (Craik & Lockhart, 1972) is perhaps the most popular current approach to the study of memory decline with advancing age. It postulates that recall is a function of the processing operation performed on the task material at the time of encoding. The various ways of encoding incoming information lie on a continuum from shallow to deep processing, with structural analyses at the shallow end and semantic analyses at the deep end. Deeper processing produces a richer or more elaborated memory trace (Craik & Tulving, 1975) which is more resistant to forgetting. According to this view, older adults may not spontaneously encode items semantically, leading to poorer recall independent of other problems of retrieval (Craik & Simon, 1980).

It has been argued that the encoding of information may be controlled by the use of incidental orienting conditions. That is, subjects perform a task that manipulates the depth of processing without knowledge of a subsequent memory test. It is expected that orienting tasks requiring shallow (structural) processing should result in poor memory performance; those requiring deep processing should produce

better memory performance.

Eysenck (1974) tested this hypothesis in young and old adults by varying incidental orienting tasks that necessitated the generation of either relatively shallow responses (counting letters or producing rhymes) or deeper responses (constructing appropriate adjectives or images) to word stimuli. Analysis of the free recall data indicated that recall was superior for the semantic processing tasks as compared to the nonsemantic orienting tasks, and there was a significant interaction between age and orienting task- the inferiority of the older adults was greater when the orienting task required more semantic processing of the material. This age by task interaction has been replicated elsewhere (Erber, Herman, & Botwinick, 1980; Mason, 1979; White, cited in Craik, 1977).

Encoding has also been manipulated by contrasting two learning conditions- intentional and incidental. Several memory studies have found a superiority of intentional as compared to incidental learning conditions (e.g., Craik & Tulving, 1975; Hyde & Jenkins, 1969; Till & Jenkins, 1973) indicating that intent to learn is a critical factor in memory performance.

Age differences in this variable have been considered by Kausler and Hakami (1983) who examined intentionality in memory for topics of conversation in young and old adults. Participants received a series of 12 topics for discussion

with three accompanying questions per topic, followed by free recall of the topics and a recognition test of the questions posed during the prior conversations. The instructions were varied such that half of the subjects in each age group were informed of the subsequent recall test (intentional memory); the remaining participants were not forewarned (incidental memory). The results indicated no effect of instructions, and the age by instructions interaction was not significant. Kausler and Hakami concluded that memory for conversational information was as adept incidentally as intentionally for both young and elderly adults.

Intentionality was also manipulated by Backman, Mantyla, and Erngrund (1984) in a study measuring the recall of 30 Swedish nouns varying age and the number of self-generated retrieval cues. It was found that intentional learning instructions produced a higher level of recall performance than incidental learning instructions, but, as in the Kausler and Hakami study just reported, this variable did not interact significantly with age. Thus, there is some tentative evidence that intentional learning instructions may not differentially affect the memory performance of young and old adults.

Research utilizing text processing, however, provides conflicting results. Simon, Dixon, Nowak, and Hultsch (1982) presented young, middle-aged, and old adults with a 500 word

narrative and asked them to perform one of four text reading tasks: syntactic, stylistic, advice, or intentional. Analysis of the written recall indicated a significant interaction of age with orienting task. Younger adults recalled the most propositions when recall was intentional or preceded by a semantic (deep) orienting task; recall was lowest in the syntactic condition. On the contrary, middle-aged and older adults performed better under the intentional learning condition than under the three incidental task conditions.

This research was extended in a cross-cultural replication with a sample of three German adult age groups (Dixon & von Eye, 1984). Utilizing the same procedures as in the above study, it was found that performance decreased across the three age groups with increasing age. In addition, subjects recalled less in the shallow condition than in the two deep conditions, which were inferior to performance in the intentional condition. The age by orienting task interaction was not significant here (as it was in the Simon et al. (1982) study), suggesting that, in general, simple instructions to learn the material will facilitate performance the most for both young and old adults.

More specifically, with respect to figurative language, encoding was directly manipulated by Marschark and Hunt (1985, Exp. 2) who presented young adults with 36 metaphors

and asked them to perform one of three orienting tasks on each metaphor: (1) rate the number of interpretations that are possible, (2) write as many interpretations as conceivable, or (3) write a single interpretation. Incidental recall of the metaphors was then requested, and the results indicated no effect of orienting task on recall.

This result contrasts with the findings of Honeck (1973) using proverbs. Students learned a series of proverbs, each accompanied by either its repetition, a grammatical alteration, a paraphrase, or a control sentence. The students attempted to recall the proverbs given the subject nouns as cues. The best recall was attained when items had been learned in the context of their paraphrases compared with the other contexts, indicating that orienting condition had an effect on recall.

Retrieval

As intimated above, there is considerable evidence indicating that older persons typically recall less information than younger persons on memory tasks requiring the reproduction of recently presented material. An alternative view of this memory deficit is that retrieval processes, and not storage factors, are responsible. In this regard, Tulving (1974) has made the distinction between trace-dependent and cue-dependent forgetting. The former refers to information retained in the memory store as a result of the original encoding of the event. Forgetting in

this instance occurs because the trace deteriorates or is lost from the store and the information becomes unavailable. The latter alludes to information present in the individual's cognitive environment at the time retrieval occurs. When the relevant retrieval information is absent, forgetting occurs because the trace information available in the store becomes inaccessible. This dichotomy of processes has been applied to the aging literature, and many investigators have postulated that older adults fail to recall as well as younger adults because they do not retrieve as efficiently. Specifically, memory information is inaccessible because it cannot be located at the time of recall.

An experimental technique that is employed to measure cue-dependent forgetting is the comparison of free (uncued) and cued recall. If forgetting is cue-dependent, it is expected that providing the cues will facilitate recall. The results of studies investigating age and cuing simultaneously are inconclusive. For example, Hultsch (1975) tested recall of four words from each of 10 categories and provided half of the subjects with category names at recall. Contrary to prediction, there were age differences in both the cued and the uncued conditions, and the provision of category names improved performance for both young and old adults. Smith (1977) has corroborated this latter result with evidence that category cues presented only at recall

benefitted all age groups equally.

Laurence (1967) also compared the recall performance of young and old subjects with and without category cues. In the uncued condition, older adults recalled less than the younger subjects; with the category cues at recall both groups performed equally well. Laurence concluded that supplying a cue at recall was beneficial in overcoming a retrieval deficit in older adults.

Category labels were used as retrieval cues in another study in which old and young participants were required to learn a set of items by searching repeatedly through a list of words to identify items specified by their category labels (Macht & Buschke, 1984). Free recall was followed by cued recall in this within-subjects design. The results were essentially identical to those of Hultsch (1975). That is, the aged recalled fewer items than the young, cued recall was superior to free recall, and the interaction between age and type of recall was not significant.

In an extension of these ideas, several investigators have examined the possible interactions between various encoding operations and retrieval processes utilizing the levels-of-processing framework previously discussed. The rationale here follows from the encoding specificity principle of Tulving and Thomson (1973) which states that, in order for a cue to be an adequate retrieval cue, it must be encoded with the item at the time of input.

Fisher and Craik (1977) tested this idea in a series of three experiments that manipulated the qualitative nature of both the encoding and retrieval cues using rhyme (shallow), category (deep), and sentence (deep) cues in Experiment 1, and rhyme (shallow) and associate (deep) cues in Experiments 2 and 3. It was found that the recall of words was highest when conditions at encoding and retrieval were compatible. Further, this compatibility was more beneficial with semantic encodings than with phonemic (rhyme) encodings. Fisher and Craik suggested that this latter result is indicative of the distinctiveness and discriminability of semantic encodings from other memory traces. That is, they are relatively unique and contain more specific information than shallow encodings which comprise more general shared information. The authors concluded that both the levels-of-processing and encoding specificity conceptions were required to explain their results.

Simon (1979) investigated these notions manipulating age in three experiments in which processing at encoding was guided by informing subjects prior to list presentation of the nature of the forthcoming retrieval cue, and then comparing the relative effectiveness of different retrieval cues. She speculated that if older adults did process less deeply, shallow cues would be more effective than deeper semantic cues.

In Experiment 1, Simon attempted to influence depth of

processing by telling subjects that they would be cued for recall with phonemic cues (the first two letters of the target words), semantic cues (synonyms), or no cues at all (free recall). It was found that subjects of all ages recalled more words when given a phonemic cue compared with free recall, but only young subjects benefitted from semantic cues. In the second experiment, a more semantic encoding was encouraged by embedding the words to be learned in sentences, and requesting subjects to read the sentences silently to study their specific meanings. The cues in this instance were either phonemes, synonyms, or a re-presentation of the sentence context with the target word missing. The results indicated that synonym and sentence context cues were beneficial for young adults, but not for the middle-aged and old adults, while phonemic cues aided these latter two age groups, but not the young. Finally, in Experiment 3, the results of Experiment 2 with older adults were replicated in young adults by reducing the presentation time of each sentence from 8 seconds to 4 seconds. Simon suggested that these experiments demonstrated the greater use of superficial memory codes by the old when compared with the young. However, this conclusion has not received unanimous support, as neither Drachman and Leavitt (1972) nor Smith (1977) found greater benefits for old participants than for young participants with first letter (shallow) cues.

White (cited in Craik, 1977) manipulated orienting task by requiring young and old subjects to make either "case", "rhyme", or "category" decisions about common nouns or simply to "learn" the words. On a subsequent free recall test, substantial age decrements for both the intentional learning and semantic orienting tasks were found. However, in an additional recognition task for the same words, the age decrement remained for the intentionally learned words, but was eliminated for words processed in the semantic condition. White concluded that the age decrement in retention was minimized, and perhaps even eliminated, under optimal input (i.e., an appropriate orienting task) and retrieval (i.e., recognition) conditions.

Other studies (Lauer, 1975, cited in Craik & Simon, 1980; Zelinski, Walsh, & Thompson, 1978) concur with this observation. However, what constitutes these optimal conditions appears to vary across studies. For example, Till and Walsh (1980) found that age differences in the recall of implicational sentences could be eliminated by using cued recall following a comprehension response task at acquisition that required subjects to write a word reflecting their understanding of each sentence's meaning.

Perlmutter (1978) provided further evidence that the age decrement could be eliminated in an experiment that examined age differences in the recall and recognition of words after intentional learning instructions or an

incidental semantic orienting task (generating free associations to each word). She reported that young and old adults performed equally well when recognition was combined with a semantic orienting task. This result was replicated in a later study (Perlmutter, 1979) demonstrating that age differences between 20- and 60- year-olds on an associative orienting task were eliminated under recognition conditions, but not for free or cued recall. Contradictory findings have been reported by Mason (1979), who obtained the greatest age differences in both recognition and recall following a semantic orienting task. However, one difference between these two studies is that Perlmutter utilized a within-subjects design, while Mason employed a between-subjects design.

Thus, it can be concluded from these studies that age differences in retention can be eliminated under certain optimal encoding and retrieval conditions.

Returning to considerations of metaphor processing, several studies manipulating retrieval support have been conducted. For instance, Harris (1979a) examined memory for sentences containing either novel metaphors, dead metaphors (those that are technically metaphorical but have come into common usage), or nonmetaphorical expressions. In one experiment, the initial noun phrase of each sentence served as a cue for recall of the sentences; another experiment employed a forced-choice recognition task with the same

sentences. Harris found that the three sentence types did not differ in their probability of being remembered using either a recall or a recognition measure, and concluded that metaphors and their literal equivalents are analogous in memory difficulty. This conclusion was further supported by Harris, Lahey, and Marsalek (1980, Exp. 1).

In a series of three experiments, Harris (1979b) extended this work to memory for Shakespearian metaphors expressed in sentences or in the context of play synopses. Metaphors were contrasted with nonmetaphors expressing the same idea. Participants listened to the stimuli and were instructed that they would be asked questions about them later. The forced-choice recognition test that followed consisted of two metaphors and two nonmetaphors as response choices for each item. The results across all three experiments were equivocal: metaphors were remembered better than nonmetaphors, and there was a greater number of false recognitions for more metaphorical than less metaphorical distractors. Harris, in accounting for the discrepant results obtained here as compared to the two studies reported above (i.e., Harris, 1979a), suggested that listeners may possess a different response set for Shakespearian language than for literal language.

Contrary evidence for the superiority of metaphors over nonmetaphors in memory performance has also been obtained (Malgady & Johnson, 1977). In this experiment, subjects

studied literal, figurative, and anomalous sentences, and were told they would be queried about the sentences. In the yes-no recognition test that followed, subjects indicated whether they had read the sentence before, and expressed their confidence in making the judgment. It was found that recognition decreased linearly with the figurativeness of the sentences, and was lowest for figurative sentences. Further, the false alarm rate was highest for metaphors. This latter finding was commensurate with that of Harris (1979b), and Malgady and Johnson (1977) suggest it is likely due to the fact that metaphors involve an additional semantic domain to falsely associate information with.

The finding that figurative sentences were remembered least often is contrary to the results of Harris (1979b), who demonstrated that metaphors were recognized more often than nonmetaphors. It is not certain why this should be the case, except perhaps that the forced-choice recognition test employed by Harris was easier, or biased response choices more toward the selection of a metaphorical alternative, than the simple yes-no recognition task utilized in the present study.

In a slightly different context, Reynolds and Schwartz (1983) investigated the role of metaphors versus their literal paraphrases as concluding statements to short prose passages. Subjects read eight short stories, each with a summarizing statement, rated them on three scales, and then

were given an unexpected immediate or delayed cued recall test. There were two types of cues. The precue consisted of the first phrase from each story; the postcues were the target sentences. It was found that when the concluding statement was expressed metaphorically rather than literally, there was increased memory for both the concluding metaphors and the preceding context. Further, there were no differences in recall with precues or postcues.

A few recent studies of metaphoric sentences have examined the differential effectiveness of topics, vehicles, and grounds as retrieval cues. In this regard, Marschark and Hunt (1985, Exp. 3) presented subjects with metaphoric sentences and asked them to rate each sentence for the number of interpretations they could think of. In a subsequent incidental cued recall test, subjects were given either the topics or vehicles as cues, and were required to provide the other term of the metaphor. There was no difference in the effectiveness of topics and vehicles as retrieval cues.

In a series of experiments, Verbrugge and McCarrell (1977) extended the range of retrieval cues to the study of grounds. (As stated previously, grounds summarize the major resemblance underlying the metaphor.) In their first experiment, two lists (A and B) of 14 metaphoric sentences of the form "Topic is (like) Vehicle" were prepared such

that the topics were identical but the vehicles were different across the two lists. Subjects were randomly assigned to one of the two list conditions, and asked to listen to each sentence and think about its meaning. An incidental cued recall test was then employed, and subjects received one of the following prompts: grounds A or B, topics A or B, or vehicles A or B. Subjects were requested to recall the entire sentence implicated by the prompt. The results indicated that recall with topic and vehicle prompts was nearly perfect, which was as expected, given that in such instances only half of the sentence remains to be recalled. Further, the grounds were almost as effective in prompting recall as the topics and vehicles. Experiments 2 and 3 confirmed that the ground can serve as an effective retrieval cue.

Light and Albertson (1987) replicated the procedure of Verbrugge and McCarrell (1977) with both young and old adults. Subjects were randomly assigned to one of the two list conditions described above, and instructed to listen to the sentences and think about the meaning of each. Incidental recall of the sentences was then examined in a within-subjects design with three cued recall tests: the first two utilized grounds as cues, and the third employed the topic of the sentence. It was found that the young adults recalled more sentences than the old adults. Topic cues were more useful than ground (meaning) cues, and there

was an interaction between age and cue type such that the effectiveness of topic cues relative to meaning cues improved recall more for the old than for the young adults.

Given this review of encoding and retrieval, the first experiment was designed to extend the methodology of Marschark and Hunt (1985), who conducted three experiments to examine memory for metaphors with respect to 10 variables previously identified as important to metaphor comprehension (Katz, Paivio, & Marschark, 1985; Marschark, Katz, & Paivio, 1983). In the first experiment, 10 groups of subjects each rated a set of metaphors on one dimension, and were subsequently presented an incidental free recall test. The results revealed that only two rated dimensions, the number of interpretations, and the imageability of the topic, reliably predicted recall.

The second experiment was designed to examine more specifically the number of interpretations dimension. In this regard, 36 metaphors, half previously rated as having many interpretations, and half as having few interpretations, were presented to subjects with orienting instructions requiring them to either: (1) rate the number of interpretations available, (2) produce as many interpretations as possible, or (3) produce a single interpretation for each metaphor. There was no significant effect of orienting condition. However, across all conditions, metaphors rated as having more interpretations

were recalled better than those rated as having fewer interpretations.

Cued recall with topic and vehicle cues was investigated in Experiment 3. The stimuli were metaphors having either a high or low number of interpretations. Subjects rated the metaphors for the number of interpretations possible, and were then presented with an incidental cued recall test given either the topics or vehicles as cues. The results indicated that topics and vehicles were equally effective as retrieval cues. Further, recall was better with metaphors having few interpretations rather than many interpretations.

Backman and Dixon (1987) extended this research to include older adults. As indicated in Chapter 1, they examined the number of interpretations dimension, and varied the procedure by presenting both metaphoric and nonmetaphoric sentences, and testing free and cued recall in a within-subjects design.

The present experiment varied encoding by requiring young and old adults to rate metaphoric sentences on one of three dimensions: imageability, number of interpretations, and semantic relatedness. A fourth intentional learning condition was also included. Retention was measured with free recall, cued recall, and recognition tests, respectively, in a within-subjects design. The cued recall situation provided an equal number of topics and vehicles as

retrieval cues.

Thus, the methodology of this investigation represented both a replication and a departure from previous research. As in the Backman and Dixon (1987) study, the inclusion of an older population permitted the investigation of age comparisons. Second, the addition of an intentional orienting condition fostered a comparison of incidental and intentional memory performance. The retrieval tests were also modified in various ways. As in the Backman and Dixon study, free and cued recall were both manipulated in one experiment; Marschark and Hunt (1985) examined them over different experiments. Further, topic and vehicle cues were a within-subjects factor (Backman & Dixon, 1987); Marschark and Hunt employed a between-subjects design in this regard. In addition, a recognition test was incorporated here to extend the amount of retrieval support beyond the cued condition in the previous research. Lastly, the inclusion of the imagery and semantic relatedness orienting conditions represented an attempt to more directly analyze two theories of metaphor comprehension. These theories will each be elaborated now before explicating more precisely their manipulation in the context of the second experiment.

Imagery

The concept of imagery has been theoretically elaborated and empirically investigated by Allan Paivio (1971), who has postulated a dual coding approach to human

language. The two codes - an imagery system and a verbal system - are assumed to be independent but partly interconnected, and to differ qualitatively in the nature of the information they control and generate. The imagery system processes nonverbal information stored in the form of images that correspond directly to concrete objects. It is synchronously organized, and specialized for the parallel processing of information. In contrast, the verbal system deals with abstract linguistic units involving discrete information that is only indirectly and arbitrarily related to objects. In addition, it operates on the sequential organization of linguistic information.

The distinction between synchronous and sequential organization has been further elaborated (Paivio, 1975). Synchronous organization refers not only to visual information, but to information specific to other senses (e.g., the sounds or smells associated with the imagined setting). Cognitive components are assumed to be simultaneously available for retrieval, although retrieval itself may proceed serially due to the limitations imposed by other systems (e.g., motor system). Sequential organization implies that multiple units of information are available successively. For instance, linguistic information is processed in this manner because the syntax that provides linguistic units with their meaning involves sequential structure. This is apparent at all levels of linguistic

analysis, from the ordering of phonemic elements (e.g., stop versus pots) to word order (e.g., "The cat chased the dog" versus "The dog chased the cat").

Paivio's (1971, 1975) theory has several additional implications for memory. First, verbal and imaginal processes are assumed to interact with the stimulus material. That is, concrete terms such as "tree" and "house" can be processed by either verbal or imaginal strategies, but only verbal mediators are available for encoding abstract items such as "love" and "truth". Thus, concrete items are easier to learn than abstract items because they can be processed potentially in either of two ways, while the latter are mediated only by verbal processes (Rowe & Schnore, 1971).

Another implication that has received empirical support is specified by the differential organization hypothesis (Paivio, 1975). It has been demonstrated that information contained in synchronously organized images requires less "storage space" than sequentially organized verbal information (Paivio, 1975). This suggests that more items can be effectively organized in memory as compound images than as verbal strings.

In a similar regard, Begg (1972) found evidence that high imagery phrases were remembered as holistic units because they required no more memory space than did their component words. Abstract low imagery phrases, on the other

hand, necessitated twice as much memory space as the individual words comprising them. This has been cited as evidence for the integration hypothesis (Paivio, 1975)- the notion that concrete phrases activate holistic long-term memory images in which the component words are organized into a single representation.

Paivio's (1975) theory has been criticized by Kirby and Das (1976), who have also postulated two types of cognitive activity referred to as simultaneous and successive processing. However, unlike Paivio, the premises of these systems originated in neurological and clinical work related to the integrative activity of the brain. Kirby and Das contended that Paivio's definition of imaginal and verbal processes was dependent on the nature of the task materials and instructions (verbal or nonverbal), while their emphasis was on the type of processes assumed to underlie task performance.

Paivio (1976) retorted by accusing Kirby and Das of overlooking the distinction between theoretical and operational definitions of concepts. He asserted that task manipulations and material variations were operational procedures that permitted the testing of theoretical implications. Further, he argued that Kirby and Das failed to recognize that they subscribed to an identical approach.

Support for the dual-coding model has been established in a concept identification task requiring participants to

attain concepts rated as either high or low in imagery (Katz & Paivio, 1975). The results indicated that high imagery concepts were attained more readily than low imagery concepts. Further, imagery instructions facilitated concept attainment when the concepts to be acquired were high in imagery value.

Belmore (1982) examined the retention of high and low imagery sentences in a continuous recognition task given either a verbal or an imaginal orienting condition. There were three types of distractors exhibiting either no change (NC), a syntactic form change (FC), or a meaning change (MC). It was found that the correct recognition of form changed items was greater for abstract than for concrete sentences. Further, the greater recognition of meaning over form changes was apparent for both sentence types. The author concluded that the data were supportive of a dual-coding interpretation.

Given this theoretical framework, a review of the empirical investigation of imagery will now be conducted.

There is considerable evidence that older adults utilize imagery less often and effectively than younger adults in their memory efforts (e.g., Hulicka, 1967; Hulicka & Grossman, 1967; Rowe & Schnore, 1971; Treat & Reese, 1976). An important factor here is the type of imagery used by a learner, and, in this regard, the distinction between interactive and noninteractive imagery has been drawn. An

interacting image depicts either an action between elements of the image (e.g., a dog chewing a ball) or structural unity between parts of the image (e.g., a dog made out of wood). A noninteracting or spatially separate image delineates two or more elements which are not in physical contact, and between which little or no action is evident (e.g., a dog sitting next to a ball) (Poon, Walsh-Sweeney, & Fozard, 1980).

In a study of paired-associate (PA) learning, Bower (1970) reported that 53% of the paired-associates were recalled using interactive images, 27% were recalled using noninteractive images, and 30% were recalled with a rote repetition learning strategy. This evidence for the efficacy of interactive images was corroborated by Begg (1973), who found that cuing facilitated recall more with integrated than with separated imagery instructions. It was also evident with picture-pairs, as learning was easier when the pair members were pictured together as unitized compounds rather than separately (e.g., Davidson & Adams, 1970; Rohwer, Lynch, Suzuki, & Levin, 1967). However, it appears from self-reports about imagery following a PA task that older adults use interactive imagery less often than this research would indicate (Hulicka & Grossman, 1967; Nebes & Andrews-Kulis, 1976; Rowe & Schnore, 1971).

Several investigators have examined whether instructions to use imagery would facilitate the memory

performance of older adults. In this regard, the paired-associate task has been utilized because mediational techniques (e.g., formation of images) play an important role in learning paired-associates. Such mediators integrate the two members of a pair into a whole so that minimal memory search is required for either of the pair members given the other.

The beneficial effects of imagery instruction with older adults have been demonstrated in several studies. For example, Hulicka and Grossman (1967) investigated the use of mediators in young and old adults in a paired-associate task under a variety of instructions. In the absence of imagery instructions, older subjects reported utilizing mediators less often and more inappropriately than did young subjects. However, when they were instructed to use a suitable mnemonic strategy, recall performance improved. The elderly improved more than the young adults, but age differences still remained.

The effect of interacting imagery instructions and cartoon mnemonics in paired-associate learning was examined in young, middle-aged, and old adults (Thomas & Ruben, 1973, cited in Poon et al., 1980). Retention was tested at one hour, and at 4 , 8 , and 16 months after learning to criterion. A substantial decline in retrieval performance was found for all groups across the retention periods. Further, subjects given imagery instructions or cartoon

mnemonics performed better than the standard instruction controls at the 4-, 8-, and 16- month retention intervals. With respect to age, under standard instructions the elderly performed at a lower level than the other age groups across all testing occasions. However, no age differences were evident with the imagery instructions or cartoon mnemonics at any test interval.

The utility of imagery instruction for old adults was evident in a study by Treat and Reese (1976) who found that, when the anticipation interval was long and subjects were instructed to use self-generated rather than experimenter-provided imagery, young and old participants performed equally well in a paired-associate task.

There is some indication, however, that the effects of imagery instruction may depend on the stimulus material utilized. For example, Poon and Walsh-Sweeney (1981) manipulated bizarreness and interaction in 20 paired-associates embedded in noun-verb-noun sentences, and instructed subjects in the experimental group to generate an image for each item during the study trial. Of the subjects who learned the sentences to criterion, it was found that interacting imagery provided the old experimental subjects with a significant advantage during learning which was not apparent at retrieval. Thus, the benefit of imagery instruction was localized to the acquisition stage of paired-associate learning.

Young, middle-aged, and old adults were given either standard free recall or imagery instructions to learn a list of 20 nouns (Mason & Smith, 1977, Exp. 2). The young recalled more words than the old, the performance in the imagery condition was superior to that in the standard instruction group, and there was a significant interaction between age and instructional condition: the imagery instructions were only effective in improving the recall performance of the middle-aged group.

In a study by Whitbourne and Slevin (1978), young and old adults generated their own word and picture mediators for abstract and concrete sentences. Three self-paced study-test trials were given, and it was found that: (a) the old recalled fewer words from the sentences than did the young, (b) concrete sentences were recalled better than abstract sentences, as predicted by a dual-coding approach, and (c) words were used more frequently than pictures in representing the meaning of the sentence. Further, older adults relied upon verbal mediators (words) for both sentence types, while the young adults used both kinds of mediators (words and pictures), particularly for the concrete sentences. In this instance, the authors concluded that old adults' recall was not benefitted by the use of mediators.

Lastly, Backman and Nilsson (1985) examined free recall in young and old adults after tasks requiring subjects to

remember lists of acts they had carried out (subject-performed tasks or SPTs), sentences with imagery instructions, or sentences without accompanying instructions. The same verbal expressions in imperative form (e.g., lift the spoon) were utilized in all three tasks. In the SPTs, subjects motorically performed the action implicated by the sentence. For the sentence condition, subjects were instructed to remember the sentences for a subsequent free recall test, but no overt action was required. In the sentence imagery task, subjects were told to visually imagine that they had actually executed the action described. It was found that older adults recalled fewer acts than younger adults for the sentence and sentence imagery conditions, but not for SPTs where recall performance was equivalent across age groups. Further, the old adults did not show any difference in recall between the two verbal tasks, suggesting that explicit instructions to utilize imagery did not differentially improve recall when compared with a no-instruction condition.

In summary, the above studies offer mixed support for the utility of imagery instructions in improving the memory performance of older adults. There is a fairly consistent trend for age differences in retention to be reduced or eliminated with imagery instructions in paired-associate learning, but not when other stimulus materials are employed. The following section will briefly address the

study of imagery in metaphoric material by discussing contradictory views of the importance of imagery to understanding figurative language, and then the specific issue of topic versus vehicle imagery as crucial to metaphor comprehension.

The significance of imaginal processes in metaphor comprehension has been considered by several investigators (e.g., Fainsilber & Kogan, 1984; Langer, 1948; Miller, 1979; Paivio, 1971, 1979; Sternberg & Nigro, 1983). For instance, Verbrugge (1977) has argued that imagery serves to "fuse" the two or more semantic domains represented in metaphor. Further, Billow (1977) has suggested that the metaphoric image may aid memory by mediating in the formation of associative connections.

Paivio (1979) has discussed the role of imagery in metaphor processing, and has outlined several contributions it, and the dual-coding approach in general, offers for this area. For instance, dual-coding enhances the probability of locating a common ground in long-term memory because the topic and vehicle may be connected in verbal memory, in the imaginal representation of the two terms, or in the integration of both processes.

The imagery system also makes specific contributions to metaphor comprehension. First, it may provide an additional (subjective) situational context for the interpretation of the metaphor. Secondly, as mentioned earlier, integrated

images allow efficient information storage because imaginal information is organized synchronously into large, integrated chunks. This implies that access to part of the structure will redintegrate the whole, and large amounts of information will become available.

Imagery also ensures processing flexibility, and, consequently, efficient memory search, because the information stored in images is not subject to sequential constraints. Rather, because it is synchronously organized, information can be retrieved in several ways, unlike the sequential constraints imposed on the retrieval of verbal information from long-term memory. This flexibility of image processing is also applicable to the construction of novel combinations from component information. Again, this is because information can be incorporated into a meaningful image in various ways, while words do not have this capability. The implication for metaphor processing is that imagery can not only increase the efficiency of the search for relevant information, but also for generating a novel, integrated representation constituting the common ground for the topic and vehicle.

Support for the role of imagery in metaphor comprehension was evident in three studies conducted by Harris, Lahey and Marsalek (1980). In the first experiment, participants listened to three types of sentences (metaphors, dead metaphors, and nonmetaphors) with either

imagery or regular instructions. They were requested to try to remember as much about the sentences as possible. A recognition test followed, and participants indicated which sentence of the three in each set they had heard, placing a "p" (for picture) beside each one if an image had been used to encode it. The results revealed no effect of the instructions or of input sentence type. However, images were reportedly used more often in encoding metaphors than nonmetaphors. Further, independent imagery ratings obtained for each sentence type indicated that metaphors were the most difficult sentence type to develop an image for (i.e., they had the lowest imageability ratings), in addition to their being reported as the sentences most frequently encoded with images.

Experiment 2 was similar to the first with the following exceptions: (a) incidental memory was tested, (b) an orienting task required participants to write the number of words in each sentence as they heard it, and (c) there were no imagery instructions. An identical forced-choice recognition test was administered along with the instruction to indicate the instances when imagery had been used. As in the first experiment, more images were reported for metaphorical sentences than for the other two sentence types. There was also a significant effect of sentence type on the recognition test. Nonmetaphors were the least correctly recognized, but they differed only significantly

from dead metaphors. In addition, there was an overall memory decrement in Experiment 2 as compared to Experiment 1, and it appeared to occur primarily for the nonmetaphors. Lastly, analysis of the counting errors indicated that there were significantly more errors for the metaphors than for the other two sentence types.

The third experiment conducted by Harris et al. (1980) did not involve memory, but attempted to discover the nature of the images evoked by metaphors. Participants were asked to form an image to sentences of all three types (i.e., metaphors, dead metaphors, and nonmetaphors), and to describe each image in words or with pictures. The images were scored for whether they contained both topic and vehicle domains, the topic domain alone, a less appropriate, but still metaphorical response, or an unclassifiable response. It was found that whenever the topic and vehicle domains were imaged together, they were not merely juxtaposed, but were intricately fused. Further observation of the responses indicated the presence of surrealistic images in participants' imagery descriptions.

Harris et al. (1980) also elaborated on the role of surrealistic imagery (i.e., the image constructed is nonveridical in that what it pictures could not literally occur in objective reality). They defended its importance in the construction of memory representations for metaphorical language, on the basis that metaphor involves two disparate

semantic domains, and is frequently a literally anomalous expression. They argued that greater consideration be given to surrealistic imagery rather than focusing on concrete, realistic imagery because imagery may be less perceptual, and more flexible, than has generally been assumed. Further, they contended that surrealistic imagery was essential to any general theory of imagery.

Other researchers, however, have questioned the significance of imagery in figurative language. For example, Reichmann and Coste (1980) argued that, in metaphor, as in literal language, there are abstract utterances that are comprehensible, but almost impossible to imagine in a perceptual manner. In this regard, they speculated that metaphor may be no different from literal speech if imagery is only applicable to concrete instances. That is, imagery may only be applicable to the comprehension of easily imagined sentences, and its role is very important in those cases. They also intimated that it is not clear whether imagery is a shallow level or deep level cognitive process (Craik & Lockhart, 1972), and whether images are stored in memory, or are transient components of the comprehension process.

Several experiments have been conducted by Riechmann and his associates to test the efficacy of an imaginal basis of figurative language comprehension. For example, Riechmann (1975) (cited in Riechmann & Coste, 1980) presented sentence

proverbs to two groups of subjects who were instructed either to form mental images of each of the proverbs, or to try to comprehend the meaning of each proverb as accurately as possible. Half of the proverbs had previously been rated as easily imageable and half as difficult to image. Immediate recognition of the interpretation of the proverbs was tested, and a significant advantage for the comprehend group relative to the imagery group was found. Riechmann construed this result as an indication that the memorial representation with instructions to comprehend was semantically more general, and thus more easily accessed, than that derived from instructions to form an image. Further, interpretations of low imagery proverbs were recognized significantly better than interpretations of high imagery proverbs.

Riechmann and van Wyk (1977) (cited in Riechmann & Coste, 1980) extended this research by first obtaining subjects' interpretations of each proverb and then selecting those proverbs with high semantic agreement across subjects as stimuli for a study identical in procedure to the one reported above. Again, the recognition of interpretations was significantly better for the comprehend group than for the imagery group, although the effect of proverb type (i.e., high imagery versus low imagery) was not significant. The authors concluded from the results of these proverb studies that imaginal representations tend to be specific

and inhibit the semantic flexibility required in many mental tasks.

With respect to metaphor, Riechmann and Coste (1978) (cited in Riechmann & Coste, 1980) examined recall of metaphoric sentences categorized as high, medium, or low in metaphoricity. There were four orienting conditions with instructions to: (a) memorize the sentences (M), (b) memorize and form mental images (MI), (c) memorize, form mental images, and rate the ease with which each image was formed (MIR), and (d) form mental images and rate the ease of doing so, but with no suggestion to memorize (IR). Written recall was cued with the subject-noun of each sentence. Verbatim recall failed to show a main effect of instruction, but a significant metaphoricity main effect indicated that high metaphoricity sentences were recalled more accurately than the other two types. Using propositional recall scores revealed a significant main effect for instructions due to the difference between the MIR and M groups only. Further, correlations between the imagery ratings and various recall measures failed to attain significance.

Riechmann and Coste (1980) contrasted the results of their metaphor study with the three experiments conducted by Harris, Lahey, and Marsalek (1980). They compared the contradictory finding of no effect of sentence type on memory performance in Experiment 1 of the Harris et al.

research to their evidence of better verbatim recall for high metaphoricity sentences. They postulated the nature of the recall task as a potential explanation. That is, they employed a one-word cue, while Harris et al. utilized a forced-choice recognition task in which the two incorrect alternatives were highly similar to the correct choice. If a more difficult retention task had been applied, whereby sentences were presented individually for recognition, Riechmann and Coste maintained that the results of Harris et al. might have been stronger, and perhaps more consistent with their own.

Riechmann and Coste (1980) also questioned the validity of the procedure Harris et al. (1980) used to acquire participants' imagery reports. By requesting subjects to place a "p" (for picture) during the recognition test beside those sentences that they had formed an image to when encoding, it was disputable whether subjects were able to discriminate between the memory of forming an image, and the inference that they must have done so. Further, they likely felt an implicit demand to place p's by a few of the sentences, and would probably have chosen the most different sentence of the three in the set (i.e., the metaphorical one).

Lastly, Riechmann and Coste (1980) maintained that the evidence in the third experiment presented by Harris et al. (1980) that subjects can form surrealistic images does not

imply that imagery is involved in the actual process of comprehension.

Riechmann and Coste (1980) concluded that, while there is a relationship between imagery and metaphor, there are boundaries to its role in the comprehension of, and memory for, metaphor. They suggested that the uniqueness of metaphor is not in the fusion of disparate domains of meaning, but merely the presence of slightly more peculiar combinations of semantic domains. They asserted that imagery is tangential to metaphor comprehension "or, at best, a process essential only at a shallow level of processing" (p. 199).

As an alternative conceptualization to imagery, Riechmann and Coste (1980) proffered the notion of a deeper-level conceptual base necessary to interpret the fusion of images, and dominant in memorial representations. The evidence for such a conceptual base was manifest in two experiments conducted by Honeck, Riechmann, and Hoffman (1975). In the first experiment, high and low imagery sentence proverbs were presented with either good, mediocre, poor, or unrelated interpretations. Each trial consisted of the proverb, its interpretation, a repetition of the proverb, and a repetition of the interpretation. Subjects were instructed to try to relate the proverb and its interpretation, and to remember as much as possible. Recall of the proverbs was subsequently tested, and the

interpretations served as retrieval cues.

It was found that recall was better with good interpretations than with unrelated interpretations, and that more content words were recalled with high imagery than with low imagery proverbs. Because they produced similar results, the good, mediocre, and poor interpretations conditions were collapsed to create a category designated "related interpretations". Analyses with this category indicated that recall was superior with related than with unrelated interpretations, but only for high imagery proverbs. The authors proposed that the subjects remembered conceptual bases, abstract mediators between the proverbs and their interpretations, that allowed them to deduce the correct proverb when provided with the interpretation.

In the second experiment, subjects wrote interpretations for high and low imagery proverbs. Retention of each proverb was then tested with a story cue or a single subject-noun cue. (The story did not comprise major vocabulary or propositional structure of the base proverb.) The results revealed that stories were better prompts for the recall of high imagery than low imagery proverbs, while the noun prompt produced equivalent recall for both types of proverbs. In addition, the correlations between the quality of proverb interpretation and recall were higher for low imagery proverbs than for high imagery proverbs.

The researchers concluded from these studies that the

results were consistent with a conceptual base interpretation. That is, people could encode an abstract characterization of a linguistic input that could be employed as a mediator in recognizing specific, conceptually related instances. The correlational evidence in Experiment 2 also suggested that the appropriate context might be necessary for the formation of this base. Thus, the conceptual base was viewed as abstract and rather general, and influenced by world knowledge. It was also imagery-free, and, as such, presented an alternative to imagery-based views of metaphor comprehension.

Further discussions of imagery have elaborated the role of topic versus vehicle imagery in understanding and remembering metaphor. In particular, Paivio (1979) has stated that the vehicle is more prepotent than the topic because, by definition, its properties are to be conveyed to the topic. In addition, the topic and vehicle of a metaphor are encoded in an analogous manner to stimuli and responses in paired-associate learning. Drawing a parallel to recall in the paired-associate task, he postulated that the concreteness of the vehicle should be crucial in metaphor comprehension, as recall accuracy in the former task depends on the concreteness of the retrieval cue. As stated earlier, a concrete term also provides rapid access to information-rich images. The vehicle (which specifies the topic) further serves as an efficient conceptual peg for metaphor

comprehension by promoting the retrieval of verbal and nonverbal information associated with the topic. While there is empirical support for this view (Backman & Dixon, 1987; Kroll & Schepeler, 1987; Verbrugge & McCarrell, 1977), contradictory evidence implicating the importance of topic imagery in recall (Marschark & Hunt, 1985, Exp. 1; Marschark, Katz, & Paivio, 1983; Nall, 1983, cited in Marschark & Hunt, 1985) has also been explicated.

Feature theory

As noted in Chapter 1, semantic feature theory is a similarity theory of metaphor comprehension. Representing a class of theory, it encompasses a variety of formulations that differ slightly with respect to their specific tenets. The development of this theory will be reviewed, beginning with the earlier research conducted by Malgady and Johnson in the mid- to late- 1970's. A more recent formulation by Ortony that burgeoned in response to criticisms will then be elaborated to indicate the current zeitgeist.

The impetus for a semantic feature based theory of metaphor processing originated with a cognitive feature model proposed by Johnson (1970) for free association. (The notion of features had already been developed by Katz and Fodor (1963) with respect to literal language.) It postulated that the basic unit of language behavior was not the word, but something more fundamental designated as an Elementary Cognitive Characteristic (ECC). The meaning of

every concept was defined as the vector of ECC elements associated with that concept. Thus, two concepts were similar in meaning to the extent that their ECC patterns overlapped.

Several experiments were conducted to determine whether this notion of similarity was applicable to metaphor processing. For example, Johnson, Malgady, and Anderson (1974) (cited in Malgady & Johnson, 1980) found evidence that the rated and ranked similarity between the topic and vehicle in metaphors and similes were predictable from the number of semantic features shared by the two terms in the figurative expression.

Similarity between the topic and vehicle terms of figurative, literal, and anomalous sentences was investigated by Malgady (1977), who examined the truth value of each proposition and the similarity of the noun pairs comprising each sentence. The results revealed that literal sentences were judged as containing highly similar terms that formed a logically true proposition. In contrast, anomalous sentences consisted of terms rated low in similarity that formed a false proposition. Metaphors were also recognized as false, but topic-vehicle similarity was intermediate to that of the other two sentence types. This provided evidence that the resemblance shared between the topic and vehicle terms of a metaphor was not sufficient to create a literal level of understanding. Further, it

suggested that literal, figurative, and anomalous sentences represented a continuum of language use, and that the best metaphors were those that expressed a moderate degree of similarity between the topic and vehicle terms.

Further support for a semantic feature interpretation of metaphor processing is manifest in studies of adjective modification. For example, Malgady (1976) presented participants with pairs of metaphors and requested them to rate which of the two metaphors in each pair contained the most similar topic and vehicle terms, or was the best figure of speech (i.e., metaphor goodness). Either the topic or vehicle noun in each metaphor was modified with one or two adjectives. The results indicated that the number of adjectives had a substantial effect on similarity and goodness judgments. Specifically, two adjectival modifiers consistently led to higher similarity and goodness judgments relative to metaphors containing only one modifier. This suggested that the number of semantic features shared by the topic and vehicle was related to metaphor understanding.

Adjectival modification was further utilized with metaphoric sentences to manipulate the features shared by the topic and vehicle nouns (Malgady & Johnson, 1976). In this experiment, the effect of different sentence contexts on metaphor comprehension was examined by generating five metaphoric versions of each sentence. Specifically, metaphors of the form "Noun A is noun B" were embedded in

adjective contexts that emphasized semantic features shared by both nouns, features distinctive to the topic and vehicle separately, features unrelated to either noun, or a control condition in which the nouns were not modified. Students either rated the five versions of each metaphor with respect to their goodness as figures of speech, or judged the similarity of the concepts compared in the metaphors without their sentence metaphoric context.

It was found that adjectives denoting features shared by both nouns in a metaphor resulted in the best rated comparisons of topic and vehicle. Modification by adjectives unrelated to the topic and vehicle was more detrimental to similarity and goodness judgments than when compared to unmodified metaphors, presumably because irrelevant information about the comparisons did not constitute part of the feature overlap defining the meaning of the metaphor. Emphasis on distinctive features of the topic and vehicle resulted in more similar concepts and better metaphors than when opposing features of the nouns were stressed. Consistent with a cognitive feature model of similarity, the authors concluded that adjectives constrained the possible meanings of the topic and vehicle through a cognitive reorganization of their semantic feature representations.

Thus, these studies provide support for a semantic feature theory specific to metaphor processing, and Johnson and Malgady (1979) have provided one such account. It is

premised on the assumption that the meaning of a word can be conceptualized as a set or vector of potential underlying feature elements (e.g., attributes, properties) that constitute all the possible meanings of a word. When words are combined in a metaphor, their meaning is determined by an additive summation of the feature sets for the individual words comprising the metaphor. Further, those feature elements that are shared by the words being combined are raised in salience in the resultant representation of the meaning of the metaphor. Johnson and Malgady (1979) have provided additional support for such a model from several kinds of data (e.g., association, similarity) collected on the topics and vehicles of metaphors in and out of sentence context.

Johnson and Malgady (1980) have also advanced a perceptual theory of metaphor comprehension that is complementary to feature models. Its emphasis is on understanding how individuals interpret metaphors, and focuses on an outside observer's point of view. But it also acknowledges that interpreting a metaphor involves experiencing a perceptual context in which the metaphor is comprehensible. Thus, it postulates that there are two possible sources of metaphorical interpretation: the material and the individual. In this regard, metaphors, like all language, do not have meanings per se, but are associated with a set of meaning possibilities potentially

realizable in any given individual.

Semantic feature theory has been criticized on several principles. These disapprobations will be elaborated first before describing a formulation that has attempted to address and circumvent these pitfalls.

First, several criticisms have been levelled at the notion of features. It has been argued that features provide only apparent precision because any given listing of features can be shown to be inadequate (Hoffman & Honeck, 1980). Similarly, the concept of common category membership provides too little a basis for metaphoricity because any two objects are common in at least one respect (Tourangeau & Sternberg, 1982). Further, semantic feature representations are posthoc and incomplete (Hoffman & Honeck, 1980; Verbrugge, 1984). That is, they partially label the results of a process, but do not begin to explain the process itself, and are unable to predict the processes underlying the comprehension of novel cases (Verbrugge, 1984).

The role of similarity in metaphor comprehension has also been questioned. Tourangeau and Sternberg (1982) noted that often the common feature between the topic and vehicle is shared only metaphorically. Ortony, Reynolds, and Arter (1978) agreed, and suggested that the distinction between literal and nonliteral similarity be made. Further, the role of comparison has been scrutinized. Ortony et al. (1978) contended that comparison is better regarded as a means of

comprehension than as the purpose of it. While this implies that metaphors cannot be explained by appeal to comparisons, it does not mean comparisons are not implicated in the comprehension process. Further to this view, Tourangeau and Sternberg (1982) have stated that metaphors make assertions and not comparisons.

Verbrugge (1984) has also noted that semantic representations for metaphor do not exhaust the plethora of experiences possible on reading them. This was corroborated by Camac and Glucksberg (1984), who asserted that the analysis of metaphor meaning should proceed at the sentence level, and not at the word level as propounded by the feature view. Further, the theory does not explicate why metaphors often seem surprising, it ignores the concept of dissimilarity (Tourangeau & Sternberg, 1982), and it implies that metaphors convey nothing new (Camac & Glucksberg, 1984).

Perhaps the greatest difficulty that has confronted feature theory is the problem of asymmetry (Camac & Glucksberg, 1984; Connor & Kogan, 1980; Malgady & Johnson, 1980; Ortony, 1979a). That is, metaphors of the form "A is B" are often not equivalent when the terms are reversed to "B is A". In fact, quite disparate meanings can result. The classic example is the metaphor "Surgeons are butchers", which becomes "Butchers are surgeons" when the terms are reversed. The former asserts a negative quality of surgeons,

while the latter describes butchers in a desirable way. Feature theory as described to this point is unable to account for this shift in meaning when the order of the terms is changed, because the same features should be activated and compared in both instances.

How can semantic feature theory contend with these objections? A formulation that has been proposed to deal with the problem of asymmetry, and that indirectly addresses some of the other issues raised, will now be elaborated.

The notion of similarity that has been discussed in the previous pages is premised on a geometric model of similarity, which espouses that the degree of similarity between two points is represented by an inverse function of the distance between them in a multidimensional space (Ortony, 1979b). Tversky (1977) presented an alternative to this approach, referred to as the contrast model, based on feature matching and able to cope with the asymmetry problem. In this model, the degree of similarity between two objects is expressed as a weighted function of the salience of their common features minus a weighted function of the features distinctive to each object. This model introduced three additional assumptions into the feature framework: (a) semantic features might be more or less representative of word meaning, depending on the context in which the word was uttered, (b) meaning could not be defined by an exhaustive list of semantic features since the feature set

was open-ended, and (c) the situational meaning of a word was also a function of factors such as the extralinguistic context and individual differences (Malgady & Johnson, 1980).

It was Ortony (1979a) who modified this contrast model and applied it to nonliteral similarity statements. One of the fundamental tenets of his model is that there is an imbalance in the salience levels of the matching attributes of the two terms in a metaphor, and that this imbalance serves as a principal source of metaphoricity. Specifically, given sentences of the form "A is B", literal constructions consist of matching attributes of equally high salience for both the A and B terms. However, in a nonliteral statement, the matching attribute is of lower salience in the A term than in the B term. (Anomalous statements consist of matching attributes that are of equally low salience for both terms, or the matching attribute is of higher salience in A than in B.)

Further, two types of similarity statements can be identified-attribute promotion and attribute introduction metaphors (Ortony, 1979a, 1979b). In the former, the hearer is assumed to possess enough knowledge about the topic to recognize that what is implicitly being said of it is true. The metaphor activates a recognition of existing similarities between concepts. Comprehension thus requires that the salience of the relevant features be promoted for

the A term through a process of feature selection (Malgady & Johnson, 1980). In the latter case, the hearer possesses very little information about the topic, and in comprehending the metaphor, novel similarities are discovered. Rather than increasing the salience of an existing feature, new features are introduced, and meaning is apprehended through a process of feature rejection (Malgady & Johnson, 1980).

Ortony (1979a) also advanced the asymmetry hypothesis, the notion that metaphors cannot be reversed without radical changes in perceived similarity and/or meaningfulness. In general, the degree of symmetry is inversely related to that of metaphoricity, such that the more metaphorical the comparison, the less symmetrical it is likely to be (Ortony, 1979a).

Ortony (1979a) has identifed another source that might enhance metaphoricity- the attribute inequality proposition. That is, attributes that are nominally the same frequently change their meanings when applied to terms in different domains. The criterion for a match in meaning, then, requires that the attributes be highly similar, rather than identical. This also suggests that domain incongruence serves to increase semantic distance, and therefore perceived similarity.

Similarly, context can also influence the salience of attributes. Ortony, Vondruska, Foss, and Jones (1985)

highlighted the distinction between context independent and context dependent properties of concepts. The former referred to properties inherent in the core meaning of a word, and activated by the word in all contexts. The latter, however, were only operative in certain contexts. When applied to metaphors, it suggests that they involve context independent (and therefore highly salient) properties for the B term, and context dependent (and therefore less salient) properties for the A term.

Given these basic suppositions, Ortony (1979a) argued that this proposed view of similarity (i.e., the imbalance model) was also consonant with a model of metaphor that espouses an interaction between the topic and vehicle terms in the specification of the ground. That is, the ground consists of shared attributes that are of high salience for the vehicle, but of low salience for the topic. Interaction is captured by this dependence of the attributes of the ground on both the topic and vehicle. Further, interaction is implicit in the constraints imposed by the topic term on the attributes of the vehicle term that can be applied.

The limitations of this imbalance model have also been discussed (Ortony, Vondruska, Foss, & Jones, 1985). First, the model is incomplete in its inability to account for the processes involved in searching for applicable attributes when judging similarity. In addition, Ortony (1979a) has suggested that attributes are subschemata that themselves

can be comprised of attributes. However, this renders it even more difficult to determine the precise nature of an attribute. Finally, the concept of salience also eludes definition. It has been characterized in numerous ways by different authors (e.g., Katz, 1982; Ortony et al., 1985), but important aspects may still have yet to be illuminated. Ortony et al. (1985) suggested that the most universal property of a salient attribute may be that it can be brought readily to mind. That is, highly salient properties are presumed to be more readily "accessible" than less salient properties.

The results of several recent experiments on metaphor are consistent with the imbalance account provided by Ortony (1979a). For example, Katz (1982) examined four indices of saliency by requiring participants to provide goodness ratings and interpretations of metaphors in their correct and reverse orders. The findings indicated that one aspect of saliency, typicality, was directly related to metaphor goodness. (Typicality was defined by the degree to which a given property matched with the prototype of the property.) Further, there was evidence for asymmetry in the metaphor goodness ratings, such that items of low topic-high vehicle salience were more metaphorical than items of high topic-low vehicle salience. In a later study, Katz (1983) collected normative data on the properties of objects with respect to two hypothesized aspects of salience, dominance

and typicality, and obtained evidence for their convergent and discriminant validity.

Further support for the salience imbalance notion was evident in a study conducted by Readence, Baldwin, Martin, and O'Brien (1984). When participants correctly interpreted a metaphor, the probability was 72% that the salient attributes were ranked in a low/high pattern. They concluded that salience imbalance generally enhanced metaphor interpretation, but was not a necessity. Lastly, Ortony, Vondruska, Foss, and Jones (1985) conducted four studies to investigate both the asymmetry hypothesis and the salience imbalance hypothesis, and obtained support for these conceptualizations.

Having reviewed the imagery and semantic feature theory formulations, their credibility was examined in the second experiment. The impetus for this research is expressed in the plea made by Paivio and Begg (1981):

> "Much of the psychological research on metaphor to date has not been directed at really fundamental problems in the area. Such work might require the systematic development of a large pool of novel metaphors that vary in type, difficulty, concreteness, and whatever other dimensions may seem relevant." (p. 287).

In this instance, novel metaphors varying on the dimensions of semantic relatedness and imagery were utilized to explore several issues pertinent to metaphor comprehension and retention.

The second experiment comprised two tasks to test the

utility of feature theory and imagery models, respectively, in memory for metaphoric sentences. The first task employed metaphoric sentences previously rated in Experiment 1 as high or low on semantic relatedness. The topic and vehicle nouns of the sentences were presented either as underlined terms in sentences (in context), or as isolated noun pairs (out of context). Young and old adults in each context condition were required to rate the similarity of the two nouns, and to list adjectives denoting properties they shared. Free recall of the topic and vehicle terms was then tested.

The design of this task was modelled on a study conducted by Johnson and Malgady (1979). They obtained several measures on the topic and vehicle nouns of metaphors in and out of sentence context, and then examined the intercorrelations between them. The results revealed a high correlation between measures taken in and out of sentence context. Further, all of the variables were interrelated. For example, highly similar words tended to have a large number of shared properties that were also highly salient. The authors concluded that the relationship between the topic and vehicle nouns in metaphors was related to the interpretation of metaphorical meaning.

Thus, the task employed here to test a semantic feature theory formulation represented an extension of the Johnson and Malgady (1979) research to a new subject population

(i.e., older adults) with an additional procedural variable (i.e., free recall). Further, the metaphors here were manipulated such that they varied on one dimension (i.e., semantic relatedness).

In the second task, metaphoric sentences formerly rated as high or low on imageability in Experiment 1 were presented auditorily to young and old adults either with or without accompanying imagery instructions to facilitate learning. Free recall, cued recall, and recognition tests were administered in a within-subjects design, analogous to the procedure of the first experiment. Cuing consisted of an equivalent number of topic and vehicle retrieval cues.

The procedure of this task did not represent an extension of any particular study, but an attempt to simultaneously combine several variables (i.e., instructions, sentence type, retention test) previously identified as pertinent to the investigation of imagery, and apply them to the memory performance of young and old adults.

Lastly, there was one additional variable that was examined in both experiments- analogical ability. The role of analogy in metaphor comprehension will now briefly be addressed before explicating more precisely the methodology of these two experiments.

There is some support for the notion that metaphorical understanding is a form of analogical thinking (e.g.,

Billow, 1975; Miller, 1979). It derives from their similarity in conceiving of something new in terms of that which is old. In particular, Sternberg and Nigro (1983) have elaborated on the theory of analogical reasoning processes proposed by Sternberg (1977), and elucidated its application to the study of metaphorical language. Specifically, they argued that a metaphor is based on an underlying analogy in which some of the terms are implicit. For example, similarity metaphors (two terms) can be contrasted with proportional metaphors (four terms) (Billow, 1977). It is the implicit nature of these terms that fosters an interaction between the topic and vehicle of the metaphor. They acknowledged that, while this view may not be applicable to all possible metaphors, its scope is comprehensive enough to be inclusive of a large and interesting variety of them.

Sternberg and Nigro (1983) presented an information-processing theory of metaphoric comprehension with components derivable from the apprehension of analogies. These components are as follows. First, the terms of the problem are encoded and identified, and relevant attributes retrieved from long-term memory. A relation is then inferred between the first two terms of the analogy. This relation is extended, or mapped, to the third term, where it is applied to generate an ideal completion of the analogy. This answer is compared with the alternative response options to

determine which is correct. One of the options is justified as better than the others, and is communicated with a response. Sternberg and Nigro (1983) argued that, although there is a correspondence in the component processes of analogies and metaphors, this does not imply an equivocal level of difficulty. With the additional verbal material present in the metaphor, the increased reading load must be balanced against the advantage of an increased mediating context.

Thus, analogical reasoning may be an important component of metaphorical understanding. For this reason, it was examined as a covariate in both experiments utilizing a written analogies test.

In summary, Experiment 1 manipulated encoding and retrieval conditions to examine memory for metaphors in young and old adults. The second experiment comprised two tasks to test the utility of semantic feature theory and imagery formulations as models for the retention of metaphoric sentences. Chapter 3 will now elaborate on the methodology employed, and the results obtained, in the first experiment. A discussion of the second experiment follows in Chapter 4.

CHAPTER 3

EXPERIMENT 1

Method

As a preface to the methodology of the first experiment, there were three orienting conditions requiring young and old adults to rate 60 metaphoric sentences for either their imageability, the semantic relatedness of the topics and vehicles, or the number of alternative interpretations possible. A fourth intentional condition requested participants to remember the sentences for a subsequent retention test. Incidental free recall, cued recall (with topics and vehicles as cues), and recognition of the sentences, respectively, were tested in a within-subjects design.

Subjects

Forty young adults aged 20-35 years (M = 25.43 years) and 40 old adults aged 60-75 years (M = 68.23 years) were recruited from the university campus and the community of Victoria, respectively, through posters, newspaper advertisements, and seniors' activity centers. They were each paid $10 for their participation in the one and one-half to two hour session. Equal numbers of males and females comprised each age group. A background questionnaire was administered to participants at the beginning of the experimental session to determine demographic characteristics such as health and number of years of

education. Further, analogical reasoning was examined with a 35-item test requiring participants to complete four-term analogy problems by choosing the correct response.

These demographic variables were analyzed for each age group. The mean age, number of years of education, and score on the analogies task is presented in Table 1 for each age group, sex, and orienting task condition. An analysis of variance on the analogies score to determine the effects of age, sex, and orienting task condition revealed no significant main effects or interactions. When the years of education variable was examined in a similar analysis, a significant main effect of age emerged, $F(1,64) = 7.565$, $p < .05$; the young had more years of education (M = 16.15 years) than the old (M = 14.53 years). A significant effect for sex was also observed, $F(1,64) = 6.662$, $p < .05$, indicating that males had more education (M = 16.10 years) than females (M = 14.57 years). Further, an age by orienting task condition interaction achieved significance, $F(3,64) = 4.089$, $p < .05$. T-tests indicated that in the imagery and intentional conditions, the young (M = 17.00 years, 16.90 years, respectively) had more education than the old (M = 14.00 years, 13.10 years, respectively); there was no difference in education between the age groups for the semantic relatedness (SRL) and interpretations (INTS) conditions.

Several self-reported indices of health were also examined. First, participants' subjective ratings of their

overall health as compared with their peers (on a scale of 1 to 5 from very poor to very good) indicated that 85% of the young rated their health as good or better, and that 92.5% of the old rated their health in this way. When compared to a perfect state of health, 87.5% of the young and 85% of the old rated their own health as good or better. Finally, whereas 20% of the young indicated that they were taking one or more drugs, 62.5% of the old indicated a similar response. Overall, the older participants in this experiment appeared to be in relatively good health.

Materials

Metaphors. Sixty metaphoric sentences of the form "(An) A is a B" were utilized. Thirty-five of these sentences (16 literary and 19 nonliterary) were selected from the normative metaphors provided by Katz, Paivio, Marschark, and Clark (1988). (These authors provide mean ratings for 204 literary and 260 nonliterary metaphors on 10 dimensions deemed relevant to metaphor comprehension.) Each was categorized as either a low imagery, high imagery, low semantic relatedness, or high semantic relatedness sentence. This was determined by selecting sentences at least one standard deviation from the scale mean (in the appropriate direction) on either the imagery (IMG) or semantic relatedness (SRL) dimension. In addition, there were other inclusion criteria the sentences were required to meet: (a) each was exclusively high or low on only one dimension, (b)

requiring participants to rate the sentences along a unique dimension. These dimensions were: (a) the ease or difficulty with which the entire sentence aroused mental imagery (imageability), (b) the number of alternative interpretations that could be furnished, and (c) the relatedness of topics and predicates (semantic relatedness). A fourth intentional condition that informed subjects to be prepared for a subsequent recall test was also used. Except for the number of interpretations and the intentional condition, the variables were rated on 7-point scales from "low" to "high." The stimuli were printed in booklets with a different cover page for each orienting condition. All instructions (see Appendix B) were accompanied by a depiction of the rating scale and three practise sentences. Each construct was defined, followed by concrete examples and precise information on the utilization of the rating scale.

Analogies Test. As discussed in Chapter 2, there is some support for the role of analogical reasoning in metaphor comprehension (e.g., Sternberg & Nigro, 1983). A measure of this ability was included in the present experiment to determine whether it might serve as a significant covariate in the recall of metaphoric sentences. The particular analogy test utilized consisted of 35 items randomly drawn from Part II of the Concept Mastery Test (CMT)-Form T developed by Terman (1956). The items

represented a wide variety of subject areas, and the level of difficulty increased progressively throughout the task. The items were presented in a booklet. For each item, three terms of the analogy, in the form a: b: as c: ?, were present. Participants were instructed to circle one of the three response choices that would best complete the analogy. An example was presented, and when all participants indicated that they understood the task, they were allowed to begin. Participants were allotted a maximum of 15 minutes to complete the task. In no instance was additional time required.

Procedure

Participants were randomly assigned to one of the four orienting conditions in a between-subjects design. Ten young adults and 10 old adults served in each condition. A maximum time of 15 minutes was apportioned for completion of the rating tests. This was followed by an unexpected written free recall test of as many sentences or parts of sentences as participants could remember. A 15 minute time limit was again imposed. Following the incidental free recall, cued recall of all of the sentences was investigated. The cues were randomly ordered in a test booklet such that half of the cues were topics and half were vehicles. Subjects were again allotted 15 minutes to recall the remaining sentence halves. Lastly, a recognition test consisting of 120 sentences (60 previous sentences and 60 distractors) was

presented. The distractors were all metaphors of the form "(An) A is a B" selected from the norms provided by Katz, Paivio, Marschark, and Clark (1988). Where possible, distractors comprised of a term identical to that of the original stimuli were included to increase test difficulty. No time limit was stipulated for this recognition task. Subjects were tested in small groups of young and old age peers. Within each testing session, participants from different orienting conditions were represented. All testing and scoring was conducted by a female experimenter.

Results

Ratings

The mean rating for each metaphor in each condition was calculated by averaging across individual participants' ratings. The first analysis was conducted to determine whether the imagery (IMG) and semantic relatedness (SRL) scale values previously reported by Katz, Paivio, Marschark, and Clark (1988) with a sample of young adults would replicate with the sample of young adults utilized here. In this regard, the mean ratings for the young adults were compared with those reported by Katz et al. (1988) for the 35 sentences selected from those norms. Utilizing a significance level of $p < .05$, t-tests indicated that 12 sentences were similarly rated in the imagery (IMG) condition, and 23 sentences for the semantic relatedness (SRL) dimension. When the mean scale ratings were compared

across studies, a significant difference for the IMG scale was evident, $t(40) = 2.70$, $p < .05$, but not for the SRL dimension. This suggests that the normative SRL scale value replicated with a different sample of young adults, while that value for the IMG scale did not.

The mean ratings for each scale were examined by sentence type to yield four mean ratings per scale: a mean rating for the high imagery sentences, the low imagery sentences, the high semantic relatedness sentences, and the low semantic relatedness sentences. These ratings are presented in Table 2 by age group and orienting condition. Analyses were conducted to determine whether the young and old adults differed in their ratings, if males and females rated differently, and if the three dimensions were similarly rated.

An analysis of variance (ANOVA) with repeated measures on the four sentence type ratings, and age (2), sex (2), and orienting condition (3) as between-subjects factors, indicated no main effects of age or sex, but a significant effect of orienting condition, $F(2,48) = 104.66$, $p < .05$. Posthoc analyses using Tukey's B procedure suggested that both the IMG and SRL scales had higher mean ratings than the interpretations (INTS) dimension (Ms = 4.61, 3.93, and 1.71, respectively). This is to be expected, as the INTS dimension did not have a similar underlying scale to the other two dimensions. In comparing the latter two scales, imagery had

a higher mean rating than the semantic relatedness dimension. The within-subjects effect of sentence type also achieved significance, $F(3,46) = 109.47$, $p < .05$. T-tests demonstrated that the mean ratings for the low imagery (M= 2.7) and low semantic relatedness (M = 2.5) sentences did not differ, and were significantly lower than for the high imagery (M = 4.2) and high semantic relatedness (M = 4.23) sentences, which were also equivalent. This provides validation for the prior categorization of the sentences into high and low subgroups.

In addition, there was a significant interaction between age and orienting condition, $F(2,48) = 4.16$, $p < .05$. Older adults had higher mean ratings than young adults on the SRL dimension, but both age groups had similar ratings on the IMG and INTS scales. Separate t-tests were conducted within each age group for the IMG and SRL dimensions, as their common underlying scale rendered them directly comparable. With respect to the young, the imagery scale had a higher mean rating than the SRL scale, $t(18) = 4.16$, $p < .05$. However, the two dimensions were equivalently rated for the old. These results indicate that, in contrast to the IMG and INTS scales, the SRL dimension was rated differently by the two age groups, and that the older adults rated the IMG and SRL scales similarly, while the young adults did not.

Lastly, there was a significant interaction between

orienting condition and sentence type, $F(6,92) = 14.62$, $p < .05$. Posthoc analyses employing Tukey's B procedure revealed that for high imagery, high semantic relatedness, and low semantic relatedness sentences, the mean rating for the INTS condition was lower than the mean for the SRL condition, which in turn was not as high as the mean for the IMG condition. For low imagery sentences, however, the IMG and SRL conditions were similarly rated, and both were greater than the INTS condition. Thus, the INTS scale was universally rated lower than the other two dimensions, regardless of sentence type, again owing to the different nature of the scale. The IMG scale was rated higher than the SRL scale for all sentence types, except low imagery sentences, where the two scales were similarly rated.

Relationship Among the Scales

In order to replicate the analyses of Katz et al. (1985) and Marschark et al. (1983), and thus provide comparability across studies, the relationship among the three scales was investigated. Recoding the data with sentence as a between-subjects factor, and values on the three scales as within-subjects factors, simple Pearson product-moment correlations were computed among the scales. The resulting coefficients are displayed in Table 3 for the young and old adults separately. All of the scales were highly intercorrelated, with R's ranging from .44 to .89. Two principal components analyses were performed on these

data, one for each age group. In both instances, only one factor with an eigenvalue greater than unity was detected, accounting for 72.6% of the common variance in the young sample, and 75% of this variance in the older group. These results were commensurate with those of Katz et al. (1985), who found that, with 10 dimensions, 77.6% of the common variance was accounted for by one factor. This connotes the presence of a single underlying dimension characteristic of all 10 scales, and of the three scales specifically utilized here.

To further clarify the relations among the three dimensions, three step-wise multiple regressions were performed for the young and for the old adults separately. In each case, one of the three scales was treated as the criterion variable, and the other two were entered as predictors. For both age groups, when the IMG scale was the dependent variable, the first significant predictor was the SRL dimension, accounting for 52.6% and 78.2% of the variance for the young and old adults, respectively. The INTS dimension was the second significant predictor, accounting for an additional 7.7% and 1.4% of the variance for the young and old, respectively. When the SRL scale served as the criterion, the only significant predictor was the IMG scale, accounting for 52.6% and 78.2% of the variance for the young and old groups, respectively. Lastly, the IMG scale again emerged as the only significant

predictor when the INTS dimension served as the criterion; it accounted for 32.7% and 24.7% of the variance for the young and old adults, respectively. These results lend further support to the notion that the IMG and SRL dimensions were highly related, as each accounted for the majority of variance in the other when one was used as the criterion; the INTS dimension appeared to play a relatively minor role in predicting performance on either of these two scales. Again, this may be due to the different scale underlying the ratings on the INTS dimension. When the INTS dimension did serve as the criterion, it was the IMG scale, rather than the SRL scale, that predicted performance.

In summary, there are several significant results pertinent to the rating of these three scales. First, the sample of young participants here rated the 35 sentences selected from previous norms with young adults similarly on the SRL dimension, but not on the IMG scale. Secondly, the overall mean scale ratings increased significantly for the INTS, SRL, and IMG conditions, respectively. This was true for the ratings of high semantic relatedness, high imagery, and low semantic relatedness sentences; only for low imagery sentences were the IMG and SRL scales rated equally, but still higher than the INTS dimension. This lower rating in the INTS condition was due to its different underlying scale relative to the other two scales. Further, the three scales loaded on a single factor, suggesting the presence of one

underlying dimension. However, the regression analyses implicated that the IMG and SRL scales were more closely related to each other than the INTS dimension, perhaps due again to the different underlying nature of the latter scale. The ratings of the young and old adults were similar on the IMG and INTS scales, but the older adults had higher ratings on the SRL scale than the young adults. Further, the older adults rated the IMG and SRL dimensions similarly, while the young adults did not. Lastly, there was a significant difference in the ratings of the high and low sentences across the IMG and SRL scales such that the former were rated higher than the latter.

Recall Analyses

The free recall and cued recall of each sentence was scored with strict and lenient criteria. First, with respect to free recall, a sentence was scored as strictly recalled if both the topic and vehicle nouns were exactly reproduced and in the correct positions. This implied that any modifiers present also be correctly included. The lenient criterion required that both topic and vehicle nouns be present, but allowed for approximations to the nouns (e.g., sword for bayonet), the interchange of topic and vehicle positions within the sentence, and generalizations of modifiers (e.g., small pillow for tiny pillow).

Cued recall was scored in five ways. If a sentence was

not responded to, it was scored as not recalled and termed an omission. Applying the same criteria as above, strict and lenient recall were scored for the one noun that was to be recalled. If the wrong noun (from a different sentence) was recalled, it was recorded as an intrusion. The recall of nouns not presented in the set of sentences were scored as constructions.

Free Recall, Cued Recall, and Recognition

The mean number of sentences recalled for each age group with free recall, cued recall, and recognition instructions is displayed in Table 4. An ANOVA with repeated measures on the three levels of recall task was conducted to investigate the effects of age (2), sex (2), and orienting condition (4) with a strict scoring criterion. There was a significant main effect for age, $F(1,64) = 28.10$, $p < .05$; the young ($M = 36.2$) recalled more sentences than the old ($M = 28.5$). The within-subjects effect of recall task was also significant, $F(2,63) = 2071.76$, $p < .05$. T-tests indicated that free recall ($M = 9.13$) was less than cued recall ($M = 31.01$) which was less than recognition ($M = 56.94$). Orienting condition and sex were not significant. There was a significant interaction between orienting condition and recall task, $F(6,126) = 2.64$, $p < .05$. A posteriori analysis with Tukey's B procedure revealed that orienting condition was significant only for free recall. That is, the intentional group remembered more than both the

interpretation condition and the imagery group. Lastly, the interaction between age and recall task was significant, $F(2,63) = 10.35$, $p < .05$ (see Figure 1). The young recalled more than the old for the free recall, cued recall, and recognition conditions ($F(1,78) = 19.44$, $p < .05$, $F(1,78) = 29.23$, $p < .05$, and $F(1,78) = 9.79$, $p < .05$, respectively). However, the difference in recall performance between the age groups was more for cued recall, less for free recall, and was minimal for the recognition task.

When a lenient criterion was employed in the same analysis, identical significant effects were obtained. However, the nature of the interaction between orienting condition and recall task for free recall was slightly modified- the intentional group only recalled more than the interpretation condition.

Proportional Recall Gain

As the previous analysis indicated, recall improved progressively over the three recall tasks as the amount of retrieval support increased. Further analyses were conducted to determine whether there were age differences in the amount of proportional gain between different levels of support.

A new proportional "gain" factor was created with two levels: (a) the ratio of cued to free recall scores, and (b) the ratio of recognition to cued recall performance. A 2 (age) by 2 (sex) by 4 (orienting condition) by 2 (gain)

ANOVA with repeated measures on the gain variable was conducted for a strict and lenient recall criterion separately. There was a significant main effect of age, $F(1,63) = 13.34$, $p < .05$; $F(1,63) = 13.37$, $p < .05$ for strict and lenient criteria, respectively. The mean proportional improvement in retention across the two levels of gain was greater for the old ($M = 4.23$) than for the young ($M = 2.93$) with a strict criterion and a lenient recall criterion ($Ms = 3.21$, 2.30 for old and young adults, respectively). The effect of orienting condition was also significant for the strict and lenient recall criteria, $F(3,63) = 3.03$, $p < .05$; $F(3,63) = 3.67$, $p < .05$, respectively. T-tests revealed that, given a strict recall criterion, the intentional group ($M = 2.67$) improved proportionally less than participants in the imagery ($M = 3.87$), semantic relatedness ($M = 3.71$), and interpretation ($M = 4.04$) conditions, which did not differ. Identical results were observed with a lenient recall criterion. In addition, a significant effect of the gain variable was evident with both the strict and lenient criteria, $F(1,63) = 78.85$, $p < .05$; $F(1,63) = 98.35$, $p < .05$, respectively. There was a greater proportional gain in recall performance between cued recall and free recall ($M = 4.87$) than between recognition and cued recall ($M = 2.25$) with a strict criterion and a lenient scoring criterion ($Ms = 3.75$ and 1.72, respectively).

Furthermore, a significant interaction was obtained between orienting condition and the proportional gain variable, $F(3,63) = 3.90$, $p < .05$; $F(3,63) = 4.20$, $p < .05$ for strict and lenient criteria, respectively. With a strict criterion, the semantic relatedness ($M = 5.58$) and interpretation ($M = 5.36$) groups improved proportionally more than the intentional condition ($M = 3.19$) from free to cued recall. However, there was no difference between the groups in improvement from cued recall to recognition. The lenient criterion indicated that the imagery ($M = 4.33$) and interpretation ($M = 4.44$) groups improved proportionally more than the intentional condition ($M = 2.52$) from free to cued recall, but there was no difference across conditions in improvement from the cued recall to recognition tasks.

These analyses of recall gain suggest that: (a) the old improved proportionally more than the young across the three retention tasks, (b) the intentional group improved proportionally the least when compared with the other three orienting conditions, (c) the mean proportional improvement between cued recall and free recall was greater than that between recognition and cued recall, and (d) there was more improvement from free to cued recall for the semantic relatedness and interpretations conditions than the intentional group with a strict criterion; the imagery and interpretations conditions improved more than the intentional group with a lenient criterion. However, when

proportional improvement from cued recall to recognition was examined, with both recall criteria, there was no difference across the four orienting conditions.

Free Recall Measures

Partial recall. A sentence was scored as partially recalled if: (a) either the topic or vehicle noun of a sentence was recalled alone, (b) two nouns from different sentences were recalled together, or (c) a correct topic or vehicle noun was paired with a new noun not in the sentence list. A 2 (age) by 2 (sex) by 4 (orienting condition) by 4 (sentence type) repeated measures analysis of variance was performed on the partial recall of high imagery, low imagery, high semantic relatedness, and low semantic relatedness sentences. There was a significant main effect for age, $F(1,64) = 7.93$, $p < .05$; the old ($M = 4.05$) partially recalled more sentences than the young ($M = 2.40$). The main effect of sex was also significant, $F(1,64) = 4.55$, $p < .05$. Females ($M = 3.85$) had more partial recalls than males ($M = 2.60$). Orienting condition and sentence type were not significant. However, the interaction between orienting condition and sex achieved significance, $F(3,64) = 3.06$, $p < .05$. T-tests indicated that females ($M = 5.90$) had more partial recalls than males ($M = 2.60$) only in the interpretations condition; the sexes did not differ in partial recall for the other three conditions. All other interactions failed to achieve significance.

Constructions. In the free recall task, the protocols of participants occasionally contained single words that did not constitute part of the original sentence set and were termed constructions. The total number produced by each participant was recorded. An analysis of variance was conducted on these constructions to examine the effects of age, sex, and orienting condition. However, there were no significant main effects or interactions of these factors on the number of constructions produced. Consequently, this variable was not considered in any further analyses.

Cued Recall

Topic and vehicle cues. The cued recall task was analyzed to determine whether there was differential recall given topics and vehicles as cues.

A repeated measures ANOVA was conducted for strict recall with topics and vehicles as cues, and age (2), sex (2), and orienting condition (4) as between-subjects variables. The main effect of age was significant, $F(1,64) = 26.72$, $p < .05$. The young ($M = 18.8$) recalled more than the old ($M = 12.21$) adults. The type of cue was also significant, $F(1,64) = 59.46$, $p < .05$. Recall was better with vehicle cues ($M = 17.2$) than with topic cues ($M = 13.8$). The main effects of orienting condition and sex were not significant, and there were no significant interactions between the factors.

An identical analysis was also performed for leniently

Constructions. In the free recall task, the protocols of participants occasionally contained single words that did not constitute part of the original sentence set and were termed constructions. The total number produced by each participant was recorded. An analysis of variance was conducted on these constructions to examine the effects of age, sex, and orienting condition. However, there were no significant main effects or interactions of these factors on the number of constructions produced. Consequently, this variable was not considered in any further analyses.

Cued Recall

Topic and vehicle cues. The cued recall task was analyzed to determine whether there was differential recall given topics and vehicles as cues.

A repeated measures ANOVA was conducted for strict recall with topics and vehicles as cues, and age (2), sex (2), and orienting condition (4) as between-subjects variables. The main effect of age was significant, $F(1,64) = 26.72$, $p < .05$. The young ($M = 18.8$) recalled more than the old ($M = 12.21$) adults. The type of cue was also significant, $F(1,64) = 59.46$, $p < .05$. Recall was better with vehicle cues ($M = 17.2$) than with topic cues ($M = 13.8$). The main effects of orienting condition and sex were not significant, and there were no significant interactions between the factors.

An identical analysis was also performed for leniently

scored cued recall with repeated measures on cue type. In addition to the two effects obtained above, there was a significant age by cue type interaction, $F(1,64) = 4.03$, $p < .05$ (see Figure 2). Posthoc comparisons applying Tukey's B procedure suggested that while the young recalled more than the old with both cue types, this difference in performance was greater with topic cues. There were no other interactions attaining significance.

Omissions, intrusions, and constructions. The mean number of omissions, intrusions, and constructions for each age group in the cued recall task is displayed in Table 5. A repeated measures MANOVA was performed to test the effects of age (2), sex (2), and orienting condition (4) on these three dependent variables by cue type. There were significant effects of age, cue type, and the age by cue type interaction, as previously discussed in the analysis of cued recall. The effect of the recall variable was also significant, $F(2,63) = 83.62$, $p < .05$. T-tests revealed that there were more omissions ($M = 17.43$) than constructions ($M = 3.34$), and more constructions than intrusions ($M = 1.43$). There were no main effects of sex or orienting condition. In addition, the interaction between age and recall variable obtained significance, $F(2,63) = 10.99$, $p < .05$. T-tests indicated that there were more omissions by old adults ($M = 23.80$) than by the young participants ($M = 11.05$) in the cued recall task. Lastly, the interaction between cue type

and recall variable attained significance, $F(2,63) = 10.41$, $p < .05$. The results of t-tests connoted that there were more intrusions with vehicle cues ($M = .89$) than with topic cues ($M = .54$), but there were more omissions given topic cues ($M = 9.40$) than given vehicle cues ($M = 8.03$).

Recognition Errors

There were two types of errors possible on the recognition test. A metaphor could have been incorrectly identified as part of the original set (false alarm), or incorrectly rejected as foreign to the set. These two types of errors were investigated in a MANOVA to test for the effects of age (2), sex (2), and orienting condition (4). There was a significant main effect of age, $F(2,63) = 4.35$, $p < .05$. However, univariate analyses indicated that this effect was significant only for errors of incorrect rejection, $F(1,64) = 8.84$, $p < .05$. The old committed more of these errors ($M = 4.70$) than the young ($M = 1.38$). There were no other significant main effects, or significant interactions.

Free and Cued Recall by Sentence Type

The free and cued recall measures were investigated by sentence type to examine whether recall was a function of the imageability and semantic relatedness of the sentences. There were four sentence types: high and low imageability, and high and low semantic relatedness. Four repeated measures ANOVAs were conducted for strict free, strict cued,

lenient free, and lenient cued recall separately. In each analysis, sentence type (4) served as the within-subjects effect, and age (2), sex (2), and orienting condition (4) as the between-subjects factors.

The strict free recall of the four sentence types was investigated first. As in a previous analysis with free recall, the main effects of age $F(1,64) = 22.16$, $p < .05$, and orienting condition $F(3,64) = 4.42$, $p < .05$, were significant. A significant main effect of sentence type also emerged, $F(3,62) = 17.82$, $p < .05$. T-tests indicated that the strict free recall of low semantic relatedness sentences ($M = 3.44$) was significantly better than that of high imagery ($M = 1.95$), low imagery ($M = 1.68$), and high semantic relatedness sentences ($M = 2.06$), which did not differ. The sex effect was not significant, and there were no significant interactions.

When a lenient criterion was employed, the main effects of age and orienting condition were again significant, $F(1,64) = 29.74$, $p < .05$; $F(3,64) = 4.46$, $p < .05$, respectively. A significant effect of sentence type was also observed, $F(3,62) = 20.43$, $p < .05$. However, t-tests revealed that recall was the lowest with low imagery sentences ($M = 2.26$), improved significantly with high semantic relatedness sentences ($M = 2.94$), and was optimal with low semantic relatedness ($M = 3.91$) and high imagery ($M = 3.99$) sentences, which were equivalent. Again, the sex

effect failed to achieve significance, as did the interactions between factors.

Analysis of the strict cued recall of these four sentence types yielded the age effect previously noted for strict cued recall. In addition, a significant within-subjects effect of sentence type was evident, F (3,62) = 74.95, $p < .05$. T-tests indicated that recall improved significantly for each sentence type given low imagery (M = 5.7), high imagery (M = 7.4), low semantic relatedness (M = 8.04), and high semantic relatedness (M = 9.88) sentences, respectively. The main effects of orienting condition and sex failed to achieve significance.

With respect to the relationships between the factors, there was one significant interaction between orienting condition and the strict cued recall of the four sentence types, $F(9,151) = 2.15$, $p < .05$. When Tukey's B procedure was employed to determine the nature of this interaction, however, the effect of orienting condition was not significant for any sentence type. Less conservative t-tests were utilized to try to elucidate this orienting condition by sentence type interaction. For all of the means that follow, the differences were significant at $p < .05$. For both the IMG and SRL conditions, strict cued recall of low imagery sentences was the poorest (Ms = 4.25 and 6.45 for IMG and SRL conditions, respectively). Performance improved for high imagery (Ms = 7.15 (IMG), 7.85 (SRL)) and

low semantic relatedness (Ms = 7.7 (IMG), 8.35 (SRL)) sentences, which did not differ. The best recall was attained with high semantic relatedness sentences (Ms = 10.0, 10.75 for IMG and SRL conditions, respectively). For the INTS condition, cued recall of low imagery sentences resulted in the worst performance (M = 5.25). Recall with high imagery (M = 6.85) and low semantic relatedness (M = 7.7) sentences improved; these two sentence types did not differ. Recall was optimal with both low semantic relatedness and high semantic relatedness (M = 8.45) sentences, which were equivalent. For the intentional condition, recall was the poorest for the low (M = 6.85) and high imagery (M = 7.75) sentences, which did not differ significantly. It improved with low semantic relatedness (M = 8.4) and high imagery sentences, which were equal. Recall was best with high semantic relatedness sentences (M = 10.3). In general, across orienting conditions, the findings suggest that strict cued recall was the poorest with low imagery sentences, improved with high imagery and low semantic relatedness sentences, and was the best with high semantic relatedness sentences. There were no other significant interactions between the factors.

When lenient cued recall was investigated by sentence type, the main effect of age was significant, $F(1,64) = 25.64$, $p < .05$, as noted in a prior analysis of cued recall. The within-subjects effect of sentence type was also

significant, $F(3,62) = 84.54$. $p < .05$. T-tests revealed that recall improved significantly over the four sentence types, respectively: low imagery ($M = 6.83$), low semantic relatedness ($M = 8.83$), high imagery ($M = 10.81$), and high semantic relatedness ($M = 11.35$). There were no other significant main effects or interactions.

These analyses of recall examined by sentence type suggest that, overall, low imagery sentences were the least frequently recalled. With free recall, low semantic relatedness sentences were the best recalled, but with cued recall, high semantic relatedness sentences were optimal.

In summarizing the recall analyses above, recall improved significantly across free, cued, and recognition tasks, respectively. Further, the young recalled more than the old adults across the three retention tasks, with the exception of partial free recall, where the old outperformed the young. The difference between the age groups was greatest for cued recall, less for free recall, and minimal with the recognition test. With respect to free recall, there was an effect of orienting condition, such that recall was better in the intentional condition than in the interpretation condition, for both a strict and a lenient criterion; the intentional condition was also better than the IMG condition with a strict criterion. Vehicles were more effective recall cues than topics in the cued recall task, and while the young recalled more than the old

with both types of cue given a lenient scoring criterion, the difference between the age groups was greater with topic cues. In further analyses of the cued recall task, there were more omissions than constructions, and more constructions than intrusions. With respect to age, there were more omissions by the old than the young adults. In addition, more intrusions were produced with vehicle cues than with topic cues, while the reverse was apparent for omissions. When the recognition test was examined for two types of errors, the old were observed to commit more errors of incorrect rejection than the young. Lastly, the effect of sex was significant only for partial free recall- females had more partial recalls than males, and in the interpretation condition when examined by orienting condition.

Ratings and Recall Data Considered Simultaneously

In order to determine whether part of the variability in recall scores might be due to the previous sentence ratings, the mean rating was analyzed as a covariate in several repeated measures analyses. The three levels of recall (free, cued, and recognition) was the within-subjects factor, and age (2), sex (2), and orienting condition (3) the between-subjects factors. The covariate was not significant when either a lenient or a strict recall criterion was utilized. However, after the adjustment with the covariate, the interaction between orienting condition

and recall task was no longer significant. Additional ANOVAs specifically testing the effects of age, sex, orienting condition, and cue type (the within-subjects variable) on strict and lenient cued recall also failed to produce a significant effect of mean rating as a covariate. Although the covariate did not change the pattern of results with the strict criterion, the interaction between age and cue type was no longer significant with a lenient criterion.

The mean rating was partialled by sentence type to create four mean rating covariates for high imagery, low imagery, high semantic relatedness, and low semantic relatedness sentences. This was done to ascertain whether prior ratings for specific sentence types would display a significant relationship with measures of recall performance. The same repeated measures ANOVAs as above were conducted, but with the four new covariates. There were no significant effects of the covariates for strict and lenient recall across the three recall tasks, or for the examination of cued recall by cue type. However, the covariates produced the same effect on the results of each analysis as reported in the preceding paragraph.

Thus, these covariance analyses indicate that the sentence ratings covariates were not significant, and that the pattern of recall results previously obtained in analyses of variance did not change. That is, the main effects of age, recall task, cue type, and sentence type

were still significant, as was the interaction between age and recall task.

Serial Position Effects

As the free recall of each sentence was coded in the precise order as its presentation in the rating task, it was possible to query for the presence of primacy and recency effects. The former refers to memory for items introduced at the beginning of a list, and the latter represents memory for items presented at the end of a list.

In this instance, the presented set of 60 sentences was evenly divided into three sections. The first 20 tested a primacy effect, the second 20 represented the asymptote, and the final 20 were indicative of a recency effect. For each list position, recall was expressed as a proportion out of twenty. Age, sex, and orienting condition effects were investigated in three separate ANOVAs with repeated measures on list position. Free recall (strict and lenient criteria separately) and partial recall were each analyzed.

First, the 2 (age) by 2 (sex) by 4 (orienting condition) by 3 (list position) ANOVA on strict free recall with repeated measures on the final factor revealed that, in addition to the significant age and orienting condition effects previously obtained, there was a significant main effect of list position, $F(2,63) = 4.89$, $p < .05$. T-tests indicated that recall of items in the middle of the list ($M = .13$) was significantly less than recall of sentences at

the beginning (M = .17) or at the end (M = .16) of the set. Sex was not significant. However, there was a significant interaction between orienting condition, sex, and list position, F(6,126) = 2.41, p < .05. T-tests revealed that in the SRL condition, females (M = .22) recalled proportionately more sentences than males (M = .07) from the middle of the list. There were no other significant interactions.

The same analysis was performed again to test for the effects of age, orienting condition, sex, and list position on lenient free recall. The main effects of age and orienting condition were significant, as in previous analyses of lenient free recall. A significant effect of list position also emerged, F(2,63) = 5.42, p < .05. T-tests demonstrated that fewer sentences in the middle of the set were recalled (M = .19) when compared with the end of the set (M = .24); there was no difference between recall of sentences at the beginning (M = .22) or in the middle (M = .19) of the set, or between those at the beginning or the end of the list. The main effect of sex was not significant.

There was a significant interaction between orienting condition and list position, F(6,126) = 2.79, p < .05 (see Figure 3). Posthoc analyses utilizing Tukey's B procedure suggested that, within the first part of the list, the intentional group (M = .33) performed better than the other three groups (Ms = .16, .19, .20). Recall of the middle part

of the set was better for those in the intentional group (M = .29) as compared with the imagery (M = .11) and interpretation (M = .16) conditions. There was no significant difference between orienting conditions in the lenient free recall of sentences in the last part of the set (Ms = .20, .24, .26, .28). Further, the interaction between age and list position was significant, $F(2,63) = 3.98$, $p < .05$; in all list positions, the young (M = .29) recalled more sentences than the old (M = .15). However, this age difference in recall increased progressively from the beginning of the set to the end (see Figure 4).

Partial recall was also investigated for possible effects of list position. However, besides the significant effects of age, sex, and the interaction between orienting condition and sex previously reported, there were no significant main effects or interactions of the list position variable.

In summary, there was some support for serial position effects in free recall. With a strict criterion, sentences from the middle of the list were recalled less frequently than those from the beginning or end of the set. A significant three-way interaction also indicated that, in the SRL condition, females recalled more sentences from the middle list position than males. With a lenient criterion, recall of sentences from the middle of the list was significantly less than from the end of the set. In

addition, there were significant interactions between orienting condition and list position, and age and list position. Specifically, the intentional condition resulted in better recall than other selected conditions for sentences from the beginning and middle of the list; there was no difference across conditions for the recall of items from the end of the list. The young recalled more than the old across all list positions, but the difference in their recall performance increased from the beginning to the end of the set.

Literary and Nonliterary Sentences

Separate analyses were conducted to examine whether the literariness of the sentences influenced rating judgments and recall performance. As described in the methodology section, there were 16 literary and 44 nonliterary sentences. The rating and recall measures for these two sentence types were represented as proportions out of 16 and 44 for the literary and nonliterary sentences, respectively.

The mean ratings for literary and nonliterary sentences were investigated first. A repeated measures ANOVA design was utilized to test the effects of age (2), sex (2), and orienting condition (3) on the mean ratings of literary and nonliterary sentences. The orienting condition effect, and the interaction between age and orienting condition were significant, as in previous rating analyses. The within-subjects effect of sentence type was also significant,

$F(1,48) = 85.57$, $p < .05$. Nonliterary sentences had higher mean ratings across the three scales ($M = 3.61$) than literary sentences ($M = 2.88$). In addition, the interaction between orienting condition and sentence type was significant, $F(2,48) = 10.90$, $p < .05$. A posteriori analyses with Tukey's B procedure indicated that for both sentence types, the mean ratings on the three scales differed significantly from each other, and increased across the INTS, SRL, and IMG conditions, respectively. However, the differences between these conditions were larger for the nonliterary than for the literary sentences.

The strict and lenient free recall of literary and nonliterary sentences were examined utilizing the ANOVA design described above. The effect of sentence type was marginally significant with a lenient criterion, $F(1,64) = 3.78$, $p = .056$, and recall was better with literary ($M = .24$) than with nonliterary ($M = .21$) sentences. Except for the significant main effects of age and orienting condition previously elaborated, there were no significant main effects or interactions.

Cued recall was studied with respect to sentence literariness in two separate repeated measures ANOVAs for the strict and lenient criteria. The effects of age (2), sex (2), orienting condition (4), and sentence type (2), with repeated measures on the final factor, were examined. In both instances, there was a significant main effect of

sentence type, $F(1,64) = 63.67$, $p < .05$ and $F(1,64) = 54.04$, $p < .05$ for strict and lenient cued recall, respectively. Cued recall was higher with nonliterary sentences than with literary sentences given a strict criterion (*M*s = .54 and .44, respectively) and a lenient criterion (*M*s = .66 and .56, respectively). As was evident from earlier analyses of cued recall, the effect of age was significant for both the strict and lenient recall criteria. There were no other main effects or interactions achieving significance.

Partial recall of these two sentence types was also investigated in a manner similar to the above analyses. The age, sex, and orienting condition by sex interaction effects were significant, as in previous analyses of partial recall. Sentence type was also significant, $F(1,64) = 19.16$, $p < .05$, but partial recall was greater for literary sentences (*M* = .08) than for nonliterary sentences (*M* = .05). Further, there was a significant interaction between sex and sentence type, $F(1,64) = 4.54$, $p < .05$. Females had more partial recalls (*M* = .08) than males (*M* = .05), but the difference between the two sexes was greater for literary sentences than for nonliterary sentences. No other main effects or interactions were significant.

These analyses of sentence literariness indicate that nonliterary sentences had higher mean ratings across the three rating scales, and were better recalled in the cued recall task, than literary sentences. Further, the

interaction between orienting condition and sentence type in the rating task revealed that ratings for both sentence types increased significantly for the INTS, SRL, and IMG conditions, respectively, but the differences were larger for nonliterary than for literary sentences. Only with partial free recall were literary sentences recalled better than nonliterary sentences. In this instance, there was also a significant interaction between sex and sentence type, connoting that, while females partially recalled more sentences than males across both sentence types, this difference was greater for literary than for nonliterary sentences. Lastly, sentence literariness was marginally significant with lenient free recall, but was not related to strict free recall performance.

Covariates

Years of Education

In an attempt to discover whether some variability in recall performance across the two age groups might be accounted for by another factor, years of education was entered as a covariate into the 2 (age) by 2 (sex) by 4 (orienting condition) ANOVAs with repeated measures on recall (free, cued, and recognition) or cue type (topics or vehicles). In each of these analyses, the pattern of results was not affected by the covariance adjustment. Years of education was a significant covariate only for constructions with topic and vehicle cues, F (1,63) = 12.91, $p < .05$. This

effect was corroborated when an analysis of variance on constructions, with education as a covariate, attained significance, $F(1,63) = 11.65$, $p < .05$. The significant correlation between the years of education variable and constructions was .37. However, when examined by age group, the correlation was .11 for the young, .55 for the old adults, and significant only for the latter group. This suggests that, for the older participants, some variability in the production of constructions in the cued recall task was due to the years of formal education attained.

Analogies

The score achieved on the analogies task was also utilized as a covariate. Pearson product-moment correlations between the analogies score and recall measures indicated several significant relationships. These correlation coefficients are displayed in Table 6 for the young and old adults separately.

In a similar manner to the education variable tested above, analogies was entered as a covariate into the repeated measures analyses for recall and cue type. It was significant for the strict criterion of free recall, cued recall, and recognition, $F(1,63) = 20.82$, $p < .05$, and for the corresponding lenient criterion, $F(1,63) = 15.59$, $p < .05$. Further, the covariance adjustment produced a significant main effect for sex with a strict criterion, $F(1,63) = 4.93$, $p < .05$; the results with a lenient

criterion were identical to those previously observed in analyses of variance. When strict and lenient recall given topic and vehicle cues were examined, analogies was also a significant covariate, $F(1,63) = 28.69$, $p < .05$, $F(1,63) = 19.73$, $p < .05$ for strict and lenient criteria, respectively. Again, a significant effect of the sex variable was obtained with a strict criterion, $F(1,63) = 4.08$, $p < .05$, but the results of the lenient recall analysis were analogous to those previously reported in analyses of variance. Further, omissions given topic and vehicle cues also displayed a significant regression effect for analogies, $F(1,63) = 19.20$, $p < .05$. This was verified in an analysis of covariance to examine the effects of age, sex, and orienting condition on omissions; analogies was a significant covariate, $F(1,63) = 21.55$, $p < .05$. In both of these covariance analyses, there was no change in the nature of the previously obtained results.

These results suggest that analogical reasoning, as measured by the test employed here, was related to the recall of metaphoric sentences. Further, adjustment with this covariate did not change the results of previous analyses of variance. The main effects of age, recall task, and cue type were still significant, as were the interactions of recall task with age and orienting condition. In addition, the correlations between the recall measures and the analogies score were positive, indicating

that a higher analogies score was related to better recall performance. Only for omissions was the correlation negative, implying that a higher analogies score was related to fewer omissions and, consequently, better recall.

Discussion

Encoding: Rating of the Three Dimensions

The results provide some evidence for the replicability of previous research with attributes of metaphor processing. First, there was support for the normative scale values specified by Katz et al. (1988), particularly for the semantic relatedness dimension. This was apparent in the ratings of the young adults on the semantic relatedness scale when compared with the Katz et al. ratings by age peers, but not for ratings on the imagery dimension. The lack of replicability with the imagery dimension may suggest that the samples of young adults were not comparable across the two investigations.

There was also evidence for the categorization of the metaphoric sentences into high and low subgroups on both dimensions. That is, significantly higher ratings were evinced for high imagery and high semantic relatedness sentences than for their low counterparts. This implies that the experimental manipulation of these dimensions is valid.

Lastly, when principal components analyses were conducted with the three dimensions, only one factor emerged, indicating a single underlying dimension of

metaphor comprehension. This finding replicates the research of Katz et al. (1985) with 10 dimensions. Considered with the evidence above for the valid categorization of sentences into subgroups, this suggests that, although each of the scales is independently manipulable in research, in composite they characterize metaphor comprehension.

With respect to the relationship between the scales, there was a greater mean rating for the imagery scale than for the semantic relatedness scale, which also corroborates previous research (Katz et al., 1985; Marschark & Hunt, 1985; Marschark et al., 1983). This was apparent across all sentence types, except low imagery sentences, where the two scale ratings were not significantly different. This connotes that it may be easier to form images than to judge similarity. That is, semantic similarity is probably based on more objective criteria than subjective impressions of imageability, and thus requires a more stringent decision criterion to warrant higher ratings. However, this assumes that the two scales had equal underlying intervals, and that each of the seven scale gradations were analogous reference points across the two scales. It also presupposes that the stimulus sentences were equally variable along both dimensions; such assumptions may be fallacious.

Lastly, the interpretations dimension, because of its different underlying scale, was rated lower than the other two dimensions. The regression analyses indicated that it

was not a significant predictor for the semantic relatedness scale, but was marginally significant for the imagery scale. This is consistent with previous studies demonstrating that the interpretations scale was not a significant predictor for the semantic relatedness dimension, but it was marginally significant with the imagery dimension (Katz et al., 1985; Marschark et al., 1983).

Concerning the age variable, there was no main effect, but a significant interaction with orienting condition was observed. The two age groups rated the imagery and interpretations dimensions similarly, but the old had higher semantic relatedness ratings than did the young adults. This suggests that the old and young participants did not perceive similarity in the same manner. Further, within each age group, the old did not differ in the imagery and semantic relatedness ratings, but the young rated the former more highly than the latter. In addition, the absence of an overall age effect was apparent in the regression analyses, indicating that there was no difference between the age groups in the order of the significant predictors that emerged. Thus, in general, the influence of age was minimal, and restricted to conceptions of semantic similarity. This suggests that the applicability of these dimensions of metaphor processing, previously normed only with younger adults, may be extended to older adults.

Retrieval

Across the three retention tasks, the young recalled more than the old with both a strict and a lenient scoring criterion, and recall improved significantly for the free, cued, and recognition tasks, respectively. With both criteria, recall improved proportionally more between free and cued recall than between cued recall and recognition, and the old improved more than the young adults across the three levels of retrieval support. While age differences were greater with cued recall, they were less with free recall, and were minimal with the recognition task. This supports the findings of Perlmutter (1978, 1979) of no difference between young and old adults in the retention of words when recognition was combined with a semantic orienting task.

The effect of orienting condition was significant solely for free recall, where the intentional group recalled more than the interpretations group with both a strict and a lenient scoring criterion, and more than the imagery condition with a strict criterion. However, when incidental recall of the three rated dimensions was compared, memory performance did not differ. This was corroborated in an analysis of proportional recall improvement across tasks, as the intentional group gained less than the other three conditions, which did not differ. This is consistent with the results of Marschark and Hunt (1985), who found that

orienting condition did not influence incidental free recall. The superiority of the intentional condition over some of the incidental conditions is consistent with the research indicating that intent to learn is an important factor in memory performance (e.g., Craik & Tulving, 1975; Hyde & Jenkins, 1969; Till & Jenkins, 1973).

Lastly, the advantage of the intentional condition was only apparent in the free recall situation, and the interaction between orienting condition and the proportional gain variable indicated that the intentional group improved less than some of the incidental conditions only from free to cued recall; there was equivalent proportional improvement from cued recall to recognition for all four orienting conditions. It is speculated that the provision of cues compensated participants in the other three conditions for the deeper processing performed by those with knowledge of a subsequent recall test. As the cues employed consisted of displaying half of each sentence, the cognitive demand was considerably reduced compared with the free recall task.

When partial free recall was examined, the age effect was reversed: the old had more partial recalls than the young adults. This indicates that the old recalled more sentence fragments, or paired more of them with each other or with new words, than the young. Thus, they stored more incomplete memory traces than the young adults. Further, females had more partial recalls than males, and sex

interacted with orienting condition such that females had more partial recalls than males in the interpretations condition. These sex effects were unexpected, and suggest that females stored more incomplete memory traces than did males.

As postulated by Paivio (1979), the cued recall situation demonstrated that vehicles were better cues than topics (Backman & Dixon, 1987; Kroll & Schepeler, 1987). That is, the vehicle guides the representation of the topic by conveying properties to the topic. This result also supports Ortony's (1979a) hypothesis of a higher salience for the vehicle term than the topic term in a nonliteral statement. Further, given a lenient cued recall criterion, the young recalled more than the old participants with both cue types, but the difference was greater with topic cues. This suggests that the combined effects of age and cue type were particularly evident in the poorer recall performance of old adults with topic cues.

When the cued recall task was further examined, there were more omissions than constructions, and more constructions than intrusions. The relatively small number of instances in the latter two categories, however, indicates that errors on the cued recall test were more often those of omission than of commission. Further, errors of commission were more likely to comprise new words rather than the confusion of words across different sentences.

There was an effect of age for omissions, as the old omitted more of the sentences in the cued recall task than the young. Previous research has also demonstrated that age differences in errors are due to more omissions by old adults when compared to young participants; age differences are not usually apparent in errors of commission (Canestrari, 1963; Eisdorfer, 1965). In addition, there were more intrusions with vehicle cues than with topic cues. However, with very small means (less than one), the practical significance of this result is questionable. Also, there were more omissions with topic cues than with vehicle cues, which is not surprising given that vehicle cues were demonstrated to be more beneficial to recall performance.

When the recognition test was examined, the old committed more errors than the young by incorrectly rejecting the correct response. This may indicate a hesitancy on the part of the old to positively identify a sentence unless they are absolutely certain it is correct (Botwinick, 1967). Thus, they may exercise a different decision criterion than the young in the recognition of sentences.

Sentence Literariness

The effect of literariness was significant in the ratings across the three dimensions. The higher ratings for nonliterary rather than literary sentences indicated that the literary metaphors were viewed as less imageable, their

nouns as less similar, and as having fewer interpretations than nonliterary sentences. This is consistent with the research of Katz et al. (1985, 1988) comparing the ratings of poetic and nominal metaphors. It is speculated that the lack of context in the presentation of the literary metaphors may account for their poorer ratings. This point will be elaborated later. Further, while the mean ratings increased across the interpretations, semantic relatedness, and imagery conditions for both literary and nonliterary sentences, the difference across the three conditions was larger for nonliterary than for literary sentences. This suggests that the different ratings across the dimensions were more pronounced for nonliterary than for literary sentences.

When retention was examined, there was no effect of sentence type for strict free recall, but significance was marginal with a lenient criterion ($p = .056$), and inspection of the means connoted that literary sentences were remembered better than nonliterary sentences.

For both strict and lenient cued recall, nonliterary sentences were remembered better than literary sentences. With the provision of cues to guide retrieval, it seems reasonable to suppose that the other sentence half would be recalled better for nonliterary sentences, where ratings indicated that the similarity between the nouns was higher than for literary sentences.

With partial free recall, literary sentences were retained better than nonliterary sentences. This might be explained with reference to the notion that if only part of a sentence is remembered in the free recall situation, or is incorrectly paired with another noun, it will probably occur for sentences that are extreme or unusual in some regard. In this instance, the literary metaphors were presented out of context, rendering some of them unusual, and thus more likely to be remembered than nonliterary sentences. Further, this might explain why only part of the sentence is recalled; if the literary sentence is quite peculiar, only part of it may be recallable. Thus, the degree of sentence bizarreness may also be important to partial free recall performance. Lastly, females recalled more sentence fragments than males, but the difference was greater for literary than for nonliterary sentences.

Free and Cued Recall by Sentence Type

For both free and cued recall, sentence type was significant, indicating that the experimental manipulation of imageability and semantic relatedness affected recall performance. First, across both recall tasks, recall was the poorest with low imagery sentences. This is consistent with Paivio's (1971) dual-coding theory, as the majority of these sentences contained abstract nouns (e.g., independence, vanity, fame) that were only amenable to processing with verbal strategies, rather than both verbal and imaginal

processes.

Secondly, the optimal sentence type for recall depended on the type of retention test. For free recall, with a strict criterion, low semantic relatedness sentences were the best; with a lenient criterion, both low semantic relatedness and high imagery sentences were recalled equally well. The proportionately high free recall of low semantic relatedness sentences may be due to their being literary sentences that, presented out of context, were unusual and therefore memorable. As mentioned in a previous analysis, the effect of sentence literariness was marginally significant for lenient free recall, and the mean appeared to be higher for the recall of literary than for nonliterary sentences. Further, of the 16 literary sentences, six were categorized as low on the semantic relatedness dimension. When combined, these findings suggest that, in the free recall task, low semantic relatedness sentences were remembered the best because they were literary sentences that, in the absence of context, were unusual and highly memorable. Further, the optimal free recall of high imagery sentences was consistent with Paivio's (1971) dual-coding theory. That is, these sentences generally contained concrete nouns that could be processed by either verbal or imaginal processes, and thus, were easier to learn and more likely to be remembered.

With cued recall, across both strict and lenient recall

criteria, low imagery sentences were the least frequently recalled. Recall improved with high imagery and low semantic relatedness sentences, and was optimal with high semantic relatedness sentences. As discussed above, the recall of low imagery sentences can be explained by Paivio's (1971) dual-coding theory. Given the superior recall of high semantic relatedness sentences, it is speculated that, when there is some degree of similarity between the two terms in a metaphor, this similarity can be employed to evoke the missing noun when one noun is presented as a cue.

Thus, this implies that when free recall is requested, participants remember those sentences that are unusual, which in this instance was low semantic relatedness literary sentences. However, with the provision of cues, the similarity of the terms in high semantic relatedness sentences is used to trigger memory for the missing noun. This further demonstrates that the demands of the recall test will influence which sentence type is most frequently recalled.

Serial Position Effects

When recall was examined for serial position effects, this variable proved to be significant. With respect to free recall, evidence for primacy and recency effects was found with a strict criterion, and somewhat less robustly for a recency effect with a lenient criterion. Also, the poorer strict recall of items from the middle of the list was

affected by sex and orienting condition such that females in the semantic relatedness condition recalled more than males. This interaction was not anticipated, and is not readily explainable. Further, with the lenient criterion, the advantage of the intentional condition over certain incidental conditions was evident for recall of sentences from the beginning and middle of the list, but not for the end of the list, where orienting condition did not affect recall performance. This implies that intent to learn is critical in the lenient free recall of items from the beginning and middle of the list, but recency effects will occur regardless of orienting condition.

Also, while the young freely recalled more than the old in all list positions, with a lenient criterion this difference increased from the beginning to the end of the set. Examination of Figure 4 indicates that there was a recency effect for the young participants, and a primacy effect for the old adults. Thus, the young and old adults were affected differently by list position effects. This is contrary to previous research in this area. For instance, two early studies (Bromley, 1958; Eisdorfer, Axelrod, & Wilkie, 1963) observed no age differences in serial position effects with free recall. In addition, Arenberg (1976) found that primacy and recency effects, when judged in relation to the asymptote, were equivalent across age groups. Only the asymptote itself was reduced in older adults as compared to

younger adults.

Other studies have specifically investigated the recency effect in free recall, as it is purported to be a measure of primary memory. The results have generally revealed no age differences in recency (Craik, 1968, 1971; Raymond, 1971), although there is contradictory evidence indicating higher recall of recent items by young rather than old adults (Salthouse, 1980). Further, methodological inadequacies have been identified in some of these studies. For example, as Raymond (1971) did not include a group of young adults, the results do not substantiate true age effects.

Thus, the majority of previous research suggests that serial position effects do not vary with the age of the participants. The lack of corroboration in the present investigation may be due to the method of stimulus presentation, as participants were free to rate the sentences in any order they wished, and could return to any items they missed.

Lastly, partial free recall was not influenced by list position, probably because of the small number of instances involved.

Covariates

A few factors were entered as covariates into the recall analyses to determine whether part of the variability in performance might be accounted for by other variables.

The results indicated that the mean ratings performed prior to recall did not affect memory performance. Thus, how a sentence was rated did not predict its subsequent retrievability. Rather, as noted in another analysis, the prior categorizaton of the sentences into certain subgroups had a greater influence on recall performance.

When the years of formal education variable was entered as a covariate, it was significant solely for constructions by older adults in the cued recall task. This suggests that, for seniors, formal education is directly related to the ability to produce novel nouns in the completion of metaphoric sentences. This provides support for the role of individual differences in recall, and connotes that metaphor comprehension and recall may be affected by subject factors that are independent of task manipulations.

This is consistent with the tetrahedral model of learning proposed by Jenkins (1979), in which four factors are presumed to influence learning. These factors are: (a) the learner's activity (e.g., elaboration, rehearsal), (b) the nature of the materials to be learned (e.g., modality, conceptual difficulty), (c) the criterial task (e.g., recall, recognition), and (d) the characteristics of the learner (e.g., skills, attitudes, knowledge), the factor of interest here. It is contended that a complete understanding of learning and memory requires experiments that are sensitive to the interaction of all of these factors.

The score on the analogies test also served as a significant covariate for both the young and old adults across several recall analyses. The correlations were significant for both age groups, but higher for the young than for the old adults. This implies that analogical ability is related to the recall of metaphoric sentences. This extends the position of Sternberg (1977) and Sternberg and Nigro (1983) by arguing that analogical reasoning is also important in metaphor recall.

CHAPTER 4

EXPERIMENT 2

The second experiment was an extension of the first in that two tasks were designed to more directly examine the role of imagery and features in the comprehension and memory of metaphors in young and old adults. Half of the subjects in each age group were randomly assigned to one of two conditions in each of the tasks, and all subjects participated in both tasks. Task 1 always preceded Task 2.

Method

Subjects

Twenty young adults aged 20-33 years (M = 26.35 years) and 20 old adults aged 60-77 years (M = 66.2 years) were recruited in a similar manner as in Experiment 1. They were each paid $10 for their participation. There was an equal number of males and females in each age group. Again, participants completed the same background questionnaire and analogies test as those in the first experiment. The mean age, number of years of education, and score on the analogies task is presented in Table 7 for each age group, sex, context, and instruction condition. An analysis of variance to determine the effects of these four factors on the years of education variable indicated no significant main effects or interactions. A similar analysis on the analogies score revealed no significant main effects. However, the three-way interaction between sex, context, and

instructions was significant, $F(1,24) = 5.678$, $p < .05$. T-tests were conducted to determine the nature of this interaction, but all comparisons between means were non-significant.

With respect to the health variables, participants rated their health on a five-point scale from very poor (1) to very good (5). When asked to rate their own health in comparison to that of their peers, 90% of the young and 100% of the old rated it as good or better. In comparing their health with a perfect state of health, 95% of the young and 100% of the old rated it as good or better. Lastly, 25% of the young and 45% of the old indicated that they were currently taking one or more drugs. As in the first experiment, the older adults appeared to be in very good health.

Task 1

Materials. Of the 60 metaphoric sentences utilized in the first experiment, a subsample of 20 were selected such that half had a high value on the semantic relatedness dimension, and half had a low value on this dimension. These values were determined by the ratings of the 10 young and 10 old participants in the semantic relatedness orienting condition of Experiment 1. The following procedure was employed to determine sentence selection. First, the 10 highest rated and the 10 lowest rated sentences on semantic relatedness were selected from each age group. Those

sentences mutually agreed upon by both young and old adults were included in the second experiment. As additional sentences were still required, an equivalent number of the highest and lowest rated sentences in each age group was added to meet the required number of 10 sentences in each of the two categories.

Context was manipulated between-subjects by presenting the topic and vehicle terms of the sentences in one of two ways: (a) as underlined terms in sentences (in context), or (b) as isolated noun pairs (out of context). Subjects in each context condition were required to perform two tasks, each one on half of the stimulus materials: (a) rate the similarity of the two terms (topic and vehicle) on a 7-point scale, and (b) list as many attributes or properties shared by both terms as they could think of using adjectives to denote the properties (i.e., feature generation). The stimuli were presented in booklets.

Design. Participants were required to: (a) rate 10 noun pairs for the similarity between the two terms, and (b) generate common attributes for another set of 10 noun pairs. Across both tasks, the terms were either presented in context (sentences) or out of context (noun pairs). The order of these two tasks was counterbalanced so that the ratings preceded feature generation for half of the participants, and followed feature generation for the other half of the participants. In addition, the 20 noun pairs

were divided into two sets varying only in their presentation order. That is, those 10 noun pairs that were to be rated in set 1 were utilized in the attribute generation task in set 2, and vice versa for the 10 remaining noun pairs in set 1. Thus, Task 1 was a 2 (age) by 2 (context) by 2 (sex) by 2 (set) by 2 (order) design.

Procedure. Half of the participants in each age group were randomly assigned to one of the two context conditions. They were allowed 15 minutes to complete the feature generation task, and five minutes for the ratings. Written incidental free recall was then tested. Participants were allotted 10 minutes to recall, from the previous two tasks, the 20 topic and vehicle noun pairs that were underlined in sentence context or that were presented as isolated pairs out of context. Testing was conducted individually and in groups of up to five age peers. Within an experimental session, all participants completed the rating and feature generation tasks in the same order with an identical set of noun pairs; only the context of the noun pairs varied between participants. The analogies test was administered after the free recall task, and served as a filler between Task 1 and Task 2. All testing and scoring was performed by a female experimenter.

Task 2

Materials. From the imageability ratings of the 10 young and 10 old adults in Experiment 1, twenty sentences

were selected such that half were rated as high on imageability and half were rated as low on this dimension. The sentences were chosen so as not to overlap with those in Task 1. The same procedure as in Task 1 was employed to select the 20 sentences. The sentences were recorded in a random order on a tape cassette by a male speaker, and auditorily presented with a tape recorder. Each sentence was spoken twice separated by 3 seconds, with a 5 second inter-sentence interval (e.g., Verbrugge & McCarrell, 1977). There were two orienting conditions requiring subjects to either: (a) listen to the sentences to remember as much about them as possible for a subsequent recall test, or (b) form an interacting mental image of the nouns in each sentence to aid memory in a recall test (e.g., Riechmann, 1975, cited in Riechmann & Coste, 1980). Subjects in each condition heard four high imagery practise sentences, and were informed that their subsequent memory for these sentences would not be tested.

Design. Participants listened to a list of 20 metaphoric sentences (10 high imagery and 10 low imagery, presented in a random order) with either standard instructions to remember the sentences, or imagery instructions to facilitate recall. Free recall, cued recall (with topics and vehicles as cues), and recognition of the sentences were tested. Thus, this second task was a 2 (age) by 2 (sex) by 2 (instructions) by 2 (sentence type) by 2

(cue type) by 3 (recall) design with repeated measures on the final three factors.

Procedure. Participants were randomly assigned to one of the two orienting conditions and tested in small groups of young and old age peers. In the imagery condition, participants were instructed to picture the two nouns in each sentence acting together in one image, rather than as separate images. To illustrate this distinction, four metaphoric sentences were verbally presented by the experimenter with both separate and interacting imaginal descriptions. The subjects were then informed that they should listen to the four practise sentences on the tape recorder to try to form an image for the two nouns in each sentence. The participants in the condition without imagery instructions were instructed to listen to the four sentences on the tape recorder to become acquainted with the speaker's voice, and the pace of sentence presentation. Subjects in both conditions were aware that 20 sentences would be presented, and that retention would subsequently be examined. After sentence presentation, free recall was tested, followed by cued recall and a recognition test identical to the procedure employed in Experiment 1. The 20 stimulus sentences were presented in random orders across the three retention tests. Recall was written, and subjects were not allowed to view previous recalls across the memory tests. Ten minutes was allotted for each of the free and

cued recall conditions. In the cued recall test, half of the sentences were cued with topics and half with vehicles. The recognition test comprised 40 sentences (20 previous sentences and 20 distractors) and did not have a time limit. All testing and scoring was conducted by a female experimenter.

Results

Task 1

Semantic Relatedness Ratings

The mean similarity ratings for each noun pair were calculated and then averaged across pairs to yield an overall mean rating. A repeated measures ANOVA was conducted to determine whether the mean ratings for the 10 high semantic relatedness pairs and the 10 low semantic relatedness pairs were influenced by the particular set of pairs rated (set), the order of the rating and feature generation tasks (order), the presentation of the noun pairs in and out of sentence context (context), age, or sex. A 2 (age) by 2 (sex) by 2 (context) by 2 (set) by 2 (order) by 2 (pair type) ANOVA, with repeated measures on the final factor, was conducted on the mean ratings. There was a significant main effect of pair type, $F(1,8) = 477.67$, $p < .05$, suggesting that higher ratings were obtained for the high semantic relatedness pairs ($M = 5.86$) than for the low semantic relatedness pairs ($M = 2.98$). This provides validation for the categorization of the pairs into high and

low subgroups. There were no other significant main effects, or significant interactions.

In order to examine the reliability of the ratings, they were compared with the semantic relatedness (SRL) ratings produced by participants in the first experiment. T-tests were calculated for each noun pair based on the mean ratings of age peers across the two experiments. With respect to the young, when a significance level of $p < .05$ was employed, 25% of the pairs (5/20) were not rated similarly across the two experiments; three of these were low semantic relatedness pairs, and two were high semantic relatedness pairs. For the older group, 15% of the pairs (3/20) were rated differently across the two studies at a criterion level of $p < .05$; two were low semantic relatedness pairs, and one was a high semantic relatedness pair. There was only one pair that was rated differently across the studies by both age groups, and it was a low semantic relatedness pair. These results indicate that the reliability of the noun pair ratings across the two studies was fairly high, and better for the old than the young adults. Further, differences in similarity ratings were more likely for low semantic relatedness pairs than for high semantic relatedness noun pairs.

It was speculated that the variability in the noun pair ratings would decrease from the first to second experiments, as those pairs included in the second study were

specifically selected from the extremes of the semantic relatedness scale in Experiment 1. Variability was examined by pair type (i.e., high versus low semantic relatedness) for each age group separately. Simple inspection of the variances indicated that, with high semantic relatedness pairs, variability decreased across the two experiments for both age groups (from 2.27 to 1.17 for the young, and from 2.40 to 1.31 for the old). When low semantic relatedness pairs were investigated, the variability of the old participants' ratings decreased (from 3.47 to 2.83), while those of the young increased (from 1.45 to 2.18). In summary, variability in the ratings appears to have decreased across experiments with the exception of one age group for one pair type- young adults given low semantic relatedness noun pairs. This suggests that, for the young adults, the similarity ratings for the low semantic relatedness noun pairs were not as extreme as they were for the young adults in Experiment 1. It further confirms the noun pairs in Experiment 2 as exemplars of high and low semantic relatedness pairs when rated by the old adults, and for the high semantic relatedness pairs with young adults.

In summary, the mean similarity ratings were not affected by age, context, sex, order, or set factors. In addition, there was evidence for the reliability and validity of the ratings. With respect to the former, approximately 80% of the 20 noun pairs were rated similarly

across the two experiments by the two age groups. With respect to validity, variability in the ratings decreased across the two experiments, with the exception of the rating of low semantic relatedness pairs by young adults. This decreased variability is to be expected, as the noun pairs selected from the ratings of Experiment 1 were those at the extreme ends of the semantic relatedness scale. Further, the higher ratings obtained for the high semantic relatedness pairs as compared to the low semantic relatedness noun pairs attests to the validity of categorization by these two pair types.

Feature Generation Analyses

Features: Individual Data

The mean number of features produced for each word pair was calculated and then averaged across word pairs to yield an overall feature generation mean. A repeated measures analysis of variance on this variable, examined by pair type, to test the effects of age (2), sex (2), context (2), set (2), and order (2) indicated a significant main effect of age, $F(1,8) = 6.66$, $p < .05$; the young produced more features ($M = 3.69$) than the old ($M = 2.61$). There was also a significant effect of the order of task completion, $F(1,8) = 8.25$, $p < .05$. More features were generated when rating preceded the feature generation ($M = 3.64$) than when the order was reversed ($M = 2.66$). Further, pair type was significant, $F(1,8) = 101.71$, $p < .05$, indicating that more

features were generated with high semantic relatedness noun pairs (M = 3.75) than with low semantic relatedness pairs (M = 2.56). The main effects of sex, context, and set were not significant. However, the four-way interaction between age, context, set, and pair type achieved significance, $F(1,8)$ = 16.64, $p < .05$, as did the interaction between context, sex, order, and pair type, $F(1,8)$ = 5.40, $p < .05$. As it was believed that the small cell size with these interactions rendered them of little practical significance, they were not examined further. There were no other interactions achieving significance.

Synonyms and Errors: Individual Data

An examination of the features generated by each participant indicated that some of the words produced did not express novel adjectives, but were similar ways of stating an identical meaning. Using a dictionary, a separate category was created for those words that expressed a true synonym of other words generated. The mean number of these synonyms produced was examined in a 2 (age) by 2 (sex) by 2 (context) by 2 (order) by 2 (set) by 2 (pair type) analysis of variance with repeated measures on the last factor. There were no significant main effects, but one significant interaction between age and the order of task completion, $F(1,8)$ = 6.04, $p < .05$. T-tests indicated no difference in the number of synonyms generated by either age group, regardless of the order of task completion. Simple

inspection of the means suggests that for the young, more synonyms were produced when rating followed the generation task (*M*s = .10 and .04 for the two orders), while the old produced more synonyms when rating preceded the feature generation task (*M*s = .10 and .03 for the two orders).

A category of errors was also examined. This variable referred to the production of inappropriate features not common to both members of the word pair presented. For example, one participant wrote the features tape, book, and paper for the word pair history and sponge. Another individual wrote the word appetite as a feature common to a table and a four-legged animal. There were very few instances of this category, as words had to be obviously incorrect to be scored as errors. A repeated measures analysis of variance on this variable with age (2), sex (2), context (2), set (2), order (2), and pair type (2) as factors indicated a significant main effect of age, $F(1,8) = 7.04$, $p < .05$; the old ($M = .18$) produced more errors than the young ($M = .02$). There was also a significant interaction between age, context, and set, $F(1,8) = 8.14$, $p < .05$. T-tests indicated that more errors were produced in set 2 ($M = .12$) than in set 1 ($M = .00$) for old adults given a noun pair context. All other main effects and interactions failed to attain significance.

Features: Group Responses

Total number of different responses. The features produced for each presented word pair were analyzed as a group for the young and old participants responding to each question. For each of the 40 pairs (20 in sentence context, and 20 in word pair context), five young adults and five older persons had the opportunity to produce features; participants did not always produce features for every noun pair they were presented.

The first variable examined was the number of different features provided by each age group. That is, for each word pair, the number of different responses produced by the five young participants was contrasted with the number of such responses produced by the five old persons. A t-test calculated for this age effect across word pairs was significant, $t(38) = 2.89$, $p < .05$. The young as a group produced a greater variety of responses ($M = 15.65$) than the older adults ($M = 11.03$). A t-test calculated to determine the effect of context failed to attain significance.

Mean correct responses. Another variable examined for each pair was the mean number of correct responses produced (i.e., features and synonyms) by each age group. A significant effect of this variable across all pairs was obtained, $t(38) = 3.26$, $p < .05$. Again, the young produced more correct responses ($M = 4.02$) than the old ($M = 2.83$). When context was examined across word pairs and age groups,

a t-test indicated no difference in the mean number of correct responses produced given noun pairs presented in and out of sentence context.

Percentage of responses: Most frequent and two largest categories. In previous research, Johnson and Malgady (1979) have measured response saliency by examining the percentage of responses accounted for by the five most frequent response categories. In their research with young adults, this variable was correlated with other measures of word association and metaphor interpretation. They found that, as predicted by semantic feature theory, all of the variables were significantly interrelated. In this study, the limited number of participants and word pairs made it possible to investigate only the most frequent response category, and the two largest response categories.

With respect to the former, the most frequently produced adjective for each word pair was recorded for the young and old participants separately. Each was examined in both age groups to probe for age effects in the frequency of its production. As percentages were being employed in the following analyses, it was deemed necessary to first equate the age groups for the number of responses produced. Several word pairs were discarded because t-tests on the mean number of correct responses produced for each pair indicated a significant effect of age; twenty-six word pairs (out of a possible 40 pairs given two context conditions) remained.

The most common response produced by the young adults for each of the 26 word pairs, expressed as a percentage of the total responses, was examined for each age group. These percentages were averaged across the word pairs to yield the overall mean percentage of responses accounted for by the most common response. When a t-test was employed to compare the mean percentages of the young adults' response across both age groups, a significant effect was obtained, $t(38) = 2.76$, $p < .05$ revealing that the young produced the response proportionately more often ($M = 16.46\%$) than the old ($M = 9.92\%$). An identical analysis was conducted for the most common response produced by the older adults for each word pair, again examined for each age group. A t-test on the mean percentages across the noun pairs achieved significance, $t(38) = 2.26$, $p < .05$. The percentage of responses accounted for by the largest category was greater for the old ($M = 15.41\%$) than for the young ($M = 9.77\%$). This suggests that: (a) the largest response category for each age group accounted for only approximately 16% of the total number of responses, and (b) the most common response tended to be different for the two age groups when averaged across the 26 word pairs.

Also examined was the percentage of responses accounted for by the two largest categories in combination. Utilizing a t-test, these categories were examined for the young, and then compared with the frequency of their occurrence for the

old across the 26 pairs. The difference was significant, $t(38) = 4.15$, $p < .05$, indicating that the percentage of responses accounted for by the two largest categories for the young was greater for them ($M = 27.89\%$) than when considered for the old ($M = 15.06\%$). This indicated that the two most common responses for the young adults were not as frequent for the old adults.

When the two largest categories were examined with respect to the older adults, and then compared with their frequency of response in the young, a significant finding emerged, $t(38) = 4.50$, $p < .05$ suggesting that the percentage of responses accounted for by the two largest categories provided by the old was greater for them ($M = 27.51\%$) than for the young ($M = 14.53\%$).

These analyses provide some suggestive evidence that, given a set of word pairs, the salient features produced by young and old adults are not identical.

In summarizing the results of the feature generation task, analyses of the individual data indicated that: (a) more features were produced for high semantic relatedness pairs than for low semantic relatedness noun pairs, (b) more features were generated when rating preceded feature generation than when the task order was reversed, and (c) the young listed more features than older adults. With respect to synonyms, the age by order interaction revealed that the young produced more synonyms when generation

preceded the rating task, while the old provided more synonyms when the order was reversed. In addition, the old committed a greater number of errors than the young participants, and more for set 2 than for set 1 given a noun pair context.

The group analyses demonstrated that the young produced more correct responses, and a greater variety of features, than the old participants. The effect of context was not significant in either of these analyses. Further, the most frequent, and two largest response categories accounted for 16% and 27.70% of the total number of responses, respectively. Lastly, the most common features produced by the young and old adults were not analogous.

Correlations Between Ratings and Features Generated

The semantic feature theory asserts that the degree of similarity between two words is a function of the number of features that they share. Thus, it predicts that the rated similarity of noun pairs should correlate positively with the number of properties (features) the words have in common (e.g., Johnson & Malgady, 1979). In order to examine this relationship, the data were recoded with word pair as a between-subjects variable. Pearson product-moment correlations between the ratings and features across the 20 pairs were computed for each age group. The resulting coefficients were positive and highly significant for the young ($r = .52$) and old ($r = .52$) groups considered

separately and simultaneously (r = .43). When the correlations between ratings and features were examined by pair type, the coefficients for the high and low semantic relatedness sentences were not significant. This provides some support for the tenets of semantic feature theory, and replicates the highly significant correlation between interword similarity and number of common properties (r = .43) found for young adults in previous research (Johnson & Malgady, 1979).

Free Recall

The free recall task was scored in the same manner as the free recall in Experiment 1. Strict, lenient, and partial recall of the nouns was examined. The mean number of noun pairs recalled for each age group in each context condition given a strict and a lenient criterion is presented in Table 8.

A 2 (age) by 2 (sex) by 2 (context) by 2 (set) by 2 (order) by 2 (pair type) repeated measures analysis of variance was conducted on the strict free recall of the high semantic relatedness and low semantic relatedness noun pairs. There was a signifiant main effect of age, $F(1,8) = 10.82$, $p < .05$, connoting that the young ($M = 8.35$) recalled more pairs than the old ($M = 5.10$). There were no other significant main effects. However, there was one significant interaction between context, set, and sex, $F(1,8) = 6.93$, $p < .05$. T-tests indicated that strict free recall was better

for set 2 (M = 8.75) than for set 1 (M = 4.50) for females given noun pairs in a sentence context.

When lenient recall was examined in a similar analysis, the main effect of order was significant, $F(1,8) = 9.45$, $p < .05$, in addition to the age effect, $F(1,8) = 33.41$, $p < .05$. Recall was better when feature generation preceded the rating task (M = 10.35) than when the order was reversed (M = 7.80), and again, the young recalled more (M = 11.45) than the old (M = 6.70). The age by order interaction also attained significance, $F(1,8) = 12.72$, $p < .05$ (see Figure 5). The old performed equivalently across both orders (Ms = 6.70). However, the young recalled more when the feature generation task preceded the rating (M = 14.00) than in the reverse order (M = 8.90). This interaction indicates that lenient recall of the young was influenced by the order of the tasks, while that of the old was not.

There was another interaction to attain significance, that between context and set, $F(1,8) = 5.85$, $p < .05$. T-tests, however, indicated that there was no difference in the recall of noun pairs given set 1 or set 2 with a sentence or noun pair context. Examination of the means suggests that noun pairs in set 1 were recalled better with a noun pair context (M = 10.8) than with a sentence context (M = 7.8). Those in set 2 did not differ given a sentence (M = 9.5) or noun pair (M = 8.2) context. Lastly, the four-way interaction between age, set, order, and sex was

significant, $F(1,8) = 10.49$, $p < .05$. However, it was believed to be noninterpretable, and not investigated further.

Partial recall was also investigated for the effects of age (2), sex (2), context (2), set (2), order (2), and pair type (2) in a similar repeated measures analysis of variance. The age effect was significant, $F(1,8) = 10.92$, $p < .05$; the old ($M = 1.00$) had more partial recalls than the young ($M = .15$). The effect of set was also significant, $F(1,8) = 5.73$, $p < .05$. That is, there was greater partial recall with the second set ($M = .85$) than with the first set ($M = .30$). In addition, an age by set interaction obtained significance, $F(1,8) = 6.87$, $p < .05$. Set was not a factor in the partial recall of the young (Ms = .20, .10 for sets 1 and 2, respectively); for the old, partial recall was better with set 2 ($M = 1.60$) than with set 1 ($M = .40$). This indicates that the partial recall of old adults was affected by the particular set of noun pairs presented, but not for the recall of young adults.

Further, there was a significant interaction between sex and pair type, $F(1,8) = 10.71$, $p < .05$. However, t-tests indicated that partial recall of high and low semantic relatedness pairs did not differ for males or females. Inspection of the means suggests that males ($M = .40$) partially recalled more than females ($M = .15$) with high semantic relatedness pairs. With low semantic relatedness

noun pairs, females (M = .40) recalled more than males (M = .20). Several three-way and four-way interactions with pair type were significant, but the small number of partial recalls renders the investigation of these interactions impractical.

In summary, young adults recalled more noun pairs than old adults with both strict and lenient free recall criteria. Only with partial recall did the old perform better than the young participants. There was one significant interaction between context, set, and sex for the strict criterion, indicating that recall was better for females given set 2 rather than set 1 with a sentence context. When a lenient criterion was employed, the main effect of order revealed that recall was higher when feature generation preceded the rating task than when the order was reversed. Further, the age by order interaction demonstrated that order was not a factor in the recall of the old, but was for the young; the lenient recall of the young participants was better when the generation task preceded similarity ratings. In addition, the interaction between context and set suggested that lenient recall of set 1 pairs was better with a noun pair than a sentence context. In contrast, the recall of set 2 pairs did not differ by context.

Lastly, when partial recall was investigated, recall was better with set 2 than with set 1. The age by set

interaction connoted that partial recall was equivalent across the sets for young adults. However, with old adults, recall of set 2 was better than that of set 1. There was also a significant relationship between sex and pair type such that males partially recalled more than females with high semantic relatedness pairs, while the opposite was apparent with low semantic relatedness pairs.

Ratings and Recall

In order to examine whether there was a relationship between the similarity ratings provided and subsequent free recall, the overall mean rating was entered as a covariate into all recall analyses. That is, the effects of age (2), sex (2), context (2), set (2), and order (2) were investigated in three separate analyses of variance on strict free recall, lenient free recall, and partial recall. However, the covariate was not significant for any of the recall measures. The covariate did change the pattern of results previously obtained for lenient and partial free recall, but not for the strict criterion. That is, with lenient free recall, the interaction between context and set was no longer significant. Further, there was no main effect of set in partial free recall.

The mean rating covariate was examined by pair type to yield mean ratings for high and low semantic relatedness pairs separately. These were both used as covariates in three additional repeated measures analyses testing the

effects of age, sex, context, set, and order on the strict free recall, lenient free recall, and partial recall of the high and low semantic relatedness pair types. The effect of the covariates was not significant in any of these analyses. However, the covariates altered the results previously observed with lenient free recall. There was no effect of order, and the interaction between context and set failed to attain significance. The strict and partial free recall analyses did not change with the addition of the covariates.

These findings indicate that the effect of the similarity ratings as covariates was not significant, and that the pattern of previously obtained ANOVA results remained essentially unchanged. That is, the main effect of age, and the interactions of age with set and order were still significant.

Feature Generation and Recall

The mean number of features produced was also examined as a separate covariate. Again, three analyses of variance were conducted for the three types of free recall; the five between-subjects factors were identical to those above, and the mean number of features produced served as the covariate. However, it was not a significant variable in any of the recall analyses. It did change the pattern of results previously obtained with lenient and partial free recall, but not with strict free recall. That is, the three-way interaction between age, set, and sex achieved significance

with a lenient criterion, $F(1,7) = 5.75$, $p < .05$. Simple inspection of the means indicated no difference in the lenient recall of set 2 noun pairs by the two sexes within each age group (Ms = 11.4, 11.8 for young males and females, respectively; Ms = 6.0, 6.2 for old males and females, respectively). With respect to set 1, for young adults, more noun pairs were leniently recalled by males (M = 13.4) than by females (M = 9.2). The reverse was apparent for the old adults, as females (M = 7.8) recalled more noun pairs than males (M = 6.8). With partial recall, there was no effect of set, and the interaction between age and set was no longer significant.

The features generated were also analyzed by pair type, and examined as covariates for the three types of free recall with high and low semantic relatedness pairs. The covariates were not significant for the three repeated measures ANOVAs conducted. However, the covariates did change the pattern of previously observed effects with strict and lenient free recall, but not for partial recall. Specifically, the main effect of age and the interaction between context, set, and sex were no longer significant for strict free recall. With a lenient criterion, there was no age effect, no interaction between age, order, set, and sex, and only a marginally significant interaction between context and set, $F(1,6) = 5.52$, $p = .057$. Further, there was a significant interaction between age, set, and sex, $F(1,6)$

= 7.77, $p < .05$, that resembled the pattern of effects reported for the same interaction in the preceding paragraph. Lastly, the interaction between context, set, and sex was also significant for lenient free recall, $F(1,6) = 7.39$, $p < .05$.

Thus, it can be concluded from these analyses that, although the feature production task was not a significant covariate, it altered the pattern of results previously described in analyses of variance. Of particular interest is the absence of an age effect obtained here with strict and lenient free recall. The insignificance observed for some of the other effects was regarded as less central to the main hypotheses under investigation. Previous results that remained significant in this analysis of covariance were the main effect of order, and its interaction with age in lenient free recall, and the main effect of age in partial recall.

Covariates

Years of Education

The years of education completed by each participant was entered as a covariate into several recall analyses testing the effects of age, sex, set, context, and order. There was no significant effect of the covariate for analyses of strict, lenient, and partial free recall, or for these analyses examined by pair type. However, the covariance adjustment changed the nature of the results

obtained in each analysis. With strict free recall, the interaction between context, set, and sex was only marginally significant, $F(1,7) = 5.14$, $p = .058$. When analyzed by pair type, this three-way interaction became nonsignificant, and the interaction between age and task order achieved significance, $F(1,7) = 6.43$, $p < .05$. Lenient recall examined by pair type indicated that the interaction between context and set was no longer significant. With respect to partial free recall, set was no longer significant, and the interaction between age and set was only marginally significant, $F(1,7) = 5.42$, $p = .053$.

These results indicate that years of education was not a significant covariate in the recall analyses. Although it produced minor changes in the findings of previous analyses of variance, the main effects of age and order, and the interaction between these two variables, remained significant.

Analogies

The score on the analogies task was also entered as a covariate in the recall analyses. It was found to be significant for strict free recall in a 2 (age) by 2 (sex) by 2 (context) by 2 (set) by 2 (order) analysis of variance, $F(1,7) = 5.77$, $p < .05$. It was also significant for a similar analysis of lenient free recall, $F(1,7) = 6.54$, $p < .05$. However, analogies was not a significant covariate for partial free recall, or for any of these dependent variables

when investigated by pair type in repeated measures analyses.

Further, these covariance analyses altered the pattern of results previously obtained for each recall variable. With respect to strict free recall alone and by pair type, the three-way interaction between context, set, and sex was no longer significant. The lenient free recall analysis indicated that the interaction between context and set was nonsignificant, and, when examined by pair type, there was no main effect of order. Lastly, partial recall alone and by pair type revealed a nonsignificant effect of set. The age by set interaction was only marginally significant, $F(1,7) = 5.47$, $p = .052$, and became nonsignificant when analyzed by pair type.

These results suggest that analogical reasoning, as measured with the present analogies test, was related to the strict and lenient free recall of the noun pairs. Further, the pattern of results previously obtained in analyses of variance was not dramatically changed, as the main effects of age, and the interaction between age and task order, remained significant. Lastly, the correlation between the analogies score, and strict and lenient free recall was .34 and .24, respectively, significant only for the former. When examined by age group, the correlations for strict and lenient free recall with analogies were higher for the old than for the young, but were not significant.

Task 2

Recall Analyses

The free and cued recall tasks were scored in an identical manner to the first experiment. The mean number of sentences recalled across the three levels of recall for the two age groups and instruction conditions is displayed in Table 9.

Free Recall, Cued Recall, and Recognition

The effects of age, sex, and instructions on strict free recall, cued recall, and recognition were examined in an ANOVA with repeated measures on the recall variable. A significant main effect of age emerged, $F(1,32) = 30.48$, $p < .05$; the young ($M = 13.38$) recalled more than the old ($M = 9.68$). Recall across the three levels was also significant, $F(2,31) = 524.90$, $p < .05$. T-tests indicated that recall improved significantly for free recall ($M = 4.68$), cued recall ($M = 10.4$), and recognition ($M = 19.53$) tasks. An age by recall interaction attained significance, $F(2,31) = 16.63$, $p < .05$ (see Figure 6). As revealed by t-tests, the young recalled more sentences than the old across all three levels of recall. However, the difference in recall between the two age groups was least with recognition, more for free recall, and was the greatest for cued recall. The main effects of sex and instructions were not significant, and there were no other significant interactions. An identical analysis conducted with the lenient recall criterion

connoted similar results.

Proportional Recall Gain

As in the first experiment, proportional gains in recall from free to cued, and from cued to recognition tests were examined for potential age differences. An ANOVA with repeated measures on these two levels of the "gain" variable was conducted to test the effects of age, sex, and instructions with each scoring criterion. There was a significant main effect of age, $F(1,31) = 20.21$, $p < .05$, $F(1,31) = 27.56$, $p < .05$, for the strict and lenient criteria, respectively. There was greater proportional improvement in performance across the three recall tasks for the old ($M = 3.41$) than for the young ($M = 2.03$) with a strict criterion and a lenient criterion ($Ms = 2.35$, 1.62 for old and young adults, respectively). There were no other significant main effects or interactions with a strict criterion. However, the gain variable was significant for the lenient criterion, $F(1,31) = 5.21$, $p < .05$; there was more proportional improvement from free to cued recall ($M = 2.18$) than from cued recall to recognition ($M = 1.76$). This was the only other effect to obtain significance with the lenient criterion.

In summary, these analyses suggest that the old improved proportionally more than the young across the three recall tasks with both scoring criteria. Further, the proportional gain variable was significant with a lenient

recall criterion, revealing that recall performance improved proportionally more from free to cued recall than from cued recall to recognition.

Cued Recall: Topic and Vehicle Cues

The mean number of sentences recalled with topic and vehicle cues is presented in Table 10 for both age groups and instruction conditions. A 2 (age) by 2 (sex) by 2 (instructions) by 2 (cue type) ANOVA with repeated measures on the final factor was performed utilizing a strict recall criterion. A significant main effect for age was evident, $F(1,32) = 34.88$, $p < .05$. The young ($M = 6.83$) recalled more than the old ($M = 3.58$). In addition, there was a significant effect of cue type, $F(1,32) = 33.12$, $p < .05$, demonstrating that recall was better with vehicle cues ($M = 6.1$) than with topic cues ($M = 4.3$). Repeating the analysis with a lenient recall criterion obtained a similar age effect, $F(1,32) = 32.88$, $p < .05$, but no effect of cue type. There were no other significant main effects or interactions in these analyses.

Cued Recall: Omissions, Intrusions, and Constructions

The mean number of omissions, intrusions, and constructions for each age group and instruction condition is presented in Table 11. A repeated measures MANOVA was performed on these three variables by cue type to examine the effects of age, sex, and instructions. As in the previous analysis of cued recall, the age effect was

significant. In addition, the main effect of the recall variable achieved significance, $F(2,31) = 64.37$, $p < .05$. T-tests revealed that there were more omissions ($M = 5.58$) than constructions ($M = 1.43$), and a greater number of constructions than intrusions ($M = .23$) in the cued recall task. There were no other significant main effects. A significant interaction between age and recall variable emerged, $F(2,31) = 12.50$, $p < .05$. T-tests demonstrated that there were more omissions for old adults ($M = 8.35$) than for young adults ($M = 2.80$). There were no other interactions obtaining significance.

Partial Free Recall

The same categories of partial recall were scored as in Experiment 1. The influences of age, sex, instructions, and sentence type on the partial recall of high and low imagery sentences were studied in a repeated measures analysis of variance. There was a significant main effect for sex, $F(1,32) = 4.22$, $p < .05$; females ($M = 1.80$) had more partial recalls than males ($M = .95$). A significant interaction between instructions and sentence type also emerged, $F(1,36) = 4.35$, $p < .05$. T-tests indicated that there was no difference in the partial recall of high ($M = .65$) and low ($M = .60$) imagery sentences given imagery instructions. However, with standard instructions, partial recall was better for low imagery ($M = 1.15$) than for high imagery ($M = .35$) sentences. No other main effects or interactions

attained significance.

Recognition Errors

As previously described in the first experiment, there were two types of errors possible on the recognition test: (a) incorrectly identifying an item that was not presented (false alarm), and (b) incorrectly rejecting an item that was presented. A MANOVA was employed to test the effects of age, sex, and instructions on these two types of errors. However, there were no significant main effects or interactions of these factors on recognition errors. The fact that recognition performance was almost perfect in all conditions may explain the absence of any significant findings here.

Free and Cued Recall by Sentence Type

The strict and lenient free and cued recall of high imagery and low imagery sentences was investigated in four 2 (age) by 2 (sex) by 2 (instructions) by 2 (sentence type) ANOVAs with repeated measures on the final factor. The results of each analysis will be discussed in turn.

Utilizing a strict free recall criterion, a significant age effect emerged, $F(1,32) = 17.41$, $p < .05$. The young (M = 3.30) recalled more than the old (M = 1.38) adults. In addition, there was a significant interaction between instructions and sentence type, $F(1,32) = 4.35$, $p < .05$. Subsequent t-tests indicated that low imagery sentences (M = 2.8) were recalled better than high imagery sentences (M =

1.85) with imagery instructions; there was no difference in the recall of high (M = 2.45) and low (M = 2.25) imagery sentences with standard instructions. There were no other main effects or interactions obtaining significance.

However, when a lenient free recall criterion was employed, there was a similar age effect, but the nature of the interaction between instructions and sentence type was altered. There was no difference in the lenient recall of high (M = 3.30) and low (M = 3.45) imagery sentences with imagery instructions. However, with standard instructions, high imagery sentences (M = 3.70) were recalled better than low imagery sentences (M = 2.75). No other significant interactions emerged.

The strict cued recall analysis of these two sentence types yielded a significant effect of this repeated measures variable, F (1,32) = 13.88, p < .05, connoting that high imagery sentences (M = 5.75) were recalled better than low imagery sentences (M = 4.65). The effect of age was also significant, as indicated in previous analyses of cued recall. In addition, there was a significant interaction between instructions and sentence type, F(1,32) = 4.85, p < .05. T-tests demonstrated that the cued recall of high imagery (M = 5.7) and low imagery (M = 5.25) sentences did not differ with imagery instructions. However, with standard instructions, high imagery sentences (M = 5.80) were recalled better than low imagery sentences (M = 4.05) in the

cued recall task. The other main effects and interactions did not achieve significance.

When a lenient cued recall criterion was employed, identical main effects of age and sentence type were obtained. However, the nature of the interaction between instructions and sentence type was modified. High imagery sentences (M = 7.45) were recalled better than low imagery sentences (M = 5.95) with imagery instructions, and with standard instructions (Ms = 7.40 and 4.75 for high and low imagery sentences, respectively). There were no other significant main effects or interactions.

In summary, the analyses of free and cued recall by sentence type (i.e., high versus low imagery) yielded a significant effect of this variable for strict and lenient cued recall. Specifically, high imagery sentences were remembered better than low imagery sentences in the cued recall task. Further, the interaction between instructions and sentence type was significant across all recall analyses, but the nature of this interaction was not uniform. With standard instructions, high imagery sentences were recalled better than low imagery sentences for both cued recall criteria, and for lenient free recall. However, there was no difference in the strict free recall of high and low imagery sentences with standard instructions, and low imagery sentences were recalled better than high imagery sentences with partial recall.

With imagery instructions, the pattern of results was still not clear. Low imagery sentences were recalled better than high imagery sentences in the strict free recall situation. With lenient free recall, strict cued recall, and partial recall, there was no difference in the recall of the two sentence types. Lastly, with lenient cued recall, high imagery sentences were remembered better than low imagery sentences.

Serial Position Effects

As in Experiment 1, the effect of list position presentation on subsequent recall was examined with respect to two memory concepts: primacy and recency effects. As stated earlier, primacy effects refer to memory for items presented at the beginning of a list, and recency to memory for items presented at the end of the list. In this instance, the set of 20 sentences was partialled into three sections: the first six sentences tested a primacy effect, the next eight metaphors comprised the asymptote, and the last six sentences represented a recency effect. Recall was measured as a proportion of these totals for each list position.

Employing a strict free recall criterion, an ANOVA with repeated measures on list position was performed to test the effects of age (2), sex (2), and instructions (2). The effect of age was significant, $F(1,32) = 18.13$, $p < .05$. Again, the young ($M = .34$) recalled more than the old ($M =$

.14). List position was also significant, $F(2,31) = 3.35$, $p < .05$. T-tests indicated that recall of sentences from the beginning of the list ($M = .26$) was better than for recall of sentences in the middle of the list ($M = .20$). There was no difference in recall of sentences from the beginning and end ($M = .25$) of the list, or from the middle and end of the list. Further, the interaction between age, instructions, sex, and list position was significant, $F(2,31) = 6.28$, $p < .05$. However, it was deemed to be noninterpretable and not examined further. Similar effects for age, and the four-way interaction were corroborated in an identical analysis with a lenient free recall criterion. However, the effect of list position was not significant, and there were no other effects achieving significance.

When partial recall was examined, a significant effect of list position was evident, $F(2,31) = 4.79$, $p < .05$. T-tests revealed that sentences at the end of the list ($M = .10$) were partially recalled better than sentences at the beginning of the list ($M = .04$). There was no difference between the partial recall of sentences at the beginning or in the middle ($M = .07$) of the list, and no difference for sentences presented in the middle or at the end of the list. The only other significant effect was that of sex, previously reported. There were no significant interactions.

In summary, evidence for a primacy effect was obtained with strict free recall, as sentences from the beginning of

the list were remembered better than those in the middle of the set. There was some suggestion of a recency effect with partial recall, as sentences from the end of the list were partially recalled better than those at the beginning of the set. There was no effect of list position on lenient free recall.

Covariates

Years of Education

In order to determine whether demographic characteristics might account for part of the variability in the recall measures, years of education was entered as a covariate into several of the analyses previously conducted. As the presence of the covariate did not change the pattern of results in most analyses, reference will only be made to those instances in which it did alter the outcome of previous analyses. When the three levels of strict and lenient recall were examined in repeated measures ANOVAs to test for the effects of age, sex, and instructions, years of education proved to be a nonsignificant covariate. The same analysis on cue type, however, yielded a significant effect of the covariate for the lenient recall criterion, $F(1,31) = 4.25$, $p < .05$. There was also a significant effect of the covariate for constructions with topic and vehicle cues, $F(1,31) = 4.61$, $p < .05$, which resulted in a significant interaction between instructions and cue type, $F(1,32) = 5.84$, $p < .05$. This covariate was confirmed to be

significant in an analysis of covariance on constructions, $F(1,31) = 4.82$, $p < .05$. Education was not a significant covariate for omissions, intrusions, or partial recall, but it rendered the sex variable nonsignificant in the latter analysis.

This indicates that, for lenient cued recall with topic and vehicle cues, and constructions, the years of education variable accounted for some of the variability in performance. Further, the covariance adjustment did not affect the results previously obtained in analyses of variance. That is, the main effects of age, recall task, and cue type, and the interaction between age and recall task, remained significant. In addition, the correlation between the years of education variable and constructions was -.34 (which was significant), and .25 between lenient cued recall and the education variable ($p = .059$). When these correlations were examined by age group, constructions was significant for the old ($r = -.47$), but not for the young ($r = -.15$), while lenient cued recall was significant for the young ($r = .38$), but not for the old ($r = .33$). This further suggests that years of formal education was related to the production of constructions for the old participants, and the lenient cued recall of young adults.

Analogies

Pearson product-moment correlations were computed between the analogies variable and several recall measures

for young and old participants separately. These correlation coefficients are presented in Table 12. All of the correlations are positive, with the exception of omissions, indicating that a higher analogies score was related to better recall performance. Further, there are more significant correlations for the old adults than for the young adults. The score on the analogies task was also entered as a covariate in the repeated measures recall analyses. It proved to be significant in the majority of cases, indicating that part of the variance in recall performance could be attributed to scores on the analogies test. Only those results affected by the covariance adjustment will subsequently be reported.

For instance, the analogies score was a significant covariate in the analysis of strict and lenient recall over the three recall tasks, $F(1,31) = 18.18$, $p < .05$ and $F(1,31) = 18.10$, $p < .05$, respectively. Further, the interaction between instructions and sex became significant with a lenient criterion, $F(1,31) = 4.50$, $p < .05$. The covariate was also significant for strict recall with topic and vehicle cues, $F(1,31) = 20.07$, $p < .05$, and for omissions by cue type, $F(1,31) = 12.34$, $p < .05$. In addition, it was significant for partial recall, $F(1,31) = 4.56$, $p < .05$, and for lenient recall given the two cue types, $F(1,31) = 19.90$, $p < .05$, However, the main effect of sex was rendered insignificant in the former analysis, and the interaction

between sex, instructions, and age became significant in the latter analysis, $F(1,31) = 4.28$, $p < .05$. Simple inspection of the means suggested that, in the absence of imagery instructions, there was no difference between the sexes in the lenient recall of young adults (*M*s = 7.8, 8.1 for males and females, respectively), or of old adults (*M*s = 3.9, 4.5 for males and females, respectively). However, given imagery instructions, the lenient recall of the young adults did not differ by sex (*M*s = 8.3, 7.7 for males and females, respectively), while for the older adults, males (*M* = 6.5) recalled more than females (*M* = 4.3). Lastly, the analogies score was not a significant covariate for constructions or intrusions.

With respect to recall analyses by sentence type, there was a significant effect of the covariate for the free recall of high and low imagery sentences when a lenient criterion was employed, $F(1,31) = 6.17$, $p < .05$, and when a strict scoring criterion was utilized, $F(1,31) = 5.55$, $p < .05$. However, this adjustment did not affect the pattern of results previously observed. Given these results, it can be concluded that the analogies task serves as a significant covariate in numerous recall analyses, suggesting that the ability to successfully complete analogies is related to the recall of metaphoric sentences. Further, the covariance adjustment did not affect the results previously described in analyses of variance, as the main effects of age, recall

task, and cue type, and the interaction between age and recall task, remained significant. In addition, from the correlations above, it would appear that the analogies score serves as a significant covariate particularly for recall by the old participants.

Discussion

Task 1

Similarity Ratings

The results of the ratings of high and low semantic relatedness noun pairs indicated no effects of age, sex, order, set, or context. This contrasts with the age effect found for the semantic relatedness dimension in Experiment 1, where the old had higher ratings than the young participants. However, it is also consistent with previous research indicating equivalent similarity ratings for noun pairs presented in and out of sentence context (Johnson & Malgady, 1979). Further, there was evidence for the validity and reliability of the ratings when compared with the semantic relatedness ratings in the previous experiment. Also, consistent with semantic feature theory, high semantic relatedness noun pairs received higher ratings than low semantic relatedness noun pairs.

Feature Generation

The findings denote that the young generated more features than the old. This may simply be an artifact of the majority of the young adults being students, and therefore

accustomed to the demands of written verbal production tasks. Further, when the features were analyzed for the group of participants responding to each noun pair, the young produced a greater variety of responses than the old adults. They also generated more mean correct responses than old participants when incorrect responses were excluded from the analyses.

The features were also analyzed as group data for the mean percentage of responses accounted for by the most frequent and two largest categories for selected noun pairs; the responses of the young and old adults were observed to differ. This is an interesting finding, because it suggests that the common attributes shared by two nouns may be perceived differently by young and old adults. That is, their representations of similarity may not be analogous. When considered with the rating analyses, the implication is that young and old adults may rate the similarity between two nouns equivalently, but the basis for this similarity may, in actuality, be quite different.

Lastly, the most frequent and two largest response categories across both age groups accounted for 16% and 28% of the total number of responses, respectively. As measures of saliency, they are relatively low, indicating that there may be several attributes that are fairly common to the two terms in each noun pair, and that no one aspect of similarity is prevalent.

There was no main effect of context in any of the feature analyses. This connotes that a sentence context does not influence the generation of adjectives common to a noun pair. It is possible that a methodological artifact may be responsible for this result. That is, in the presentation of the stimulus materials, the nouns were underlined in each sentence. This may have had the effect of de-emphasizing the sentence context.

The main effect of order was significant in the individual feature generation analyses. More features were generated when rating preceded the feature generation task than when the order was reversed. This may be because having already rated the similarity of noun pairs, participants were introduced to the type of noun pairs they would be producing features for, and the notion of semantic similarity upon which both tasks were predicated.

Further, given the individual feature data, there was some support for feature theory in the finding that more features were generated for high semantic relatedness pairs than for low semantic relatedness pairs. That is, as predicted by the theory, the greater the similarity between two objects, the more features they should have in common.

When the individual feature data were examined for synonyms and errors, it was found that the young produced more synonyms when generation preceded the rating task, than in the reverse order. In contrast, the old generated more

synonyms when rating preceded the generation task than when the opposite order prevailed. However, the few number of synonyms produced renders the significance of this result moot. With respect to errors, the old committed more errors than the young, and more for set 2 than for set 1 with a noun pair context. Again, there were very few errors, intimating that these findings must be interpreted with caution.

When the correlation between ratings and features was examined, support for semantic feature theory was obtained. That is, similarity ratings were significantly and positively correlated with the number of shared features generated by both young and old participants. This replicates earlier findings by Johnson and Malgady (1979) with young adults, and extends this phenomena to older adults. When examined by pair type, the correlations were not significant, probably because the small number of noun pairs of each type (i.e., ten) elevated the significance level such that a very high correlation would be needed to achieve significance.

Recall

The results indicated that the young recalled more noun pairs than the old for both strict and lenient incidental free recall. This reveals that the typical memory deficits obtained for the old when compared with the young were also apparent in the present investigation. Only with partial

recall did the old recall more than the young adults. That is, they were more likely to remember one member of a pair, or to match single nouns across different pairs, suggesting that their memory representations were not as complete as the young participants in many instances. With strict free recall, there was also a three-way interaction between sex, context, and set connoting that recall of set 2 noun pairs was greater than those in set 1 for females given a sentence context. Set was not anticipated to be significant as a main effect or in interaction with other variables, and thus this three-way interaction was contrary to expectation.

When lenient free recall was examined, there was an effect of order, such that recall was better when generation preceded the rating task than when the order was reversed. This is probably because the rating task was completed in a short amount of time and subjected to less processing than the generation task. As a consequence, it was more likely to be influenced by interference when it was followed by the feature generation task. Thus, in order to maximize retrieval across both tasks, the better order was the one in which there was no interference from the generation task.

There was also an interaction with age, revealing that recall by the old was not influenced by task order, while the young remembered more in the preferred order just reported. It is speculated that task order was irrelevant for the old because they remembered more noun pairs from

whichever task was immediately prior to recall. Further, with free recall, it cannot be established whether the old lost information from the first task from the memory store, or whether it was simply inaccessible. There was also an interaction connoting that more noun pairs were recalled from set 1 out of sentence context than in sentence context, while for set 2, pairs were equally remembered regardless of context. This finding is surprising, as the two sets were equated for the number of high and low semantic relatedness pairs, and had no a priori reason to be differentially affected by context.

Lastly, when partial recall was investigated, set was significant, such that partial recall of noun pairs was better with set 2 than with set 1. Again, set was not expected to influence recall performance, as an identical number of high and low semantic relatedness noun pairs comprised each set. There was an interaction with age, indicating that set did not influence recall of the young, but noun pairs in set 2 were partially recalled better than those in set 1 for the old participants. Based on the findings from the first experiment, it could be speculated that, for old adults, there was something more unusual about the pairs in set 2 than in set 1 that rendered some of their individual nouns more memorable. Lastly, high semantic relatedness noun pairs were partially recalled more by males than by females, and the reverse was obtained for low

semantic relatedness pairs. Again, there is no obvious explanation for this result.

Covariates

Several variables were examined as covariates to determine whether they might account for some of the variability in the strict, lenient, and partial free recall of these noun pairs. The mean similarity ratings, the average number of features generated, and the years of formal education attained by participants were nonsignificant for the three types of recall, and when these analyses were investigated by pair type. However, the score on the analogies test was a significant covariate for strict and lenient free recall, but not for partial recall, or for these analyses by pair type. Correlations suggested that the relationships were stronger for the old than for the young adults, but they did not achieve significance. This supports a role for analogical reasoning in the recall of these noun pairs, and alludes to the importance of considering individual differences in recall.

Task 2

Instructions

There was no main effect of the instructions on recall performance. That is, retention did not differ with standard or imagery instructions. This is probably because the imagery condition was not a strong enough manipulation to produce an effect. In this experiment, imagery was briefly

described and then illustrated with a few metaphoric examples.

Recall

As in the first experiment, the young recalled more than the old across the three recall tasks with both a strict and a lenient recall criterion. Further, retention increased for the free, cued, and recognition tasks, respectively. This interacted with age, such that, while the young performed better than the old across all recall conditions, this difference was least for recognition, more for free recall, and was the greatest for cued recall. When improvement over the three recall tests was examined, the old improved proportionally more than the young adults with both a strict and a lenient recall criterion. Further, with a lenient criterion, the gain variable was also significant, indicating that there was greater proportional improvement from free to cued recall than from cued recall to recognition.

Considered together, these results lend support to the notion of cue-dependent forgetting (Tulving, 1974), in which memory is a function of the information present in the individual's cognitive environment at the time of retrieval. Forgetting occurs because the relevant retrieval information is absent and the memory store is inaccessible. This is contrasted with trace-dependent forgetting, in which memory is a function of the original stimulus encoding, and

forgetting occurs because the trace decays or is lost from the memory store. In the present instance, the greater proportional improvement of the old adults across the three recall tasks when compared to the young adults, and the near-perfect recall performance of the old participants on the recognition test, attest to the fact that information in the memory store was not lost, but was merely inaccessible.

When partial free recall was examined, there was a significant effect of sex indicating that females partially recalled more sentences than males. This replicates the sex effect obtained for partial recall in Experiment 1.

In the cued recall task, there was evidence with a strict scoring criterion that vehicle cues facilitated recall more than topic cues. Similar to Experiment 1, this lends credence to Paivio's (1979) contention that the vehicle is more critical than the topic in metaphor comprehension and recall because it guides the representation of the topic. Further, the salience imbalance hypothesis (Ortony, 1979a) is supported. This formulation states that, given nonliteral statements of the form "A is B", the matching attribute shared by both terms is of lower salience in the A term than in the B term. The absence of this cue type effect with a lenient criterion was surprising and the reason for it not readily obvious.

Further, analogous to the results obtained in Experiment 1, there were more omissions than constructions,

and more constructions than intrusions in the cued recall task. In fact, there were very few intrusions, indicating that cued recall errors were either those in which questions were left unanswered, or new nouns were created to complete the sentences. Also, there were more omissions for the old than for the young adults, indicating that the old did not respond to as many of the questions as the young participants, and thus, their errors were more often those of omission than of commission (Canestrari, 1963; Eisdorfer, 1965).

Lastly, there was no difference between the young and the old adults in the types of errors committed on the recognition test, and the effects of age, sex, and instructions were not significant. The high level of recall performance here may have precluded these factors from attaining significance.

Free and Cued Recall by Sentence Type

The results indicated some support for an effect of imagery, as high imagery sentences were recalled better than low imagery sentences with strict and lenient cued recall. This is consistent with Paivio's (1971) dual-coding approach, as high imagery (concrete) sentences can be processed by either verbal or imaginal strategies, while low imagery (abstract) sentences are mediated solely by verbal processes.

Further, there was an interaction between instructions

and sentence type, but its nature depended on the recall test employed. Extrapolating from dual-coding theory (Paivio, 1971), it was expected that high imagery sentences would be recalled better than low imagery sentences with standard instructions. However, with imagery instructions, low and high imagery sentences would not differ. This pattern of results was obtained for lenient free recall, and for strict cued recall. For lenient cued recall, high imagery sentences were recalled better than low imagery sentences with both types of instructions. However, the pattern of results was in the expected direction: the recall of the low imagery sentences improved with imagery instructions, but not enough to attain the level of recall with the high imagery sentences.

In contrast, the findings were not as predicted for strict free recall and partial recall. With respect to the former, the two sentence types did not differ with standard instructions. Given imagery instructions, recall of the low imagery sentences was better than that of the high imagery sentences. This result is inconsistent with dual-coding theory (Paivio, 1971), and is not readily interpretable. When partial free recall was examined, the low imagery sentences were partially recalled better than high imagery sentences with standard instructions. With imagery instructions, there was no difference in the recall of the two sentence types, but recall of both was not as high as

that of the low imagery sentences with standard instructions. Thus, it is not directly apparent why recall decreased for low imagery sentences given imagery instructions.

Serial Position Effects

The results indicated some support for serial position effects in the retention of metaphoric sentences for two of the recall variables. That is, a primacy effect was apparent for strict free recall, and partial recall displayed a recency effect. There was no effect of list position for lenient free recall. These results connote that, with a strict free recall criterion, sentences heard at the beginning of the tape were more likely to be remembered than later sentences. However, sentences that were partially recalled were more often those most recently heard from the end of the set.

Covariates

In order to examine whether part of the variability in recall performance might be attributed to individual differences, the years of formal education attained and the score on the analogies test were each entered separately as covariates into the recall analyses. The education variable was observed to be significant for lenient cued recall with topic and vehicle cues, and for constructions in the cued recall task. Correlations by age group, however, indicated that constructions were significant only for the old, while

lenient cued recall was significant for the young adults. Further, the direction of the correlation was negative for constructions, and positive for lenient cued recall. These results suggest that more years of formal education for the old participants was related to the production of fewer constructions in the cued recall task. This negative relationship seems counter-intuitive, and does not support the direct relationship obtained for these two variables in the first experiment. Also, higher lenient cued recall was related to more years of formal education for the young adults. It is not known why this relationship was specific to lenient cued recall.

When the score on the analogies test was entered as a covariate into the recall analyses, it was significant for strict and lenient recall across the three retention tasks, and for strict and lenient free recall by sentence type. It was also significant for strict and lenient recall by cue type, for omissions given topic and vehicle cues, and for partial recall. There were more significant correlations between these variables and the analogies score for the old than for the young adults, and the direction of the correlations was positive, with the exception of partial recall and omissions. Thus, better analogies scores were related to higher recall, less partial recall, and fewer omissions, particularly for the old adults.

Kausler (1982) discussed the notion of functional

reasoning, or reasoning through the use of analogy, and concluded on the basis of a few studies of metaphor interpretation that older adults were proficient in this type of reasoning. However, an explanation for the relationship between analogy and recall with old adults in the present instance is not readily apparent. It could be due to factors that evince no decline, and sometimes even improvement, with increasing age, such as verbal ability or crystallized intelligence. The higher correlations between the recall variables and the analogies score for the young as compared to the old adults in the first experiment suggest that the results here be interpreted with caution. That is, they may also be an artifact of smaller sample size or a reduced stimulus set. As in the first experiment, this significant effect of analogies extends previous theorizing (Sternberg, 1977; Sternberg & Nigro, 1983) about its role in metaphor comprehension to metaphor recall.

The final chapter will now summarize the results of these two experiments, and discuss other related issues.

CHAPTER 5

GENERAL DISCUSSION

Two experiments were conducted to examine the roles of imagery and semantic feature models in the retention of metaphoric sentences by young and old adults. Specifically, the first experiment required participants to rate a list of metaphors on one of three dimensions--imageability, semantic relatedness, or the number of interpretations possible--or to intentionally learn the sentences for an impending recall test. Free recall, cued recall with topic and vehicle cues, and recognition were then tested in a within-subjects design.

The second experiment was comprised of two tasks. In the first task, feature theory was investigated. This theory postulates that word meaning is a function of an underlying feature set that constitutes all possible meanings of a word. When words are combined to form a metaphor, their meaning is determined by an additive summation of their individual feature sets. Those features that are shared by the words being combined are raised in salience in the resultant meaning of the metaphor. Semantic relatedness ratings from the previous experiment were utilized to select a set of sentences high and low on this dimension. The nouns in each sentence were either presented as noun pairs or underlined in sentence context. Participants rated the similarity of the nouns in each pair, and generated

adjectives common to both pair members. Free recall of the noun pairs was then tested. Commensurate with semantic feature theory, it was expected that: (a) the rated similarity of the noun pairs and the number of adjectives (features) generated would be significantly and positively correlated, and (b) higher ratings and more features would be evident for high semantic relatedness noun pairs as compared to low semantic relatedness noun pairs.

The second task examined imagery by requiring young and old adults to listen to a set of sentences given either standard memory instructions or instructions to use imagery to remember the sentences. Half of the sentences were previously rated in Experiment 1 as high on imageability, and half as low on this dimension. Free recall, cued recall, and recognition of the sentences was tested in a manner identical to that in the first experiment.

There were four purposes of the present investigation. The first aim was to provide evidence for the replicability of previous research with these dimensions, and to extend their application to older adults. Second, manipulations of encoding and retrieval conditions were conducted to expand the study of memory in this area. The third intention was to examine imagery and semantic feature theories more directly through the dimensions of imageability and semantic relatedness, respectively. Lastly, age differences between young and old adults on the retention of metaphoric

sentences were analyzed. The results of these two experiments will now be discussed with respect to each of these four goals.

Replicability of the Dimensions

There was evidence across both experiments for the replicability of previous rating research utilizing these dimensions of metaphor processing (e.g., Katz et al., 1985; Marschark et al., 1983). The reliability and validity of the imagery and semantic relatedness normative scale values indicated that the dimensions were independently manipulable, and yet characteristic of metaphor comprehension. Thus, it is possible to elucidate dimensions of metaphor processing, and further experimentation with these scales is warranted.

In addition, the effect of age was not significant in the ratings of imageability, or in the number of interpretations possible. The old had higher semantic relatedness ratings than the young participants in Experiment 1, but not in the second experiment, where these ratings did not differ by age group. Further, the regression analyses revealed no age differences in the order of the significant predictors that emerged, and separate principal components analyses indicated the presence of a single underlying dimension of metaphor comprehension for both young and old adults. Thus, the influence of age in performing scale ratings was minimal, and where apparent,

limited to conceptions of semantic similarity. This provides evidence for the applicability of these dimensions of metaphor processing with both young and old adults.

However, the use of rating data, in general, has been qualified by some researchers. For instance, McCabe (1983) argued that rating the similarity between two items was dependent on the other pairs of items concurrently being rated. She asserted that this effect of context implies that absolute, unchangeable resemblance between any two things is not possible. In addition, Harris, Lahey, and Marsalek (1980) maintained that, while ratings are useful, caution must be exercised in their interpretation because their unidimensional focus most likely involves different processes than those in normal comprehension. Katz et al. (1988) stated, however, that better experimental control is permitted with unidimensional scales because the manipulated variables are not confounded with other uncontrolled item characteristics, and stronger theoretical claims can subsequently be made. Further, they argued that their 10 dimensions of metaphor processing load onto a single factor, indicating that each is measuring an aspect of the same process.

Memory: Manipulations of Encoding and Retrieval

Encoding was manipulated in the first experiment with the three rating tasks and the addition of an intentional orienting condition. The results indicated that the

intentional group recalled more than certain other incidental groups, but only in the free recall situation. This substantiates previous studies indicating a superiority of intentional as compared to incidental learning conditions (Backman, Mantyla, & Erngrund, 1984; Craik & Tulving, 1975; Dixon & von Eye, 1984; Hyde & Jenkins, 1969; Till & Jenkins, 1973). The advantage of the intentional condition disappeared with cued recall, as the provision of cues constituted half of each sentence and considerably reduced the memory demand on participants in the incidental conditions. Further, recall across the three incidental conditions did not differ, probably because they were all tasks requiring deep (semantic), rather than shallow (syntactic), processing (Craik & Lockhart, 1972). That is, each required participants to process the meaning of the nouns in each sentence. This is consistent with the research of Marschark and Hunt (1985), who found that orienting condition did not affect incidental free recall performance. Lastly, age did not interact with orienting condition in the recall analyses, indicating that retrieval was not differentially influenced by the encoding of the young and old adults.

The recognition test was employed to extend retrieval support beyond the cued condition previously investigated in metaphor recall (Backman & Dixon, 1987; Marschark & Hunt, 1985). The results of both experiments revealed that recall

improved significantly for free, cued, and recognition tasks, respectively, Thus, retention was observed to improve with increasing levels of retrieval support. However, this improvement was proportionally greater from free to cued recall than from cued recall to recognition.

This latter finding should be interpreted with caution, as it is most likely due to a ceiling effect. That is, there could not have been the same, or greater, proportional improvement from cued recall to recognition as there was between free and cued recall because of the small number of sentence stimuli employed. For example, in the second experiment, the mean amount of proportional gain from free to cued recall was 2.18, and the mean number of sentences recalled in the cued recall task was 12.8. To obtain an equivalent amount of proportional gain from cued recall to recognition, at least 28 sentence stimuli would be needed (i.e., eight more than the actual number employed). Thus, only with more sentences would it have been possible to witness equivalent proportional improvement across the three recall tasks, or even greater improvement between cued recall and recognition than between free and cued recall.

These results are consistent with the notion of cue-dependent forgetting (Tulving, 1974), in which poor memory performance is due to the absence of relevant information in the individual's cognitive environment at the time of retrieval. Thus, the information available in the memory

store is rendered inaccessible. In contrast, with trace-dependent forgetting, information that is retained in the memory store as a result of the original encoding of the event becomes unavailable because the trace deteriorates or is lost from the store. In the present instance, recall gains across the three tasks indicated that information was not lost from the memory store, but was merely inaccessible.

Cue type was varied in a within-subjects manner to determine whether topic or vehicle cues were most facilitative of cued recall. Consistent with Paivio's (1979) theorizing, the salience imbalance hypothesis (Ortony, 1979a), and previous research (Backman & Dixon, 1987), vehicles were demonstrated to be better cues than topics in both experiments. This suggests that the vehicle is more salient than the topic in the recall of metaphoric sentences. In contrast, Marschark and Hunt (1985, Exp. 3) obtained no difference in the effectiveness of the two cue types, but they utilized a between-subjects design. This attests to the importance of methodology in affecting the empirical outcome of research.

Partial free recall was scored (e.g., Marschark & Hunt, 1985, Exp. 2) in both experiments when either the topic or vehicle noun of a sentence was recalled alone, or a correct topic or vehicle noun was paired with a new noun, or a noun from a different sentence. It indicates that an individual's memory representation for a sentence is incomplete when

compared to the recall of an entire sentence. However, it may be better than not remembering anything at all because retention of the one noun can be used to trigger memory for the remaining sentence half, given the additional retrieval support provided in the cued recall and recognition tasks.

Partial recall was determined to be affected by the sex of the participants, as females partially recalled more sentences than males. This was the only significant sex effect across both experiments. The sex variable was included based on past research (e.g., Backman & Dixon, 1987) indicating that females remembered more than males in the free and cued recall of metaphoric and literal sentences. However, this effect did not replicate in the present instance, as the greater partial recall by females compared to males indicates that females had more incomplete memory traces than did males.

There was evidence in the first experiment that, although the mean ratings in the incidental conditions did not influence retrieval, the four sentence types did. Across free and cued recall, low imagery sentences were the least frequently recalled, consistent with Paivio's (1971) dual-coding approach. That is, these sentences contain abstract nouns that can only be processed with verbal strategies, in contrast to concrete nouns that can be processed either verbally or with imaginal strategies. With free recall, low semantic relatedness sentences were optimally recalled; with

cued recall, high semantic relatedness sentences were most often remembered.

It was speculated that recall of the low semantic relatedness sentences could be explained by their uncustomary presentation: they were literary sentences, presented in the absence of their original context. It was this unusual presentation that rendered them memorable in the free recall situation. With the provision of cues, however, the basis for recall shifted and became the similarity between the noun terms in the sentence. That is, the high degree of similarity between the nouns in high semantic relatedness sentences was most conducive to recall given half of the sentence as a cue. In addition, the age variable did not interact with sentence type, indicating that the hypothesized bases of retrieval in the free and cued recall situations (peculiarity and similarity, respectively) were identical for young and old adults. Further, dual-coding theory was supported for both age groups in the poor recall of low imagery sentences.

Retention was also extended to the examination of serial position effects in free recall. Across both experiments, there was evidence for primacy and recency effects in strict free recall, and for recency effects in lenient and partial free recall. This suggests that with free recall, metaphoric sentences will be remembered from both the beginning and end list positions. However, with

partial recall, a recency effect predominates, such that nouns from recently presented sentences are remembered.

Lastly, the literariness of the sentences in the first experiment was investigated. This variable is of interest because it highlights the issue of ecological validity. That is, the pattern of results with constructed metaphors may not be representative of those obtained with more natural metaphors created by poets. It was found that nonliterary sentences had higher mean ratings across the three incidental conditions than literary sentences, probably because, as stated by Katz et al. (1985), "presentation of poetic comparisons without context. . . puts those metaphors at a disadvantage relative to the constructed ones when one attempts to interpret them" (p. 380). Further, nonliterary sentences were remembered better than literary sentences in the cued recall task.

With partial free recall, the reverse effect was observed, as literary sentences were retained better than nonliterary sentences; this effect also approached significance with lenient free recall. It is postulated that this latter result also indicates the relevance of context, as some of the literary sentences were rendered unusual and more memorable by the absence of their original contexts. However, it is the degree of peculiarity that determines whether the entire sentence or merely part of it will be freely recalled. Specifically, the greater the peculiarity,

the more likely partial recall will be evident. With the provision of cues, however, there is a broader basis for retrieving items from the memory store, and peculiarity no longer serves as the basis of retrieval. Further, there was no interaction of age with sentence literariness in any of these analyses revealing that the absence of context with literary sentences similarly influenced the rating and recall performance of young and old adults.

Imagery and Semantic Feature Theories

Semantic feature theory was manipulated with the semantic relatedness dimension in task 1 of the second experiment. The theory proposes that word meaning is a function of an underlying feature set containing all possible meanings of a word. The meaning of a metaphor is determined by the feature overlap of the individual words comprising the metaphor; those features that are shared by both words are raised in salience in the resultant meaning of the metaphor.

The results indicated support for the theory. For both age groups, high semantic relatedness noun pairs were rated higher than low semantic relatedness noun pairs, and more features were generated for the former than for the latter pair type. Further, the correlation between ratings and features was significant for both young and old adults, indicating that highly similar words had more properties in common. These findings demonstrate that the tenets of

semantic feature theory were substantiated with both young and old adults, thereby extending the applicability of the theory to an older population. There was no effect of context on the rating or generation tasks, consistent with previous research (Johnson & Malgady, 1979). This is probably because underlining the topic and vehicle nouns in the sentence presentation format de-emphasized the remaining sentence context.

Lastly, the salience of the largest and two most common categories of features generated across the noun pairs was calculated, and observed to be fairly low in both instances. This suggests that, overall, similarity between the nouns in each pair was based on many properties, and not on a single, dominant attribute. Further, it should be noted that this measure of saliency across noun pairs is not related to that specified by Ortony (1979a), which concerns the relative salience of the topic and vehicle terms within a metaphor.

Imagery was examined in task 2 of the second experiment by manipulating the recall instructions with high and low imagery sentences. However, there was no effect of imagery instructions on recall. The absence of an instruction effect corroborates previous research (Harris, Lahey, & Marsalek, 1980, Exp. 1; Riechmann & Coste, 1978, cited in Riechmann & Coste, 1980), but it is also contrary to studies with proverbs indicating the advantage of comprehension instructions over imagery instructions in the recognition of

proverb interpretations (Riechmann, 1975, cited in Riechmann & Coste, 1980; Riechmann & van Wyk, 1977, cited in Riechmann & Coste, 1980). However, the dependent variable in the latter studies was the recognition of interpretations, and not the recognition of the original stimuli, as in the Harris et al. (1980) research, which might account for the discrepant findings.

The absence of an age by instruction interaction in the recall analyses is contrary to studies of paired-associate learning revealing beneficial effects of imagery instruction for old adults (e.g., Thomas & Ruben, 1973, cited in Poon, Walsh-Sweeney, & Fozard, 1980; Treat & Reese, 1976). However, the present results are consistent with studies of sentence memory indicating that the performance of the old adults did not improve with instructions to image (e.g., Backman & Nilsson, 1985; Whitbourne & Slevin, 1978). This extends previous findings with literal sentence stimuli to metaphoric sentences, and suggests that the advantage of imagery instruction for old adults may depend on the nature of the stimulus material employed.

An alternative interpretation is that the imagery instruction manipulation with sentence stimuli is not as strong as the mnemonics typically employed with paired-associates and list learning tasks (e.g., method of loci, peg-word technique). This study is illustrative in this regard, as the imagery instruction group received very brief

training in the formation of interactive images, and was not allowed any extensive practise. Further, sentence presentation was paced, and did not afford participants much time to form very elaborate or memorable images. Past research has ascertained that older adults are generally more disadvantaged than younger adults when the presentation rate is rapidly paced (e.g., Canestrari, 1963).

Also, high imagery sentences were recalled better than low imagery sentences with strict and lenient cued recall. This is consistent with a dual-coding position (Paivio, 1971), as high imagery sentences can potentially be processed with either verbal or imaginal strategies, while low imagery sentences are mediated solely by verbal processes. In addition, there was no interaction of age with sentence type, providing evidence of the utility of dual-coding theory with both young and old adults.

Lastly, there was an interaction between sentence type and instructions. It was expected that high imagery sentences would be recalled better than low imagery sentences with standard instructions, as predicted by dual-coding theory (Paivio, 1971). However, the two sentence types were anticipated to be similarly recalled with imagery instructions. These results were apparent for lenient free recall, strict cued recall, and in the expected direction for lenient cued recall; contrary findings were obtained for strict free recall and partial recall. Thus, there was some

support for the role of imagery in the recall of metaphoric sentences, contrary to the conceptual base hypothesis proffered by Honeck, Riechmann, and Hoffman (1975). However, this latter position is founded on evidence utilizing proverbs and their interpretations, and it is suggested that the processes involved in metaphor and proverb comprehension may not be directly comparable. As stated by Kausler (1982), "Conceivably, proverbs and metaphors involve different forms of functional reasoning that differ in their age sensitivity" (p. 552).

Age Differences

In examining retention across the free, cued, and recognition tasks in both experiments, the young recalled more than the old adults at all levels of retrieval support. However, the interaction between age and recall task indicated that the difference in their performance was almost eliminated with the recognition test. This is consistent with previous research indicating equivalent recall for young and old adults when a semantic orienting task was combined with recognition (Perlmutter, 1978, 1979). However, this beneficial effect of recognition may only be evinced utilizing within-subjects designs, as contrary evidence has been obtained with a between-subjects procedure (e.g., Mason, 1979). Further, the old improved proportionally more than the young across all levels of retrieval support. As stated previously, this substantiates

the notion of cue-dependent forgetting (Tulving, 1974), as recall improvement with cues and recognition indicates that information in the memory store was not lost, but was merely inaccessible. Lastly, the absence of an interaction between age and orienting condition connotes that the poorer memory performance of the old adults was not due to the encoding of the stimuli, as advanced by some researchers (e.g., Craik & Simon, 1980; Eysenck, 1974).

The old recalled more parts of sentences than the young adults in the free recall situation in the first experiment, and more words from noun pairs in the first task of the second experiment. This may indicate that, for some sentences/noun pairs, the old stored more incomplete memory representations than the young adults. When Marschark and Hunt (1985, Exp. 2) examined partial recall, they concluded that memory for metaphors was wholistic, and did not involve the "retention of discrete associations between subjects and predicates" (p. 418). However, the present findings suggest that this may not always be the case, particularly for older adults.

With respect to cued recall, the interaction between age and cue type, given the lenient criterion in Experiment 1, indicated no effect of cue type in the recall performance of the young adults, but better retention with vehicle cues than with topic cues for the old participants. This provides some suggestive evidence that, for old adults, the vehicle

specifies the topic (Paivio, 1979), and is of higher salience than the topic (Ortony, 1979a). However, as this interaction was not apparent with a strict criterion in Experiment 1, or in the second experiment with either criterion, it should be viewed as tentative until corroborated in further investigation. Further, across both experiments, the old omitted completing more of the sentences, when provided with half of the sentence as a cue, than the young participants. There were also more errors of incorrect rejection on the recognition test in the first experiment for the old than for the young adults. This may indicate a hesitancy in not positively identifying a sentence unless they are absolutely certain it was contained in the original stimuli (Botwinick, 1967).

Lastly, there was an interaction between age and list position for lenient free recall in Experiment 1, indicating that, while the young retained more than the old across all list positions, this difference increased from the beginning to the end of the list. Specifically, there was a recency effect for the young adults, and a primacy effect for the old adults (see Figure 4). This suggests that list position effects in the recall of metaphoric sentences may vary with the age of the participants, contrary to previous research indicating no age differences in serial position effects (Bromley, 1958; Eisdorfer, Axelrod, & Wilkie, 1963).

There were also age effects in the rating and feature

generation tasks in the second experiment. There was no difference between the young and old adults in ratings of noun pair similarity. However, the young produced more features, a greater variety of responses, and more correct responses, than the old participants. When the most common and two largest categories of response were examined, the young and old adults were observed to generate different features. Thus, while both age groups perceived the same degree of similarity between the nouns in each pair, the actual basis of this similarity was different. This result also argues that rating data alone may present an incomplete account of the processes involved in metaphor comprehension.

In addition, lenient free recall of the noun pairs indicated an interaction between age and the order of task completion. Specifically, recall by the old was not influenced by the order of the rating and feature generation tasks. However, for the young adults, retention was better when generation preceded rating than when the order was reversed. It was speculated that the young performed best in the order that minimized interference effects across tasks. However, the fact that the old were not similarly affected suggests that they remembered noun pairs from whichever task was most recent. Information from the first task was thus either lost from the memory store, or was inaccessible with the free recall test employed.

The issue of cohort effects arises here because these

were cross-sectional studies in which individuals of different ages were compared at the same point in time. That is, individuals of varying ages are also members of distinct cohorts, and it is highly likely that they can be distinguished in many respects besides age. There are diverse societal and cultural factors that influenced their growth. For example, historic events such as war and economic depression, and sociocultural evolution such as changes in education and health care produce divergence between cohorts. These cultural or generational differences may influence the variables of interest. Internal validity is thus jeopardized because observed age differences may actually be descriptive of cohort differences. For example, in the present research, there may have been age-related differences in beliefs about memory, conceptions of the tasks, and criteria of performance. Salthouse (1982) asserted that the best way to deal with this potential confound is to demonstrate that the majority of the observed age differences are determined by experimentally manipulable variables. In the present instance, this explanation of age differences in terms of manipulated variables cannot be fully demonstrated, and the possibility of cohort differences may not be rejected.

Further Issues

Given the results of this investigation, several related issues will now be addressed. These include the

notions of context, individual differences, possible directions for future study, and the relevance of this research to cognitive theories of information processing. Each of these areas will be addressed in turn.

The importance of context was postulated in the present research with respect to literary sentences, and was manipulated in the presentation of noun pairs in the second experiment. Several researchers have discussed the role of context in metaphor comprehension. For example, Hoffman (1984) has stated that almost any sentence can be regarded as metaphorical given the appropriate context. He provided evidence (McCabe, 1980, 1981, cited in Hoffman, 1984) indicating that if metaphors are presented in the context of meaningful conversation or prose, they may become more memorable. Anderson and Ortony (1975) also argued for the role of context in metaphor comprehension, stating that context imposes an interpretation, and that because it determines semantic relationships, it is indispensable. Camac and Glucksberg (1984) concurred in their assertion that, in some cases, there may be several features that the topic and vehicle terms share, but without additional context, the intended meaning is ambiguous.

The role of context has recently been elaborated in a theory advanced by Marschark (1985). Termed the "context availability model", it was evoked to explain findings inconsistent with dual-coding theory. Specifically, dual-

coding was robust at demonstrating the advantage of concrete over abstract sentences with list learning and mental manipulation (e.g., mental rotation, symbolic comparisons) studies, but it could not account for some evidence from sentence and prose research. The context availability model asserts that in list learning tasks, item-specific processing occurs. With individual sentences, however, two types of processing are required: the item-specific processing of concepts, and the relational processing among these concepts. With multi-sentence prose, there is even greater relational processing, as a variety of contextual information must be considered.

Employing terminology from discourse models (e.g., Kintsch & van Dijk, 1978), comprehension and memory for lists of unrelated sentences involves microstructure (i.e., the structure of individual propositions and their relations). However, related sentences in text also involve macrostructure, the relations among propositions and the topic of discourse. Concrete and abstract sentences involve similar processing at the macrolevel (e.g., prose), but concrete materials benefit more at the microlevel (e.g., lists) from the activation of perceptual information. The theory thus predicts that the superior recall of concrete materials observed with word and sentence lists may not generalize to sentences in paragraph contexts. That is, providing a context may increase the relational processing

of abstract materials and thus improve their recallability to a level commensurate with that of concrete sentences.

Marschark (1985) conducted three experiments testing this hypothesis by manipulating the ordered presentation of sentences in a list or in paragraph form, presenting sentences randomly or in order, and utilizing comprehension or intentional memory instructions. The results were consistent with a context availability model, indicating no difference in the recall of abstract and concrete sentences with paragraph presentation, but superior recall of concrete sentences compared to abstract sentences when sentences were presented as randomized lists. It was concluded that providing contexts for abstract materials increased the relational processing of them.

Thus, the implication of Marschark's (1985) research for the present findings is that if the metaphors had been presented in paragraph form, the effect of sentence type in both experiments may have been reduced. Specifically, recall of high and low imagery sentences would not have differed. This would have diminished support for dual-coding theory, or at least required that it be qualified on the basis of considerations of context. Further, this discussion of context suggests that caution must be exercised in generalizing to metaphors in more extended contexts, where the additional macrostructure also contributes to metaphor processing. Likewise, the present findings cannot be too

readily applied to figurative language in general, because, as previously noted, there is no conclusive evidence that similar mechanisms underlie the ability to comprehend different types of nonliteral speech (e.g., Kausler, 1982). In addition, these results may be relevant only to metaphors focusing on the relation between the topic and vehicle nouns, and not to those involving other speech classes. For example, in the metaphor "Inflation is eating our savings," the source of the metaphoricity is the verb and not the comparison of the topic and vehicle terms.

Another consideration that was represented in the present research was that of individual differences. Hoffman (1984) argued that awareness of metaphoric meaning also depended on individual differences in cognitive style. One of the conclusions of the research by Pollio and Smith (1979) was that perceptions of anomaly and metaphor were highly dependent on individual judgments, and that the influence of individual differences must play a key role in any theory of metaphor. Katz et al. (1988) also obtained evidence for considerable idiosyncracy across individuals' metaphor ratings, and stated that, "Variation in individual reactions to metaphors . . . remains a central theoretical problem that is only now beginning to receive empirical study" (p. 23).

In fact, Johnson and Malgady (1980) have specified a perceptual theory of metaphoric comprehension that focuses

on how individuals interpret metaphors (see Chapter 2). They hypothesized that there were two possible sources of the content of metaphoric interpretations: the material and the individual. More specifically, metaphors were viewed as associated with a set of meaning possibilities potentially realizable in any given individual. By implication, group data could not be utilized to discover what a metaphor meant to a specific person. Katz et al. (1988) concurred with this viewpoint in asserting that "individual differences almost certainly will interact with the structural characteristics of the metaphors themselves" (p. 23).

In the present research, individual differences were examined in analyses of covariance with the education variable and the score on the analogies test. With respect to the former, it was significant in predicting the number of constructions by the old participants in the cued recall task, although the direction of the relationship depended on the experiment. Further, it was related to the lenient cued recall of the young adults in Experiment 2. With respect to analogies, it was significant in several recall analyses across both experiments. However, the correlations were higher for the young than for the old in Experiment 1, and exactly the reverse for the second experiment. These findings support Sternberg's (1977) role of analogical reasoning in metaphor comprehension and extend it to the recall of metaphoric sentences. Future research should

attempt to identify additional factors (e.g., verbal ability, creativity) that might help to predict individual performance in the recall of metaphoric sentences, as this will aid in elucidating the mechanisms involved in the comprehension and recall of figurative language.

In addition, there are several possible directions future research might follow. First, more refined operational definitions of constructs are now possible. For example, the research here conceptualized features as adjectives. However, this may be a limited perspective, as there are also word taxonomies predicated upon verbs and nouns (e.g., robin-bird). Thus, future investigation of features may encompass a broader definition of the term than was employed here.

In this regard, Connor and Kogan (1980) have noted the difficulty in operationally defining features, and have devised a more qualitative approach utilizing the Metaphoric Triads Task (MTT). The MTT is a nonverbal instrument comprised of 29 triads of chromatic pictures. Two are metaphorically related, and the third shares a nonmetaphoric relation with each of the other two. The participant's task when presented with a triad of pictures is to choose the two that "go together" best, and to explain the basis for the pairing. The participant has the opportunity to form other pairings and describe them; interest is on whether the metaphoric pairing is recognized and clearly interpreted.

The authors collected data establishing the convergent and discriminant validity of the MTT. Greater use of the MTT may help to ascertain whether it describes features better than a verbal presentation.

Also, over a decade ago, Honeck, Riechmann, and Hoffman (1975) argued for a more formal specification of semantic similarity, and Ortony (1979a) has since offered the distinction between literal and figurative similarity. That is, the topic and vehicle of a metaphor may share a common feature only figuratively (Tourangeau & Sternberg, 1982). For example, the instructions in the feature generation task of the present research stressed the literal similarity between the topic and vehicle terms. However, participants may have thought of additional features based on figurative similarity that they did not list because the task did not seem to require this. Thus, future research will have to be explicit in the definition of similarity being utilized, and if comprehensiveness is desired, both types of similarity will have to be distinguished.

In addition, the concept of imagery needs to be more extensively investigated. For example, Harris, Lahey, and Marsalek (1980) have discussed the notion of surrealistic imagery, and its potential importance to imagery theories. However, such qualitative distinctions have not been subjected to empirical validation. For instance, the imagery dimension rated in the present research emphasized the

ease/difficulty of forming images. It did not analyze the complexity, quality, or nature of the images formed, or what domains were being utilized in imagery formation (e.g., topic, vehicle, or both). Further, no training with imagery was involved, nor were various imagery mnemonics contrasted for potentially different effects on ratings. All aspects of imagery need to be examined to provide a full account of its role in theories of metaphor comprehension and retention.

In addition, it would have been advantageous to also have included in Experiment 1 a metaphor goodness dimension to be rated. This is because the researcher noted in selecting sentences low on the semantic relatedness and imagery scales (Katz et al., 1988) that these sentences were also rated low on the dimensions of metaphoricity and metaphor goodness. These refer to the degree to which sentences are considered to be literally or figuratively true, and to how good or apt the metaphor appears, respectively. Thus, this presented somewhat of a confound, as sentences low on these dimensions were also less metaphorical, and perhaps less comprehensible. Inclusion of the metaphor goodness dimension would have permitted investigation of this potential confound.

An explanation for this observed relationship between metaphor goodness and the relative position of sentences on the imagery and semantic relatedness scales is apparent in the research of McCabe (1983). She conducted four

experiments to examine the relation between similarity and metaphor quality. The type of stimulus material was varied over the studies from word pairs, to sentence metaphors of the form "(An) A is a B", to metaphors in extended passages. There were three conditions requiring participants to either: (a) rate the similarity of noun pairs out of sentence context, (b) rate the quality of metaphors in sentences of the form "(An) A is a B", or (c) rate the quality of the metaphors in extended passages.

In general, it was found that the similarity of noun pairs was related to metaphor quality in isolated sentences, but not to quality in extended contexts. Further, there was no significant correlation between the quality of metaphors in isolated and extended contexts. McCabe concluded that when there is nothing else to go on, as in isolated sentences, metaphor quality depends on the similarity between the topic and vehicle terms. However, when metaphor occurs in a longer, natural context, similarity is overridden by that context, influencing metaphoric quality.

Thus, with respect to the sentences utilized in the present research, their metaphor quality was based on the similarity between the topic and vehicle terms. Those sentences displaying low similarity between the terms (i.e., low imagery and low semantic relatedness stimuli) would therefore be simultaneously viewed as poor in metaphor quality. Katz, Paivio, and Marschark (1985) have also noted,

however, that metaphor goodness ratings may involve both aesthetic and nonaesthetic properties, and that this dimension should be investigated more explicitly in future research.

Lastly, further studies should be conducted to determine the influence of age in the recall of both metaphor meaning and the verbatim recall of metaphor structure. The present research revealed age differences favoring the young adults with the latter type of retention. However, age effects may be reduced or nonexistent when memory for meaning is analyzed. Such a result was observed by Labouvie-Vief, Campbell, Weaver, and Tannenhaus (1979) (cited in Kausler, 1982) in a study with two conditions requiring young and old adults to either recall the verbatim content of metaphors, or to recall their meanings. The young adults performed better in the verbatim recall condition, while the old adults remembered more in the meaning condition.

Ultimately, the results of this metaphor research must be incorporated into cognitive theories of information processing and memory if the theories claim to have any generality at all. In this regard, there appears to be fairly widespread agreement that network models based on spreading activation mechanisms in a semantic "space" are not adequate in accounting for metaphor (Malgady, 1976; Malgady & Johnson, 1980; Ortony, Reynolds, & Arter, 1978).

These theories posit that concepts are represented as nodes in a semantic network. Movement within the network is through a spreading activation mechanism based on the notion of an intersection, in which a node is activated by two or more concepts in the input string.

The difficulty these models have in explaining the processing of metaphors is that they fail to account for the asymmetrical nature of metaphors. For example, the metaphor "Surgeons are butchers" has an entirely different meaning than "Butchers are surgeons", yet a spreading activation mechanism would result in the same intersection for both, even though the grounds of the two metaphors are different. However, Clevenger and Edwards (1988) conducted two experiments with the semantic differential procedure and concluded that proximity in semantic space was a major factor in selecting metaphoric vehicles for figurative topics.

There is also some consensus that propositional models (e.g., Kintsch, 1974) can not adequately account for metaphor (Belmore, 1982; Ortony, Reynolds, & Arter, 1978). These models assume that comprehended sentences are represented in the form of propositions. In this view, metaphors are seen as semantically anomalous, because they do not fit the rigid word meanings in propositions, and as such must be converted into explicit comparisons. Thus, in these models, it is impossible to distinguish metaphors from

anomalous sentences, and it is not evident that every metaphor can be readily converted into an explicit comparison (Ortony et al., 1978).

Ortony, Reynolds, and Arter (1978) suggested that schema models may provide a better understanding of metaphor comprehension. In this approach, knowledge is represented as associations elicited by things to which the words refer; in contrast, knowledge in propositional models is expressed in the core meanings of words. Thus, the information that is stored in schema representations is what is usually or normally the case, rather than what is necessarily the case. More specifically, the structure of a schema is of a series of variables together with relationships among them. This model is more flexible in matching incoming information to previously stored knowledge, and does not require that metaphors be reinterpreted, as in propositional models. Thus, schema theories may offer one plausible account of the cognitive processes involved in metaphor comprehension and retention. However, whatever the model, it is clear that "any psycholinguistic theory that does not handle metaphor is incomplete, and any that cannot is inadequate" (Ortony et al., 1978, p. 934).

References

Anderson, R. C., & Ortony, A. (1975). On putting apples into bottles - A problem of polysemy. Cognitive Psychology, 7, 167-180.

Arenberg, D. (1976). The effects of input condition on free recall in young and old adults. Journal of Gerontology, 31, 551-555.

Backman, L., & Dixon, R. A. (1987). Adult age differences in memory for metaphoric and nonmetaphoric sentences. Unpublished manuscript.

Backman, L., & Nilsson, L. (1985). Prerequisites for lack of age differences in memory performance. Experimental Aging Research, 11, 67-73.

Backman, L., Mantyla, T., & Erngrund, K. (1984). Optimal recall in early and late adulthood. Scandinavian Journal of Psychology, 25, 306-314.

Begg, I. (1972). Recall of meaningful phrases. Journal of Verbal Learning & Verbal Behavior, 11, 431-439.

Begg, I. (1973). Imagery and integration in the recall of words. Canadian Journal of Psychology, 27, 159-167.

Belmore, S. M. (1982). The role of imagery in recognition memory for sentences. Acta Psychologica, 50, 107-115.

Billow, R. M. (1975). A cognitive-developmental study of metaphor comprehension. Developmental Psychology, 11, 415-423.

Billow, R. M. (1977). Metaphor: A review of the psychological literature. Psychological Bulletin, 84, 81-92.

Black, M. (1962). Models and metaphors: Studies in language and philosophy. Ithaca, NY: Cornell University Press.

Boswell, D. A. (1979). Metaphoric processing in the mature years. Human Development, 22, 373-384.

Botwinick, J. (1967). Cognitive processes in maturity and old age. New York: Springer.

Bower, G. H. (1970). Analysis of a mnemonic device. American Scientist, 58, 496-510.

Bromley, D. B. (1958). Some effects of age on short term learning and remembering. Journal of Gerontology, 13, 398-406.

Camac, M. K., & Glucksberg, S. (1984). Metaphors do not use associations between concepts, they are used to create them. Journal of Psycholinguistic Research, 13, 443-455.

Campbell, P. N. (1975). Metaphor and linguistic theory. The Quarterly Journal of Speech, 61(1), 1-12.

Canestrari, R. E. (1963). Paced and self-paced learning in young and elderly adults. Journal of Gerontology, 18, 165-168.

Cassirer, E. (1953). The philosophy of symbolic forms. New Haven, CT: Yale University Press.

Clevenger, T., & Edwards, R. (1988). Semantic distance as a predictor of metaphor selection. Journal of Psycholinguistic Research, 17, 211-226.

Connor, K., & Kogan, N. (1980). Topic-vehicle relations in metaphor: The issue of asymmetry. In R. P. Honeck & R. Hoffman (Eds.), Cognition and figurative language. Hillsdale, NJ: Erlbaum.

Craik, F. I. M. (1968). Short-term memory and the aging process. In G. A. Talland (Ed.), Human aging and behavior. New York: Academic Press.

Craik, F. I. M. (1971). Age differences in recognition memory. Quarterly Journal of Experimental Psychology, 23, 316-323.

Craik, F. I. M. (1977). Age differences in human memory. In J. E. Birren & K. W. Schaie (Eds.), Handbook of the psychology of aging. New York: Van Nostrand Reinhold.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. Journal of Verbal Learning & Verbal Behavior, 11, 671-684.

Craik, F. I. M., & Simon, E. (1980). Age differences in memory: The roles of attention and depth of processing. In L. W. Poon, J. L. Fozard, L. S. Cermak, D. Arenberg, & L. W. Thompson (Eds.), New directions in memory and aging: Proceedings of the George A. Talland memorial conference. Hillsdale, NJ: Erlbaum.

Craik, F. I. M., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. Journal of Experimental Psychology: General, 104, 268-294.

Davidson, R. E., & Adams, J. F. (1970). Verbal and imagery processes in children's paired-associate learning. Journal of Experimental Child Psychology, 9, 429-435.

Dixon, R. A., & von Eye, A. (1984). Depth of processing and text recall in adulthood. Journal of Reading Behavior, 16, 109-117.

Drachman, D. A., & Leavitt, J. (1972). Memory impairment in the aged: Storage versus retrieval deficit. Journal of Experimental Psychology, 93, 302-308.

Edie, J. M. (1963). Expression and metaphor. Philosophy and phenomenological research, 23, 538-561.

Eisdorfer, C. (1965). Verbal learning and response time in the aged. Journal of Genetic Psychology, 107, 15-22.

Eisdorfer, C., Axelrod, S., & Wilkie, F. L. (1963). Stimulus exposure time as a factor in serial learning in an aged sample. Journal of Abnormal and Social Psychology, 67, 594-600.

Erber, J. T., Herman, T. G., & Botwinick, J. (1980). Age differences in memory as a function of depth of processing. Experimental Aging Research, 6, 341-348.

Eysenck, M. W. (1974). Age differences in incidental learning. Developmental Psychology, 10, 936-941.

Fainsilber, L., & Kogan, N. (1984). Does imagery contribute to metaphoric quality? Journal of Psycholinguistic Research, 13, 383-391.

Fisher, R. P., & Craik, F. I. M. (1977). Interaction between encoding and retrieval operations in cued recall. Journal of Experimental Psychology: Human Learning and Memory, 3, 701-711.

Harris, R. J. (1979a). Memory for metaphors. Journal of Psycholinguistic Research, 8, 61-71.

Harris, R. J. (1979b). Memory for literary metaphors. Bulletin of the Psychonomic Society, 13, 246-249.

Harris, R. J., Lahey, M. A., & Marsalek, F. (1980). Metaphors and images: Rating, reporting, and remembering. In R. P. Honeck & R. Hoffman (Eds.), Cognition and figurative language. Hillsdale, NJ: Erlbaum.

Hoffman, R. R. (1984). Recent psycholinguistic research on figurative language. Annals of the New York Academy of Sciences, 433, 137-165.

Hoffman, R. R., & Honeck, R. P. (1980). A peacock looks at its legs: Cognitive science and figurative language. In R. P. Honeck & R. Hoffman (Eds.), Cognition and figurative language. Hillsdale, NJ: Erlbaum.

Honeck, R. P. (1973). Interpretive versus structural effects on semantic memory. Journal of Verbal Learning & Verbal Behavior, 12, 448-455.

Honeck, R. P., Riechmann, P., & Hoffman, R. R. (1975). Semantic memory for metaphor: The conceptual base hypothesis. Memory & Cognition, 3, 409-415.

Hulicka, I. M. (1967). Age differences in retention as a function of interference. Journal of Gerontology, 22, 180-184.

Hulicka, I. M., & Grossman, J. L. (1967). Age-group comparisons for the use of mediators in paired-associate learning. Journal of Gerontology, 22, 46-51.

Hultsch, D. F. (1975). Adult age differences in retrieval: Trace dependent or cue-dependent forgetting. Developmental Psychology, 11, 197-201.

Hyde, T. S., & Jenkins, J. J. (1969). The differential effects of incidental tasks on the organization of recall of a list of highly associated words. Journal of Experimental Psychology, 82, 472-481.

Jenkins, J. J. (1979). Four points to remember: A tetrahedral model and memory experiments. In L. S. Cermak & F. I. M. Craik (Eds.), Levels of processing in human memory. Hillsdale, NJ: Erlbaum.

Johnson, M. G. (1970). A cognitive-feature model of compound free associations. Psychological Review, 77, 282-293.

Johnson, M. G. (1975). Some psychological implications of language flexibility. Behaviorism, 3, 87-95.

Johnson, M. G., & Malgady, R. G. (1979). Some cognitive aspects of figurative language: Association and metaphor. Journal of Psycholinguistic Research, 8, 249-265.

Johnson, M. G., & Malgady, R. (1980). Toward a perceptual theory of metaphoric comprehension. In R. P. Honeck & R. Hoffman (Eds.), Cognition and figurative language. Hillsdale, NJ: Erlbaum.

Katz, A. N. (1982). Metaphoric relationships: The role of feature saliency. Journal of Psycholinguistic Research, 11, 283-296.

Katz, A. N. (1983). Dominance and typicality norms for properties: Convergent and discriminant validity. Behavior Research Methods and Instrumentation, 15, 29-38.

Katz, A. N., & Paivio, A. (1975). Imagery variables in concept identification. Journal of Verbal Learning & Verbal Behavior, 14, 284-293.

Katz, A. N., Paivio, A., & Marschark, M. (1985). Poetic comparisons: Psychological dimensions of metaphoric processing. Journal of Psycholinguistic Research, 14, 365-383.

Katz, A. N., Paivio, A., Marschark, M., & Clark, J. M. (1988). Norms for 204 literary and 260 nonliterary metaphors on 10 dimensions. Unpublished manuscript.

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

Kausler, D. H. (1982). Experimental psychology and human aging. New York: John Wiley & Sons.

Kausler, D. H., & Hakami, M. K. (1983). Memory for topics of conversation: Adult age differences and intentionality. Experimental Aging Research, 9, 153-157.

Kintsch, W. (1974). The representation of meaning in memory. New York: Academic Press.

Kintsch, W., & van Dijk, T. A. (1978). Toward a model of text comprehension and production. Psychological Review, 85, 363-394.

Kirby, J. R., & Das, J. P. (1976). Comments on Paivio's imagery theory. Canadian Psychological Review, 17, 66-68.

Klatzky, R. L. (1980). Human memory: Structures and processes. San Francisco: W. H. Freeman and Company.

Kroll, N. E. A., & Schepeler, E. M. (1987). The comprehension and recall of similes as a function of context and cue. Journal of Psycholinguistic Research, 16, 101-132.

Langer, S. (1948). Philosophy in a new key. New York: Mentor.

Laurence, M. W. (1967). Memory loss with age: A test of two strategies for its retardation. Psychonomic Science, 9, 209-210.

Light, L. L., & Albertson, S. A. (1987). [Metaphoric comprehension (young versus older adults)- A replication of Verbrugge and McCarrell (1977)]. Unpublished raw data.

Macht, M. L., & Buschke, H. (1984). Speed of recall in aging. Journal of Gerontology, 39, 439-443.

Malgady, R. G. (1976). Category size, feature comparison, and the comprehension of figurative propositions. Perceptual and Motor Skills, 42, 811-818.

Malgady, R. G. (1977). Discriminant analysis of psychological judgments of literal and figurative meaningfulness and anomaly. Journal of Psychology, 95, 217-221.

Malgady, R. G., & Johnson, M. G. (1976). Modifiers in metaphors: Effects of constituent phrase similarity on the interpretation of figurative sentences. Journal of Psycholinguistic Research, 5, 43-52.

Malgady, R. G., & Johnson, M. G. (1977). Recognition memory for literal, figurative, and anomalous sentences. Bulletin of the Psychonomic Society, 9, 214-216.

Malgady, R. G., & Johnson, M. G. (1980). Measurement of figurative language: Semantic feature models of comprehension and appreciation. In R. P. Honeck & R. Hoffman (Eds.), Cognition and figurative language. Hillsdale, NJ: Erlbaum.

Marschark, M. (1985). Imagery and organization in the recall of prose. Journal of Memory and Language, 24, 734-745.

Marschark, M., & Hunt, R. R. (1985). On memory for metaphor. Memory & Cognition, 13, 413-424.

Marschark, M., Katz, A. N., & Paivio, A. (1983). Dimensions of metaphor. Journal of Psycholinguistic Research, 12, 17-40.

Mason, S. E. (1979). The effects of orienting tasks on the recall and recognition performance of subjects differing in age. Developmental Psychology, 15, 467-469.

Mason, S. E., & Smith, A. D. (1977). Imagery in the aged. Experimental Aging Research, 3, 17-32.

McCabe, A. (1983). Conceptual similarity and the quality of metaphor in isolated sentences versus extended contexts. Journal of Psycholinguistic Research, 12, 41-68.

Miller, G. (1979). Images and models, similes and metaphors. In A. Ortony (Ed.), Metaphor and thought. Cambridge: Cambridge University Press.

Nebes, R. D., & Andrews-Kulis, M. E. (1976). The effect of age on the speed of sentence formation and incidental learning. Experimental Aging Research, 2, 315-331.

Ortony, A. (1979a). Beyond literal similarity. Psychological Review, 86, 161-180.

Ortony, A. (1979b). The role of similarity in similes and metaphors. In A. Ortony (Ed.), Metaphor and thought. Cambridge: Cambridge University Press.

Ortony, A., Reynolds, R. E., & Arter, J. A. (1978). Metaphor: Theoretical and empirical research. Psychological Bulletin, 85, 919-943.

Ortony, A., Vondruska, R. J., Foss, M. A., & Jones, L. E. (1985). Salience, similes, and the asymmetry of similarity. Journal of Memory and Language, 24, 569-594.

Paivio, A. (1971). Imagery and verbal processes. New York: Holt, Rhinehart & Winston.

Paivio, A. (1975). Imagery and synchronic thinking. Canadian Psychological Review, 16, 147-163.

Paivio, A. (1976). Concerning dual-coding and simultaneous-successive processing. Canadian Psychological Review, 17, 69-72.

Paivio, A. (1979). Psychological processes in the comprehension of metaphor. In A. Ortony (Ed.), Metaphor and thought. Cambridge: Cambridge University Press.

Paivio, A., & Begg, I. (1981). Psychology of language. Englewood Cliffs, NJ: Prentice-Hall.

Perlmutter, M. (1978). What is memory aging the aging of? Developmental Psychology, 14, 330-345.

Perlmutter, M. (1979). Age differences in adults' free recall, cued recall, and recognition. Journal of Gerontology, 34, 533-539.

Pollio, H. R., Barlow, J. M., Fine, H. J., & Pollio, M. R. (1977). Psychology and the poetics of growth: Figurative language in psychology, psychotherapy, and education. Hillsdale, NJ: Erlbaum.

Pollio, H. R., & Smith, M. K. (1979). Sense and nonsense in thinking about anomaly and metaphor. Bulletin of the Psychonomic Society, 13, 323-326.

Poon, L. W., & Walsh-Sweeney, L. (1981). Effects of bizarre and interacting imagery on learning and retrieval of the aged. Experimental Aging Research, 7, 65-70.

Poon, L. W., Walsh-Sweeney, L., & Fozard, J. L. (1980). Memory skill training for the elderly: Salient issues on the use of imagery mnemonics. In L. W. Poon, J. L. Fozard, L. S. Cermak, D. Arenberg, & L. W. Thompson (Eds.), New directions in memory and aging: Proceedings of the George A. Talland memorial conference. Hillsdale, NJ: Erlbaum.

Raymond, B. J. (1971). Free recall among the aged. Psychological Reports, 29, 1179-1182.

Readence, J. E., Baldwin, R. S., Martin, M. A., & O'Brien, D. G. (1984). Metaphorical interpretation: An investigation of the salience imbalance hypothesis. Journal of Educational Psychology, 76, 659-667.

Reynolds, R. E., & Schwartz, R. M. (1983). Relation of metaphoric processing to comprehension and memory. Journal of Educational Psychology, 75, 450-459.

Richards, I. A. (1936). The philosophy of rhetoric. New York: Oxford University Press.

Riechmann, P. F., & Coste, E. L. (1980). Mental imagery and the comprehension of figurative language: Is there a relationship? In R. P. Honeck & R. Hoffman (Eds.), Cognition and figurative language. Hillsdale, NJ: Erlbaum.

Rohwer, W. D., Lynch, S., Suzuki, N., & Levin, J. R. (1967). Verbal and pictorial facilitation of paired-associate learning. Journal of Experimental Child Psychology, 5, 294-302.

Rowe, E. J., & Schnore, M. M. (1971). Item concreteness and reported strategies in paired-associate learning as a function of age. Journal of Gerontology, 24, 470-475.

Salthouse, T. A. (1980). Age and memory: Strategies for localizing the loss. In L. W. Poon, J. L. Fozard, L. S. Cermak, D. Arenberg, & L. W. Thompson (Eds.), New directions in memory and aging: Proceedings of the George A. Talland memorial conference. Hillsdale, NJ: Erlbaum.

Salthouse, T. A. (1982). Adult cognition: An experimental psychology of human aging. New York: Springer-Verlag.

Simon, E. (1979). Depth and elaboration of processing in relation to age. Journal of Experimental Psychology: Human Learning and Memory, 5, 115-124.

Simon, E. W., Dixon, R. A., Nowak, C. A., & Hultsch, D. F. (1982). Orienting task effects on text recall in adulthood. Journal of Gerontology, 37, 575-580.

Smith, A. D. (1977). Adult age differences in cued recall. Developmental Psychology, 13, 326-331.

Sternberg, R. J. (1977). Component processes in analogical reasoning. Psychological Review, 84, 353-378.

Sternberg, R. J., & Nigro, G. (1983). Interaction and analogy in the comprehension and appreciation of metaphors. Quarterly Journal of Experimental Psychology, 35(A), 17-38.

Till, R. E., & Jenkins, J. J. (1973). The effects of cued orienting tasks on the free recall of words. Journal of Verbal Learning & Verbal Behavior, 12, 489-498.

Till, R. E., & Walsh, D. A. (1980). Encoding and retrieval factors in adult memory for implicational sentences. Journal of Verbal Learning & Verbal Behavior, 19, 1-16.

Tourangeau, R., & Sternberg, R. J. (1981). Aptness in metaphor. Cognitive Psychology, 13, 27-55.

Tourangeau, R., & Sternberg, R. J. (1982). Understanding and appreciating metaphors. Cognition, 11, 203-244.

Treat, N., & Reese, H. W. (1976). Age, pacing, and imagery in paired-associate learning. Developmental Psychology, 12, 119-124.

Tulving, E. (1974). Cue-dependent forgetting. American Scientist, 62, 74-82.

Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. Psychological Review, 80, 352-373.

Tversky, A. (1977). Features of similarity. Psychological Review, 84, 327-352.

Verbrugge, R. R. (1977). Resemblances in language and perception. In R. Shaw & J. Bransford (Eds.), Perceiving, acting, and knowing: Toward an ecological psychology. Hillsdale, NJ: Erlbaum.

Verbrugge, R. R. (1984). The role of metaphor in our perception of language. Annals of the New York Academy of Sciences, 433, 167-183.

Verbrugge, R. R., & McCarrell, N. S. (1977). Metaphoric comprehension: Studies in reminding and resembling. Cognitive Psychology, 9, 494-533.

Whitbourne, S. K., & Slevin, A. E. (1978). Imagery and sentence retention in elderly and young adults. The Journal of Genetic Psychology, 133, 287-298.

Zelinski, E. M., Walsh, D. A., & Thompson, L. W. (1978). Orienting task effects on EDR and free recall in three age groups. Journal of Gerontology, 33, 239-245.

APPENDIX A

Stimulus Metaphors Used in Experiment 1

High Semantic Relatedness

1. A fan is a private breeze.
2. A playpen is a sandbox.
3. A lantern is a flashlight.
4. The soul is a voiceless thought.
5. Waves are the sailor's steed.
6. Billboards are yellow pages.
7. Smoke is a fire's calling card.
8. The automobile is a mechanical horse.
9. A blackmailer is a leech.
10. A dam is a fence.
11. Cigarettes are pacifiers.
12. The tongue is a bayonet.
13. A sewing needle is a porcupine quill.
14. Tweezers are pincers.
15. A swimming pool is a chlorinated pond.

Low Semantic Relatedness

1. Death is a drummer.
2. A table is a four-legged animal.
3. An envelope is a casket.
4. Mysticism is a figleaf.
5. The moon is a dancing cat.
6. A television antenna is a rabbit.
7. Night is a castle.
8. The environment is a chamber pot.
9. Sleep is an ocean.
10. Noon is a rising dolphin.
11. Money is blood.
12. A fisherman is a spider.
13. Laughter is a laxative.
14. A forest is a chameleon.
15. Autumn is a hawk.

High Imagery

1. A pimple is the skin's volcano.
2. Adventure is a roller-coaster.
3. Skyscrapers are giraffes.
4. An egg is an incubator.
5. Drug addicts are hollow shells.
6. A marshmallow is a tiny pillow.
7. A tie is a noose.
8. The mind is a mountainous landscape.
9. Clouds are wooly worlds.
10. A rooster is an alarm clock.

11. A trained fighter is an armored knight.
12. Earth is a buzzing beehive.
13. Snow is falling dandruff.
14. An anthill is a crawling colony.
15. Twinkling stars are flickering candles.

Low Imagery

1. Vanity is a charade.
2. History is a sponge.
3. Independence is a false liberty.
4. The human form is a fiery forge.
5. Mankind is a preface.
6. Religion is society's lifeblood.
7. Extravagance is poison ivy.
8. Anarchy is a destructive watering can.
9. Maturity is wrecked happiness.
10. Anticipation is spring seeding.
11. Experience is fermented wine.
12. Mathematics is unadorned beauty.
13. Jealousy is an unannounced competition.
14. Fame is an intrusion.
15. A shoelace is a garden hose.

APPENDIX B

Task Instructions Used in Experiment 1

Number of Alternative Interpretations

In this task, you will be asked to judge sentences with regard to the number of possible interpretations that you can give to each. Consider the following sentence: "When the wind blew, the house shook." You probably understand or comprehend the meaning of this sentence. Some sentences, however, appear to be nonsense or take a long time to understand. Consider, for example, the following sentence: "When the show was over the gruesome split." This sentence is probably nonsense to you, although with effort you could probably make some sense out of it. For example, if you thought of "The Gruesome" as the name of a travelling theatrical group, the sentence could be quite understandable.

Some sentences can be given more than one interpretation. Consider some possible interpretations of the sentence: "Truth is a sword." Some possible interpretations could be: 1) truth can be a weapon, 2) truth is long and straight, and 3) truth cuts through lies. You might be able to think of other interpretations.

In this task we present you with a list of nonliteral sentences that are metaphors and ask you to state the number of different interpretations that you can think of for each sentence. If no interpretation comes to your mind, please

write a 0 (zero) in the space provided at the beginning of the sentence; if you think of one, write 1 in the space; if you think of two, write 2, and so on. Don't be concerned about how often you use a particular number, as long as each is your true judgment.

Below are several examples. Indicate the number of different alternative interpretations you can give to each sentence by writing the appropriate number in the space provided.

Semantic Relatedness

In this task, you will be asked to judge how similar in meaning the two objects of a sentence are. For example, consider the following two sentences: 1) Cities are beehives; and 2) The city is a library of lights. In each of these sentences, two objects are compared and related to each other in a nonliteral way. You will be presented with a set of such sentences, all of which are metaphors. Your task will be to decide how closely related in meaning the two objects of each sentence are. In a sentence like "A city is a large town," for example, CITY and LARGE TOWN probably would be considered fairly close in meaning. However, in a sentence like " A city is a swamp of individuals," CITY and SWAMP OF INDIVIDUALS would probably be considered to have very little in common. Your ratings should be made on a seven-point scale where 1 is the LOW SIMILARITY end and 7 is the HIGH SIMILARITY end. A depiction of the rating scale is:

1 2 3 4 5 6 7

LOW SIMILARITY HIGH SIMILARITY

Again, you are to decide how close in meaning the two objects of each sentence are. You are not to rate how good or bad each sentence is, or whether you like it or not, or how pleasant it is. You are simply to rate each sentence for the similarity in meaning of the two objects in the sentence.

In the space provided beside each sentence, please write in the number from 1 to 7 that best estimates your feeling as to the similarity in meaning of the two objects in each metaphor. If, in your opinion, the two objects of a metaphor are very distant in meaning, then a rating of 1 should be given. If, however, you think the two objects are very close in meaning, then a rating of 7 should be given. For those metaphors in which the meaning of the objects is intermediate to these extremes, an appropriate rating from 1 to 7 should be given. Don't be concerned about how often you use a particular number as long as it is your true judgment.

Below are several examples. Indicate your judgment of the similarity between the two objects of each metaphor by writing in the number from 1 to 7 in the space provided beside each sentence.

Imageability

Sentences are known to differ in their ability to arouse mental images. Some sentences arouse mental images (that is, sensory experiences such as mental pictures) very quickly and easily, whereas other sentences may only do so with difficulty or not at all. For example, you might find it fairly easy to make a mental picture for the sentence: "Tree trunks are straws for thirsty leaves and branches." You might picture the leaves and branches as cartoon persons sucking on long straws on a hot day. The purpose of this task is to determine the ease with which mental images are formed for different sentences. You will be presented with a set of sentences, all of which are metaphors. You are to rate on a 7-point scale the ease or difficulty with which they arouse mental images. If you find it easy to form an image to a particular sentence, then you should give it a HIGH IMAGERY rating. If, however, you find it difficult to form an image to the sentence, then it should be given a LOW IMAGERY rating. A depiction of the rating scale is:

1 2 3 4 5 6 7

LOW IMAGERY HIGH IMAGERY

In the space provided beside each sentence, please write in the number from 1 to 7 that best estimates your feeling as to the ease with which a mental picture can be

formed for each metaphor. If, in your opinion, it is extremely difficult to form an image to the sentence, then a rating of 1 should be given. If, however, you think it is very easy to form an image to the sentence, then a rating of 7 should be given. For those metaphors in which your ability to form a mental picture is intermediate to these extremes, an appropriate rating from 1 to 7 should be given. Don't be concerned about how often you use a particular number as long as it is your true judgment.

Below are several examples. Indicate your judgment of the ease with which a mental picture can be formed for each metaphor by writing in the number from 1 to 7 in the space provided beside each sentence.

Intentional

In this task, you will be presented with a list of sentences, all of which are metaphors. A metaphor is a type of sentence in which two objects are compared in a nonliteral way. For example, consider the following metaphoric sentence: "Cities are beehives." This sentence is a metaphor, because, literally speaking, a city is not a beehive. However, a city is like a beehive because it is full of activity, because it is noisy, and so on. You are to read the metaphoric sentences and try to remember as much about them as you can for a memory test that will follow. You aren't expected to remember everything. Just do your best and try to remember as much as you can.

Table 1

Mean Age, Years of Education, and Analogies Score by Age Group, Sex, and Orienting Condition

	Demographic Variables		
Condition	Mean Age	Education	Analogies[a]
Young Adults			
Overall	25.43	16.15	24.73
	(3.14)	(1.85)	(3.81)
Males	25.25	16.65	25.75
	(3.19)	(1.92)	(3.17)
Females	25.60	15.65	23.70
	(3.09)	(1.79)	(4.46)
Imagery	25.10	17.00	24.10
	(2.13)	(2.31)	(4.10)
Males	26.80	18.40	25.20
	(0.84)	(2.61)	(2.78)
Females	23.40	15.60	23.00
	(1.52)	(0.55)	(5.20)
Semantic Relatedness	26.10	16.40	27.00
	(4.68)	(1.84)	(2.06)
Males	22.80	15.60	26.60
	(2.17)	(1.14)	(2.19)
Females	29.40	17.20	27.40
	(4.16)	(2.17)	(2.07)
Interpretations	24.80	14.30	22.10
	(4.39)	(1.64)	(5.41)
Males	23.40	14.40	24.60
	(3.44)	(1.52)	(2.97)
Females	26.20	14.20	19.60
	(5.17)	(1.92)	(6.43)
Intentional	25.70	16.90	25.70
	(4.95)	(2.69)	(4.30)
Males	28.00	18.20	26.60
	(6.29)	(2.39)	(4.72)
Females	23.40	15.60	24.80
	(1.52)	(2.51)	(4.15)

Old Adults			
Overall	68.23	14.53	23.53
	(3.97)	(2.79)	(5.01)
Males	69.20	15.55	23.30
	(4.00)	(3.63)	(4.88)
Females	67.25	13.50	23.75
	(3.93)	(1.95)	(5.14)
Imagery	70.10	14.00	22.80
	(3.64)	(3.74)	(6.34)
Males	72.40	14.40	23.80
	(2.07)	(5.41)	(6.54)
Females	67.80	13.60	21.80
	(3.49)	(1.34)	(6.72)
Semantic Relatedness	67.50	15.10	25.00
	(6.04)	(2.28)	(4.00)
Males	68.00	16.80	27.00
	(5.70)	(0.84)	(2.56)
Females	67.00	13.40	23.00
	(7.00)	(1.95)	(4.42)
Interpretations	68.10	15.90	23.60
	(3.48)	(3.45)	(5.85)
Males	67.20	16.60	20.80
	(4.15)	(4.67)	(5.81)
Females	69.00	15.20	26.40
	(2.83)	(1.92)	(4.88)
Intentional	67.20	13.10	22.70
	(3.80)	(3.25)	(4.47)
Males	69.20	14.40	21.60
	(4.09)	(3.58)	(4.62)
Females	65.20	11.80	23.80
	(2.39)	(2.59)	(4.55)

Note. The values in parentheses are standard deviations.

[a]Maximum score = 35.

Table 2

Mean Ratings for Each Sentence Type by Age Group and Orienting Condition

Sentence Type	Orienting Condition		
	IMG	SRL	INTS
Young Adults			
Overall	4.9	3.6	1.7
High Imagery	6.1	4.5	2.0
Low Imagery	3.5	2.8	1.5
High Semantic Relatedness	5.9	4.8	1.8
Low Semantic Relatedness	4.0	2.2	1.6
Old Adults			
Overall	4.4	4.3	1.7
High Imagery	5.6	5.3	1.9
Low Imagery	3.3	3.6	1.6
High Semantic Relatedness	5.6	5.4	1.9
Low Semantic Relatedness	2.9	2.8	1.5

Table 3

Correlations Among the Scales by Age Group

Scales	2	3
	Young Adults	
1. Imagery	.7307	.5817
2. Semantic Relatedness	-----	.4425
3. Interpretations		-----
	Old Adults	
1. Imagery	.8866	.5100
2. Semantic Relatedness	-----	.4444
3. Interpretations		-----

Table 4

Mean Number of Sentences Recalled by Age Group and Retrieval Condition

Age Group	Retrieval Condition		
	Free	Cued	Recognition
Strict Criterion			
Young Adults	12.45	37.60	58.63
Old Adults	5.80	24.43	55.25
Lenient Criterion			
Young Adults	17.43	44.63	58.63
Old Adults	8.78	31.00	55.25

Note. The total number of sentences is sixty.

Table 5

Mean Number of Omissions, Intrusions, and Constructions by Age Group

	Cuing Variables		
Age Group	Omissions	Intrusions	Constructions
Young Adults	11.05	1.17	3.15
Old Adults	23.80	1.67	3.52

Table 6

Correlations Between Analogies Score and Selected Recall Variables

	Free Recall	Cued Recall	Topic Cues	Vehicle Cues	Omissions
Young Adults					
		Strict Criterion			
Analogies	.41	.57	.54	.52	-.46
		Lenient Criterion			
Analogies	.42	.56	.55	.53	-.46
Old Adults					
		Strict Criterion			
Analogies	.35	.44	.34	.49	-.41
		Lenient Criterion			
Analogies	.28	.38	.28	.46	-.41

Note. All correlations are significant at $p < .05$.

Table 7

Mean Age, Years of Education, and Analogies Score by Age Group, Sex, Context, and Instruction Condition

Context	Demographic Variables		
	Mean Age	Education	Analogies[a]
	Imagery Instructions		
Young Adults			
Overall	27.42	14.55	25.55
	(3.57)	(1.77)	(3.16)
Males	28.42	15.25	26.59
	(2.32)	(2.54)	(4.29)
Females	26.42	13.84	24.50
	(4.82)	(1.00)	(2.03)
Sentences	25.42	15.59	25.75
	(4.82)	(0.65)	(3.10)
Males	26.50	16.50	27.50
	(2.12)	(0.71)	(3.54)
Females	24.33	14.67	24.00
	(7.51)	(0.58)	(2.65)
Noun Pairs	29.42	13.50	25.34
	(2.32)	(2.89)	(3.22)
Males	30.33	14.00	25.67
	(2.52)	(4.36)	(5.03)
Females	28.50	13.00	25.00
	(2.12)	(1.41)	(1.41)
Old Adults			
Overall	67.38	14.88	26.71
	(4.20)	(2.69)	(4.31)
Males	69.67	15.59	28.17
	(3.17)	(4.09)	(3.99)
Females	65.09	14.17	25.25
	(5.23)	(1.29)	(4.63)
Sentences	65.84	14.92	24.50
	(4.17)	(3.06)	(7.10)
Males	66.00	16.50	28.00
	(2.83)	(4.95)	(5.66)
Females	65.67	13.33	21.00
	(5.51)	(1.16)	(8.54)
Noun Pairs	68.92	14.84	28.92
	(4.23)	(2.32)	(1.51)

Males	73.33	14.67	28.33
	(3.51)	(3.22)	(2.31)
Females	64.50	15.00	29.50
	(4.95)	(1.41)	(0.71)
	Standard Instructions		
Young Adults			
Overall	25.47	14.09	26.71
	(4.16)	(1.24)	(3.71)
Males	25.34	13.84	25.42
	(4.36)	(0.77)	(6.91)
Females	25.59	14.34	28.00
	(3.95)	(1.71)	(0.50)
Sentences	27.59	13.84	26.17
	(4.71)	(2.18)	(3.02)
Males	26.67	13.67	23.33
	(3.06)	(1.53)	(6.03)
Females	28.50	14.00	29.00
	(6.36)	(2.83)	(0.00)
Noun Pairs	23.34	14.34	27.25
	(3.60)	(0.29)	(4.39)
Males	24.00	14.00	27.50
	(5.66)	(0.00)	(7.78)
Females	22.67	14.67	27.00
	(1.53)	(0.58)	(1.00)
Old Adults			
Overall	64.38	14.13	22.63
	(2.95)	(1.28)	(4.16)
Males	64.67	15.59	26.25
	(3.22)	(1.51)	(3.14)
Females	64.09	12.67	19.00
	(2.67)	(1.04)	(5.16)
Sentences	63.92	13.84	22.50
	(3.58)	(1.16)	(4.91)
Males	65.33	14.67	23.00
	(5.03)	(2.31)	(5.57)
Females	62.50	13.00	22.00
	(2.12)	(0.00)	(4.24)
Noun Pairs	64.84	14.42	22.75
	(2.32)	(1.40)	(3.40)
Males	64.00	16.50	29.50
	(1.41)	(0.71)	(0.71)
Females	65.67	12.33	16.00
	(3.22)	(2.08)	(6.08)

[a]Maximum score = 35.

Table 8

Mean Number of Noun Pairs Recalled by Age Group and Context Condition

Condition	Recall	
	Strict	Lenient
Young Adults		
Sentences	8.00	11.10
Noun Pairs	8.70	11.80
Old Adults		
Sentences	4.20	6.20
Noun Pairs	6.00	7.20

Note. The total number of noun pairs is twenty.

Table 9

Mean Number of Sentences Recalled by Age Group and Instruction Condition

Condition	Recall		
	Free	Cued	Recognition
Strict Criterion			
Young Adults			
Imagery	6.1	13.6	19.8
Standard	7.1	13.7	20.0
Old Adults			
Imagery	3.2	8.3	19.6
Standard	2.3	6.0	18.7
Lenient Criterion			
Young Adults			
Imagery	8.5	16.0	19.8
Standard	9.3	15.9	20.0
Old Adults			
Imagery	5.0	10.8	19.6
Standard	3.6	8.4	18.7

Note. The total number of sentences is twenty.

Table 10

Mean Number of Sentences Recalled by Age Group, Instruction Condition, and Cue Type

Condition	Cue Type	
	Topics	Vehicles
Strict Criterion		
Young Adults		
Imagery	6.2	7.4
Standard	5.7	8.0
Old Adults		
Imagery	3.3	5.0
Standard	2.0	4.0
Lenient Criterion		
Young Adults		
Imagery	8.1	7.9
Standard	7.7	8.2
Old Adults		
Imagery	5.3	5.5
Standard	3.9	4.5

Table 11

Mean Number of Omissions, Intrusions, and Constructions by Age Group and Instruction Condition

Condition	Cuing Variables		
	Omissions	Intrusions	Constructions
Young Adults			
Imagery	3.10	.20	.70
Standard	2.50	.10	1.50
Old Adults			
Imagery	6.70	.50	2.00
Standard	10.00	.10	1.50

Table 12

Correlations Between Analogies Score and Selected Recall Variables

	Free Recall	Cued Recall	Topic Cues	Vehicle Cues	Omissions
Young Adults					
		Strict Criterion			
Analogies	.45*	.45*	.41*	.38*	-.28
		Lenient Criterion			
Analogies	.38	.37	.36	.32	-.28
Old Adults					
		Strict Criterion			
Analogies	.31	.75*	.57*	.74*	-.62*
		Lenient Criterion			
Analogies	.45*	.72*	.59*	.74*	-.62*

*$p < .05$.

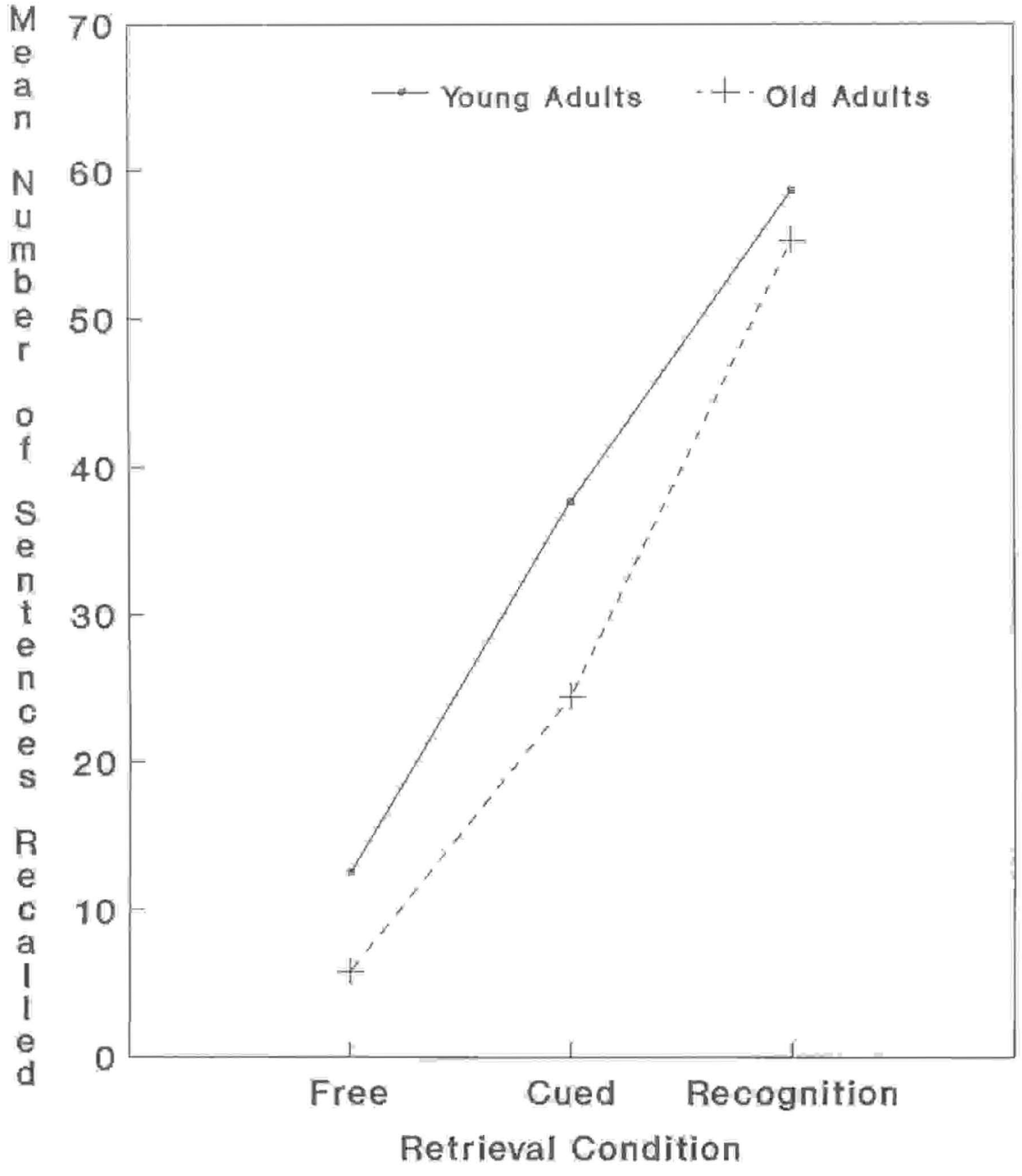

Figure 1. Mean number of sentences strictly recalled in Exp. 1 as a function of age and retrieval condition.

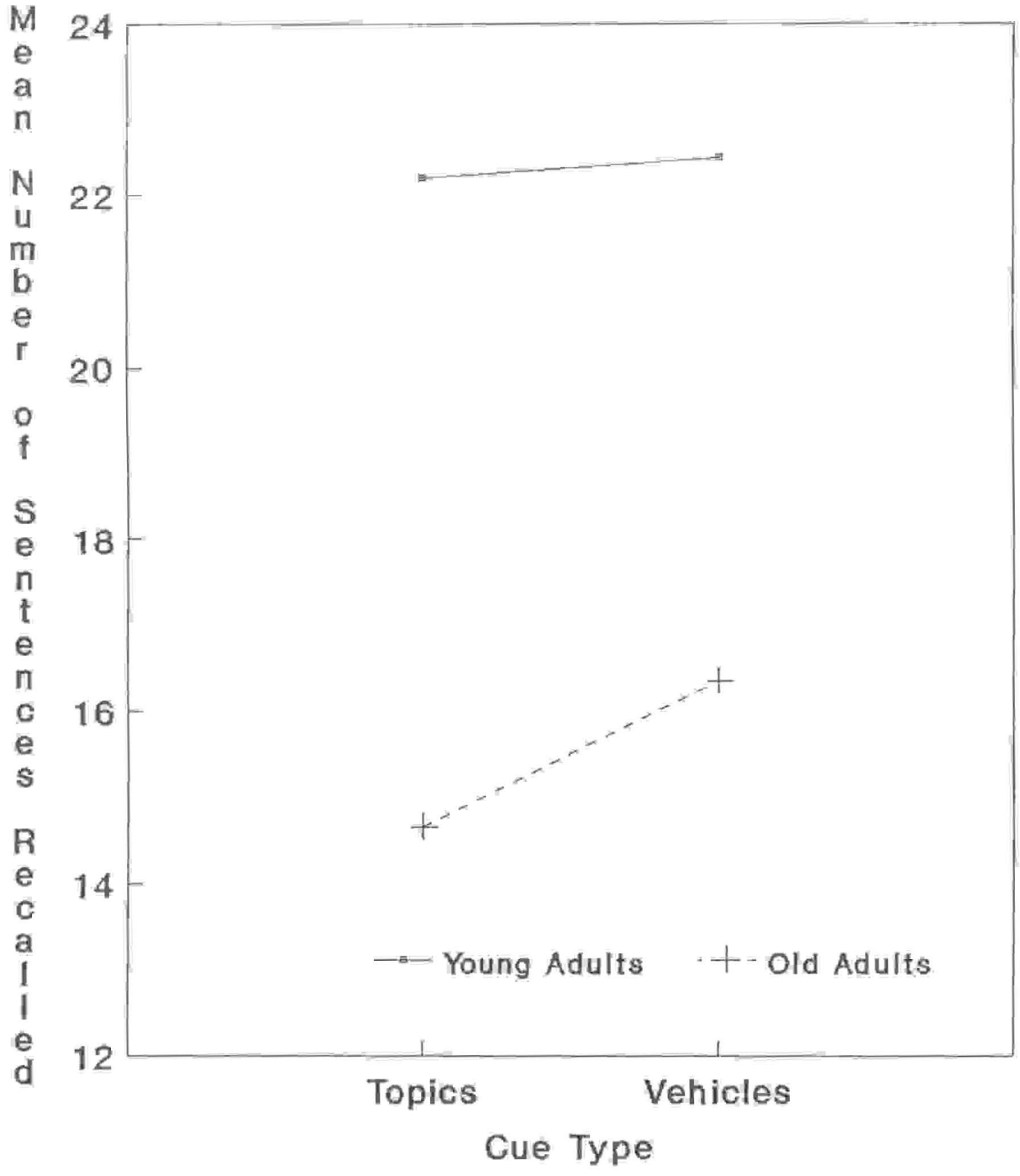

Figure 2. Mean number of sentences leniently recalled in the cued recall task as a function of age and cue type.

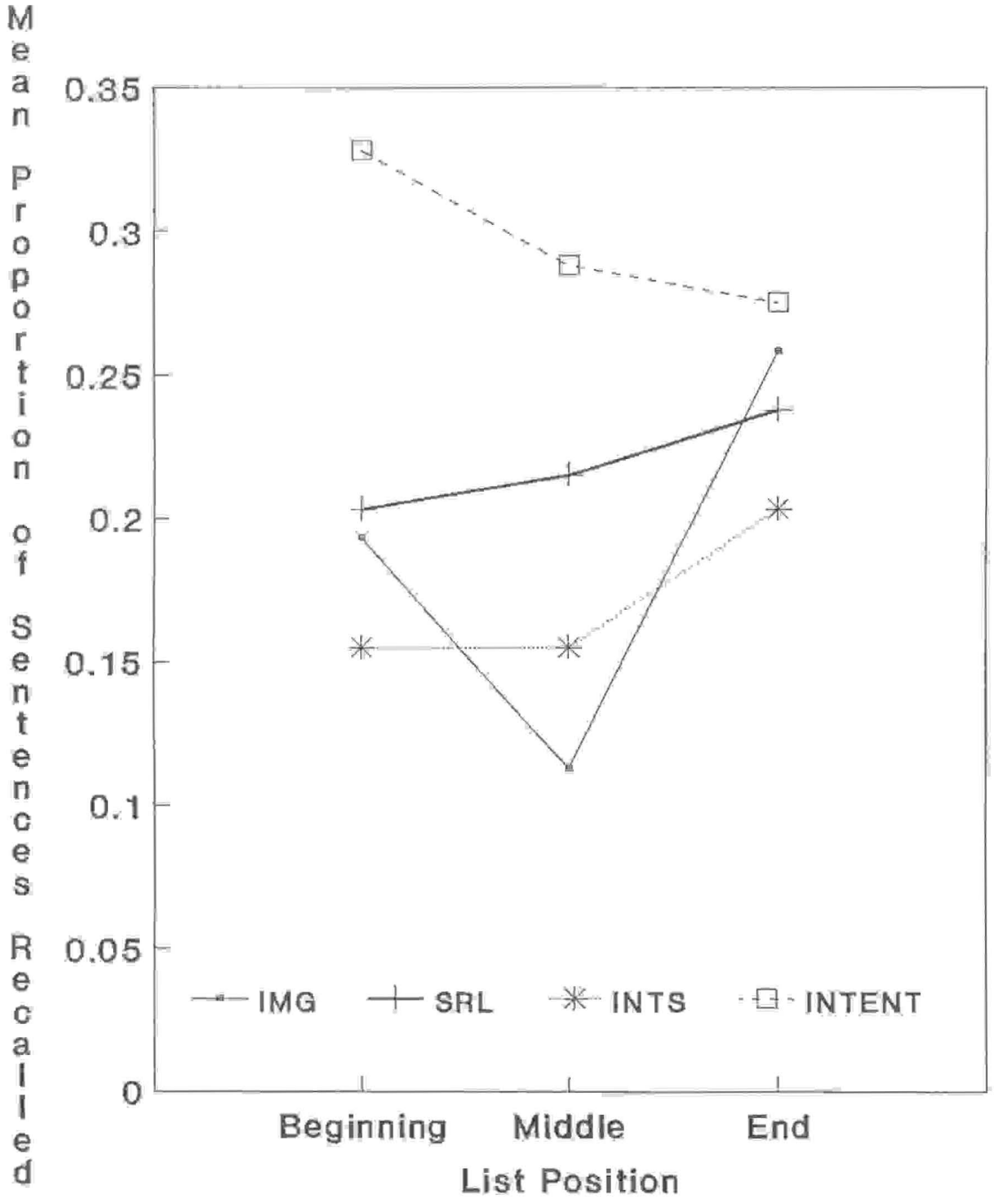

Figure 3. Mean proportion of sentences freely recalled (lenient criterion) by orienting condition & list position.

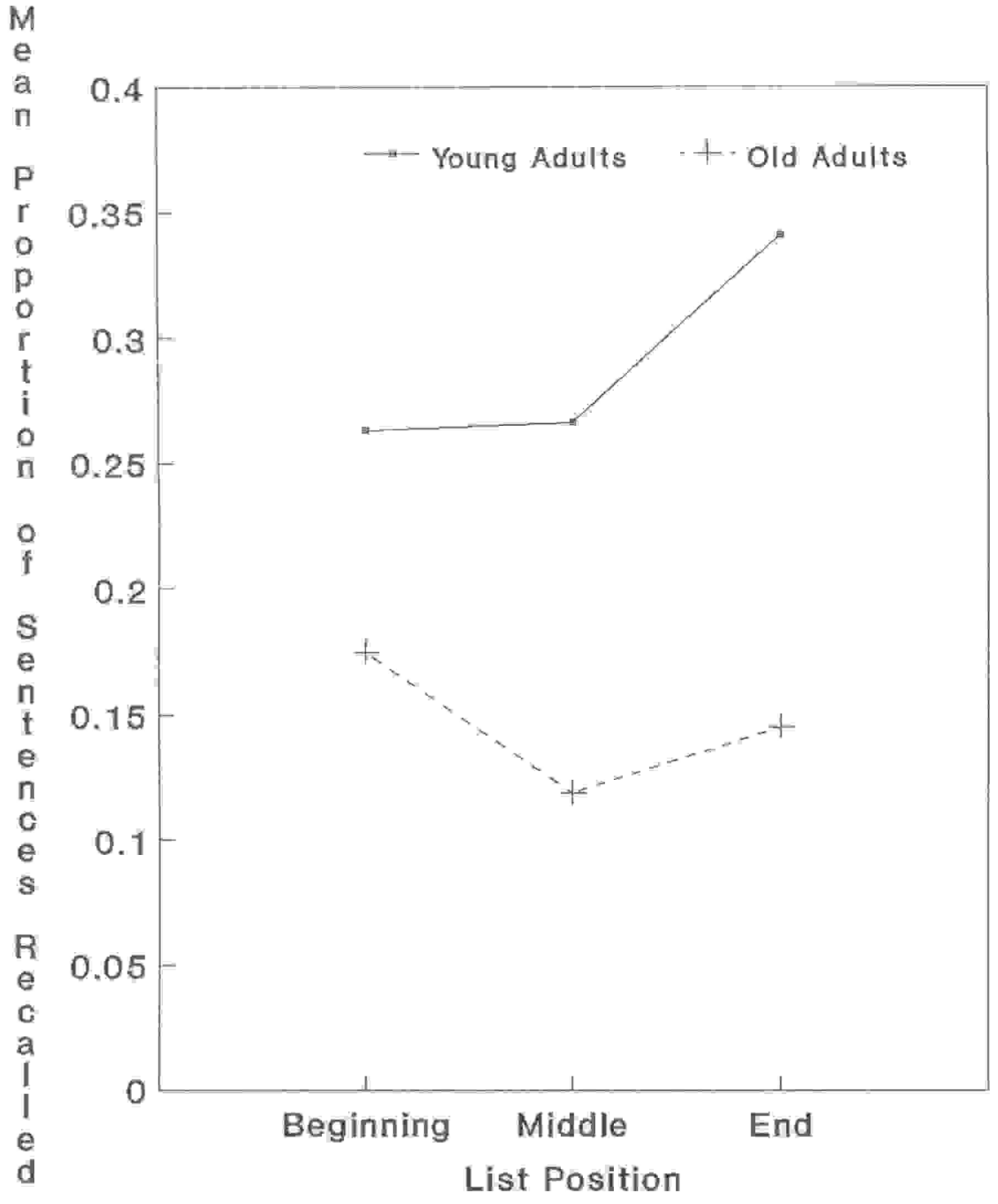

Figure 4. Mean proportion of sentences freely recalled (lenient criterion) as a function of age and list position.

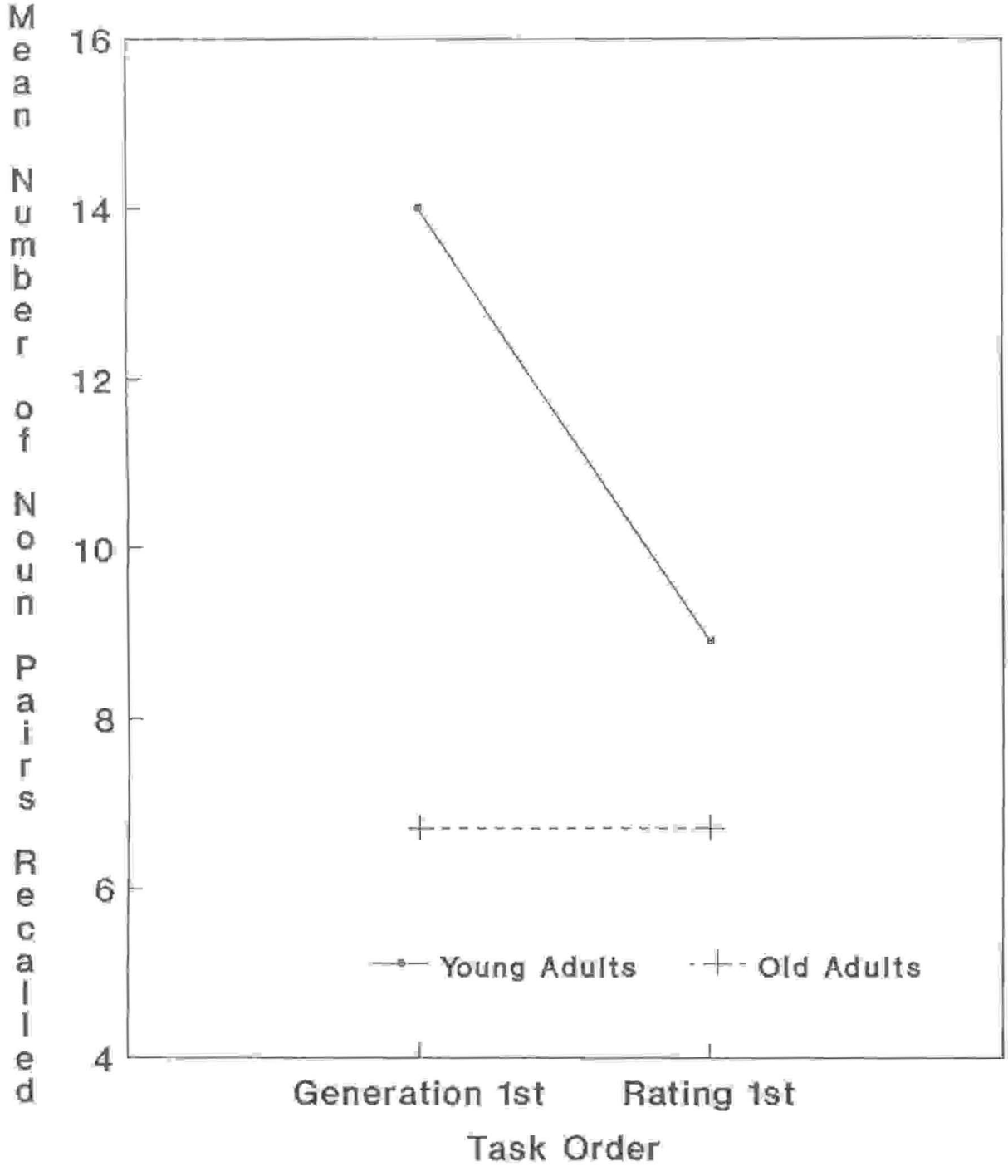

Figure 5. Mean number of noun pairs freely recalled (lenient criterion) as a function of age & task order.

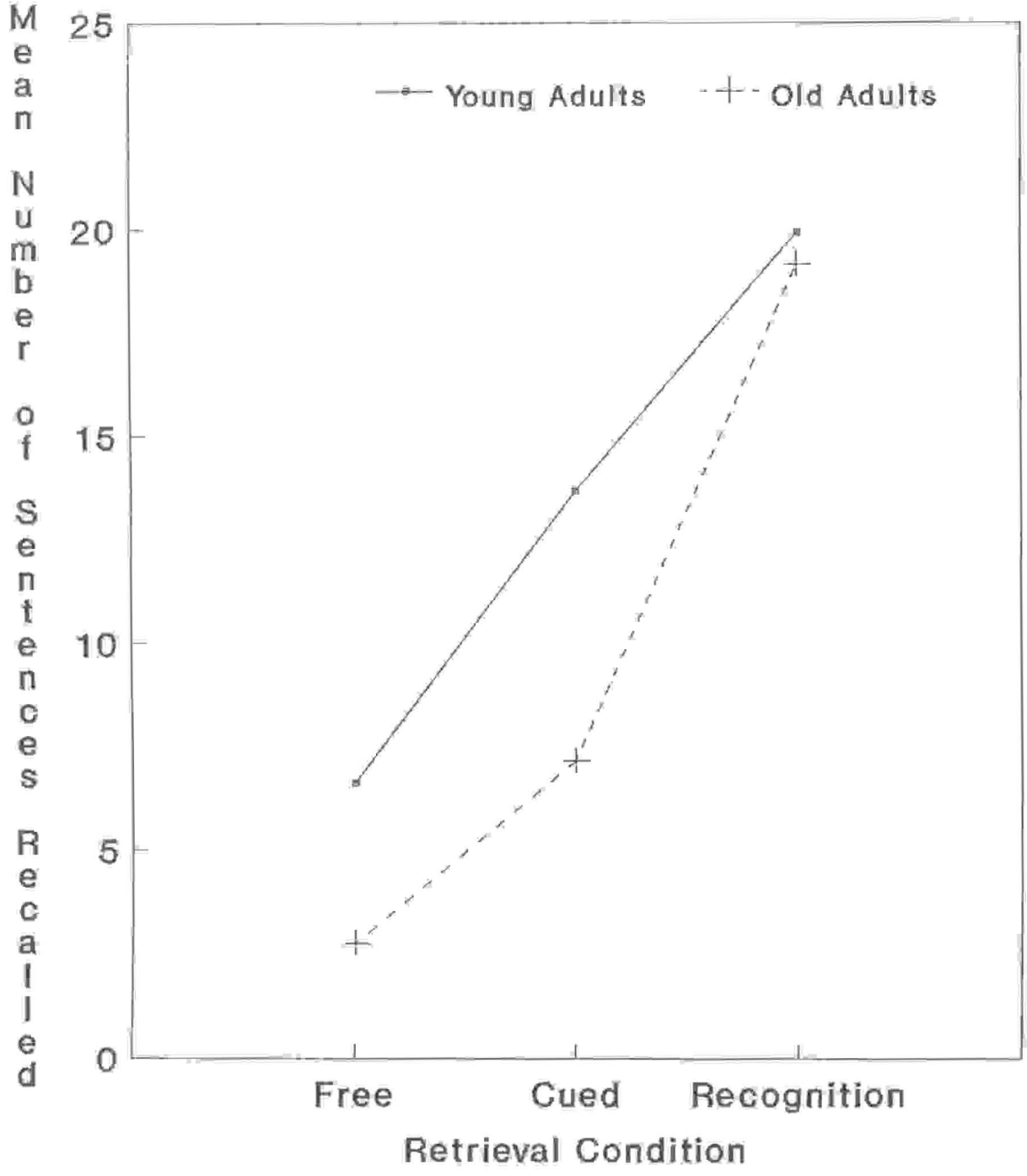

Figure 6. Mean number of sentences strictly recalled in task 2 as a function of age and retrieval condition.

VITA

Name:	Diane Patricia Fox
Place of Birth:	Calgary, Alberta
Date of Birth:	April 10, 1964

Educational Institutions Attended:

UNIVERSITY OF CALGARY	1982-1986
UNIVERSITY OF VICTORIA	1986-1989

Degrees Awarded:

B.Sc. (Honors)	1986	University of Calgary
M.A.	1989	University of Victoria

Awards:

Louise McKinney Scholarship	1983/84
Southam Communications Award	1984/85
Viscount Bennett Award	1984/85
University of Calgary Merit Award	1984/85
Louise McKinney Scholarship	1985/86
Faculty of Social Sciences Gold Medal	1986
Natural Sciences and Engineering Research Council of Canada Postgraduate Scholarship	1986-1988

PARTIAL COPYRIGHT LICENSE

I hereby grant the right to lend my thesis to users of the University of Victoria Library, and to make single copies only for such users or in response to a request from the Library of any other university, or similar institution, on its behalf or for one of its users. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by me or a member of the University designated by me. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Title of Thesis:

ADULT AGE DIFFERENCES IN MEMORY FOR METAPHORIC SENTENCES:

THE ROLE OF IMAGERY AND SEMANTIC FEATURE MODELS

Author [redacted]

DIANE FOX

February 9, 1989

ISBN 0-315-50126-X