THE UNIVERSITY OF MANITOBA

DEPTH OF PROCESSING: IS THERE A SEMANTIC CONTINUUM?

by

Donald Wilson McMurray

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE

DOCTOR OF PHILOSOPHY

DEPARTMENT OF PSYCHOLOGY

WINNIPEG, MANITOBA

December, 1978

DEPTH OF PROCESSING: IS THERE A SEMANTIC CONTINUUM?

BY

DONALD WILSON McMURRAY

A dissertation submitted to the Faculty of Graduate Studies of the University of Manitoba in partial fulfillment of the requirements of the degree of

DOCTOR OF PHILOSOPHY

ABSTRACT

The levels of processing approach to memory proposed by Craik and Lockhart (1972) has recently come under attach because of the circularity that has been inherent in relating depth to the strength of the memory trace (Nelson, 1977). Seamon and Virostek (1978), however, derived separate theoretical and empirical orderings of a range of classification tasks from across the various processing domains. A significant rank correlation between the ordered scale and stimulus word free recall strengthened the claim of earlier studies (Craik & Tulving, 1975; Hyde & Jenkins, 1973) that memory is a by-product of perceptual/cognitive analyses performed on the stimulus at input. This research sought to extend Seamon and Virostek's findings by operationalizing the term 'depth' and testing its predictive capabilities within a single domain in a series of incidental learning studies.

The framework for the six experiments reported here was the recent revision of the depth of processing idea advanced by Lockhart, Craik and Jacoby (1976), i.e., the notion of hierarchically organized domains of processing. Specifically, it was hypothesized that within the semantic domain more elaborate semantic analyses, defined in terms of the qualitative nature of the mental operations required to perform the task, would yield superior recall.

Experiment I involved subjects rating a variety of semantic or meaning producing activities against criteria that were postulated as underlying depth of processing. Subjects assessed each task in terms of the amount of conscious effort or attention required to carry out the task. A depth-ordered scale was derived from paired-comparison judgments collected on nine semantic tasks. The tasks ranged from judging whether a word was living or non-living to providing a short definition to determining how two unrelated words were similar.

In two follow up experiments (Experiment II & III) independent groups of subjects were unexpectedly tested for recall of lists of 24 unrelated words after engaging in one of the scale-ordered tasks that required them to generate their own encodings. Predictions based on the scale were empirically validated for most tasks when stimuli were low (Experiment II) or medium (Experiment III) in meaningfulness.

Additional experiments explored the generalizability of the findings of the first two studies by equating for study time in a within subject design (Experiment IV) and introducing a self-paced procedure in Experiment V. Experimental demands were also varied so that subjects rated words in all cases. Under these circumstances the relationship between the depth-ordered scale and the retention functions associated with each task broke down.

Experiment VI employed the original experimental procedures (i.e., subject generated encodings) but retained highly meaningful stimuli as in Experiments IV and V. Although order of recall correlated well with the scale the effect was not as strong as in Experiments II and III and differences across conditions were statistically indistinguishable.

It was concluded that the effectiveness of depth of processing within the semantic domain is a function of level of meaningfulness of the material under study vis à vis the stored contents of existing cognitive structures, the qualitative nature of the encoding operations and the extent to which the learner is actively engaged in those operations. Results were interpreted in the light of Craik and Jacoby's (1975) two process model of short-term retention.

ACKNOWLEDGEMENTS

I extend a deep feeling of gratitude to John McIntyre who supervised the development of this research project in his capacity as chairman of the dissertation committee. In addition to offering advice and continued encouragement at every stage, John gave freely of that which I have come to know and value over the years, his self. This provided me with the necessary resolve and direction to complete this enterprise.

For their valuable criticisms and suggestions I acknowledge a debt to Dr. Murray Singer and to Dr. John Whiteley, both of whom contributed substantially in setting the study in perspective in terms of its overall scope and objectives. My thanks also go to Dr. Joanne Rogan for her comments relating to the analysis and presentation of the data.

I am extremely grateful to Joanne Stothers for the innumerable hours she devoted to typing, proofreading and correcting the various drafts of the manuscript.

Finally, my deepest thanks are for Anne who stood at my side, and whose constant emotional support and solace were a source of strength and inspiration to me through all those times when the way ahead seemed unclear.

TABLE OF CONTENTS

TABLE OF CONTENTS (CONTINUED)

TABLE OF CONTENTS (CONTINUED)

LIST OF TABLES

DEPTHS OF PROCESSING: IS THERE A SEMANTIC CONTINUUM?

The levels of processing view propounded by Craik and Lockhart (1972) has exerted a considerable influence on the direction of recent research in human memory. Although originally intended as an alternative conceptual framework to multistore models of memory it has assumed the status of a theory especially as it relates to control processes in short-term retention. In the original paper, memory for an event was conceived of as a by-product of perceptual analyses with higher levels of recall reflecting deeper levels of processing at input. Preliminary shallow analyses were thought to be concerned with physical or structural features of the stimulus. Progressively deeper analyses were more concerned with extracting meaning attributes and/or the developing of associative relationships with the stored contents of existing cognitive structures. The depth to which a stimulus was analyzed was, therefore, thought to be a function of the type of analysis or level of encoding employed as well as the attention required of the central processor for the analysis. This too, varied according to the familiarity or level of meaningfulness of the to-be-remembered (TBR) item. It was initially thought that highly meaningful material was processed much faster than low meaningful material.

The memory trace was the end-product with deeper initial encodings leading to richer, more durable traces. This notion supplanted the need to postulate separate memory stores along with a variety of mechanisms for transferring material from one store to another. Memory was simply characterized as the by-product of various information processing activities or stimulus analyzing operations.

The depth of processing approach, although partially formulated on the basis of a review of some of the earlier studies of Jenkins and his associates (Hyde & Jenkins, 1969; Johnston & Jenkins, 1971), generated an enormous amount of research following the Craik and Lockhart paper. Most of the findings provide experimental support for the notion that deeper encodings result in better retention for both free recall (Craik & Tulving, 1975; Epstein, Phillips & Johnson, 1975; Hyde & Jenkins, 1973) and recognition (Craik & Tulving, 1975; Elias & Perfetti, 1973; Seamon & Murray, 1976). Most of these studies involved the use of an incidental learning paradigm where an attempt was made to gain control over the encoding strategies adopted by the subject. Usually the subject was required to perform an orienting task by judging selected characteristics of the stimulus materials. It is assumed that engaging a subject in an orienting task provides a fairly strong means of controlling that subject's encoding operations (Postman, 1976). Typically, in a levels of processing experiment, the material is held constant and the task demands are varied for each of several groups of subjects. For example, Goldman and Pellegrino (1977) presented three different orienting tasks to subjects. If the stimulus to be processed was COPPER then the three tasks thought to represent progressively deeper levels of analysis were as follows: (a) "Does the word contain the letter p?" (b) "Does the word rhyme with 'hopper'?" and (c) Gold is a metal. Is ________?" Results indicated that the deeper processing suggested by task (c) produced better memory than did the more superficial, less meaningful, processing entailed in tasks (a) and (b).

This new look in memory research represents a partial shift away from traditional variables (e.g., stimulus meaningfulness, study time and associative strength between items) to a focus on the qualitative nature of the mental operations at input. Deeper levels of processing presumably yield richer, more durable memory traces. The approach has not been without its detractors, however, and as we shall see there are still problems, especially as they relate to determining what criteria can be found that will predict, a priori, that a particular level of encoding will induce a better memory trace than another.

Recently minor modifications have been made to the concept of depth of processing. Factors such as stimulus elaboration (Craik & Tulving, 1975), distinctiveness of encoding (Moscovitch & Craik, 1976; Lockhart, Craik & Jacoby, 1976), precision of encoding (Stein, 1978) attempt to focus more closely on the mechanisms mediated by the various orienting tasks operating at input.

The Craik and Tulving (1975) series of experiments, as Postman (1976) points out, represents the most detailed attempt to date to systematically explore the relationship between levels of processing of TBR words and their subsequent retention. These authors concluded that the notion of "depth" was too simplistic as an explanatory concept to account for the complete range of perceptual/cognitive analyses that might be performed on a stimulus during encoding. Instead they opted for the concept of "domains" of processing (Lockhart, Craik & Jacoby, 1976), a term suggested by Sutherland (1972).

Additonally, they speculated that even within a domain depth ought to be replaced by "elaboration" or "spread" of encoding since these terms were more descriptive of the nature of the analyses carried out at input. Nevertheless, the stress still remained very much on the qualitative nature of the encoding operations. In fact, Craik and Tulving (1975) adduced strong evidence to show that a fairly rapid and "minimal semantic analysis is more beneficial than an elaborate structural analysis (Experiment 5)" (p.289).

In adopting the idea of domains, Lockhart, Craik and Jacoby (1976) suggested that various characteristics of words (e.g., orthographic, phonemic or semantic) can be best thought of as existing at separate levels. Hence the concept of depth was retained in the sense that the domains can be hierarchically arranged. The important question became one of whether stimulus processing necessarily proceeds through each of these domains until the level of analysis demanded by the encoding task is reached. The authors themselves are somewhat vague on this point. At one stage (p.78) it is claimed that with practice a complete domain can be circumvented as when an accomplished reader bypasses the phonemic stage completely (something the present author agrees with entirely), yet later in the same paper it is stated (p.80) that input from the phonemic domain is a necessary prerequisite for word perception. This is difficult to reconcile with evidence (Cattell, 1885) that subjects are capable of perceiving whole words faster than they can a single letter. Obviously word perception does not depend totally on either letter-by-letter identification or some grapheme-phoneme recoding mechanism.

More recently Graesser and Mandler (1975) conducted a study that also indicated that domains may be bypassed with the learner proceeding straight to the target domain. These authors showed that subjects do not retain memory for the surface structure of sentences when forced to concentrate on semantic analyses. Mistler-Lachman (1975) also demonstrated that subjects can resolve deep structure or semantic ambiguity before detecting and resolving ambiguities at the surface structure level of sentences. Subsequently, one of the co-authors in the paper under review conceded that deeper semantic analyses can occur without any prior structural analyses (Moscovitch & Craik, 1976).

In conclusion, the status of depth as a concept descriptive of the types of analyses that exist either within or across domains remains unclear. The present study explores the plausibility of utilizing depth as an explanatory concept within a domain (e.g., the semantic domain) and to assess whether qualitatively deeper stimulus analyses will yield a richer memory trace.

Certainly there is some experimental support for Craik and Tulving's (1975) suggestion that within a particular domain depth may be better characterized by the "spread" or number of encoded features rather than some ordered levels of analysis. Within the semantic domain, for example, Klein and Saltz (1976) found that subjects encoding along two uncorrelated dimensions evidenced better recall than subjects who elaborated the stimuli in terms of a single encoding dimension. In attempting to examine what kinds of existing cognitive structures might be involved in semantic deep-level processing these authors initially estimated correlations between the dimensions "happy-sad," "big-little," "fast-slow," and "pleasant-unpleasant," based on subjects' ratings in a pilot study. Subsequently, under incidental learning conditions, subjects evidenced better recall after rating words on two dimensions than on a single dimension, especially if the dimensions were uncorrelated, e.g., "pleasant-unpleasant" and "big-little" (where the correlation between the ratings was .05).

However, the phrase "greater depth" implies qualitatively deeper levels of cognitive analysis, especially within the semantic domain. Moreover, it is intuitively appealing to think in terms of a semantic continuum since the concept of semantic processing is so diverse and hitherto generally not well understood. Added to this is the idea that it is highly improbable that all semantic processing is equivalent. Given a gradation of memory performance contingent on type of depth of processing, the continuum might eventually be used to examine more closely the psychological nature of semantic processing itself (Baddeley, 1978).

There is some empirical support for a depth-ordered semantic continuum. For example, Epstein, Phillips and Johnson (1975) demonstrated that asking subjects to find similarities between pairs of words with different meanings or differences between words that were similar in meaning led to better recall than indicating similarities for similar words or differences for different word pairs. The implication is that encoding deeper semantic features produced a more elaborate memory trace. These results are, of course, entirely compatible with Lockhart, Craik and Jacoby's (1976) contention that "a reasonably familiar pattern or stimulus - response sequence [i.e., judging similarities for similar words] will be analyzed and encoded by a moderate number of analytic operations and will result in a moderately rich memory trace . . . Conversely, if the stimulus is novel, or difficult to process. . . more analyses are carried out and a richer memory trace results" (p. 79). Notice too, that the different semantic tasks in the Epstein, Phillips and Johnson study could be characterized as requiring differing amounts of conscious effort or attention in order that the task be successfully carried out. This is important because memory then becomes not only a function of the number of encoded features but it also is dependent on the qualitative nature of the cognitive analyses, i.e., deeper analyses require more conscious attention and hence produce richer and longer lasting memory traces.

Defining Depth of Processing

There has been a good deal of criticism aimed at the levels position in terms of the circularity that exists with respect to an adequate definition of depth of processing (Baddeley, 1978; Goldman & Pellegrino, 1977; Nelson, 1977). Generally, depth has been defined in terms of degree of stimulus elaboration and is reflected in increased levels of recall or recognition. Similarly, it is argued that well-remembered stimuli must have been more deeply processed. Processing tasks are chosen because researchers intuitively feel task A requires more elaborate processing than task B. Most of the processing activities fall into semantic and non-semantic categories, but as some authors suggest (Postman, 1975; Postman & Kruesi, 1977) no independent criteria for distinguishing semantic from non-semantic tasks have been developed. Consequently, until recently, the only validation for any "levels" notion has been the memorial consequences of the various orienting tasks themselves.

In an attempt to induce depth of processing and to break the circularity surrounding the concept, Craik and Tulving (1975, Experiment 2) explored the possibility that response latency might be associated with depth of encoding. Subjects were asked one of three types of questions: 1) "Is the word in capital letters?" 2) "Does the word rhyme with PALACE?" 3) "Does the word fit the sentence 'He lived with his mother and father in a ____________'?" Results confirmed that as decision latency increased for Tasks 1), 2), and 3) (thought to reflect ever deeper levels of processing) so did recognition performance. It could be argued that these findings are predictable on the basis of total study time. In re-analysis of the data from Experiment 2 and a subsequent study (Experiment 5) the same authors found a negative correlation between study time and recognition performance and hence concluded that retention is not solely a function of study time. (For a fuller discussion of the Craik and Tulving, 1975, series of experiments see Appendix A.) That processing time is not a major determinant of memory performance has also been suggested by other studies (Carpenter, 1974; Gardiner, 1974; Goldman & Pellegrino, 1977; Seamon & Murray, 1976).

In response to Nelson's (1977) argument that depth as a construct was scientifically meaningless due to the lack of suitable definition Seamon and Virostek (1978) derived an ordering of classification tasks based on subject-defined depth of processing. Subjects were instructed to order thirteen tasks (suggested by Nelson as being representative of the various processing domains) in terms of the depth of processing required by each task. Depth was defined as the degree of difficulty associated with each task. The subject-defined processing depth was tested for its predictive capabilities in an incidental learning paradigm that employed the same questions with different subjects. A significant correlation between the median ranks of the classification questions and their associated free-recall scores strengthened the claim that memory performance is a positive function of increasing depth of stimulus processing. Additionally the requirement that depth of processing be defined independently of memory performance was met at least for a range of tasks drawn from across the orthographic, phonemic and semantic domains. It remains true that no one has produced an independent depth-ordering of processing tasks within a single domain.

Levels within Levels

Baddeley (1978) lamented the fact that there has been no real attempt to delineate levels within any one domain be it structural, phonemic or semantic. Within the semantic domain, for example, there is a need to further differentiate levels of analysis and to assess the relative effects of various kinds of semantic processing tasks on memory (Craik & Tulving, 1975; Klein & Saltz, 1976; Schulman, 1974; Seamon & Murray, 1976). As stated earlier, it is the intention of the present study to explore the proposition that within a single domain different encoding processes will produce different memory traces.

Despite Baddeley's (1978) claim that he is unaware of any evidence suggesting that deeper levels of processing within the semantic domain lead to better retention both the Epstein, Phillips and Johnson (1975) and the Klein and Saltz (1976) studies described above tend to support the notion. In the latter case the authors anticipated a richer memory for the two-dimension rating group because it was reasoned that more semantic attributes of the stimulus were actively encoded. This additive notion of encoding of stimulus attributes has received experimental support elsewhere, at least for free recall (Goldman & Pellegrino, 1977). Others too, have found that variable encoding of the same items tends to facilitate retention (Martin, 1968; Nelson & Hill, 1974). Within the phonemic domain, however, Nelson (1977) found that subjects required to make two different decisions for each of two repetitions retained these words no better than subjects making the same decision on each of two repetitions. Also, Hyde (1973) found that subjects who processed words on two semantic dimensions showed no better retention than either of the two groups who processed TBR words on a single semantic dimension only.

Despite the kinds of objections registered by Nelson (1977) and Nelson and Vining (1978) depth of processing continues to enjoy a great deal of attention and the amount of research the concept is generating, shows no sign of abating. While it is generally conceded that the reliability of depth of processing as a phenomenon, especially across domains, has been adequately demonstrated it is equally certain that the validity of the concept has not been fully documented. Because of this, the concept lacks both explanatory and predictive powers. For example, how do we determine, a priori, whether generating a free associate in response to a stimulus or examining it to see if it fits meaningfully into a given sentence frame will lead to better memory for that stimulus? If depth of processing within the semantic domain is to have any construct validity then two things must occur:

1. The term "depth of processing" must be operationalized by clearly specifying the operations used to measure it. If depth implies ordinality then there must exist an independently devised depth-ordering of processing tasks within the semantic domain.

2. The resulting taxonomy of processing tasks must be tested empirically to see if the rankings can predict previously obtained retention differences.

It may be said that apart from establishing some predictive capability for the ordering of depth of processing tasks, this type of exercise will reveal little of the nature of depth of processing as a psychological concept. Yet it remains true that the theoretical principle that retention increases as function of depth of stimulus processing has not been fully explored. It is the contention here that deeper semantic encodings (to be independently and operationally defined) will lead to richer, more persistent memory traces.

Therefore, as a preliminary exercise, the present study began by scaling a variety of semantic or meaning-producing activities by having subjects rate all tasks in a series of paired-comparisons using criteria thought to encompass the concept of depth of processing. Following this the scale was empirically validated in a series of incidental learning experiments. If tasks rated as requiring deeper semantic analyses in fact lead to stronger memory traces then an appealing notion would be confirmed.

EXPERIMENT I

The specific purpose of the initial experiment was to establish an ordinal scale for nine tasks requiring processing English words and which have been designated as involving semantic operations by 'levels' researchers in recent years (Bellezza, Cheeseman & Reddy, 1977; Craik & Jacoby, 1975; Craik & Tulving, 1975; Hyde & Jenkins, 1969, 1973; Mistler-Lachman, 1975; Nelson, Wheeler, Borden & Brooks, 1974; Shulman, 1971).

The nine tasks selected were not exhaustive of the kinds of semantic tasks employed but were chosen so as to be representative of the range. Intuitively they were thought to vary with respect to the complexity of the cognitive operations involved.

Method

Subjects. Twelve undergraduate students from a second year experimental psychology course were recruited to perform the rather demanding rating task which involved Fechner's (1860) method of paired comparisons (see Torgerson, 1958). It was reasoned that as psychology majors the subjects would be more committed to the task demands of the experiment than any random sample of psychology undergraduates. Subjects received partial course credit for their participation.

Rating tasks. The nine tasks that were selected for rating were as follows:

1. Write a few words saying how the two words are different. (DIFF)
2. Living-Nonliving. Subjects were asked to circle each word as 'Living' if the word described something that was in their opinion living, otherwise they were to circle the word 'Nonliving.' Similar

judgments were required for tasks 3 and 7. (L-NL)

3. Strong-Weak (S-W) - see Task 2.
4. Write a single meaningful sentence using both words. (2-WS)
5. Write two free associates for each word, i.e., which are the first two words that come to mind when you think of the word. (2-FA).
6. Write a meaningful sentence incorporating each word. (1-WS)
7. Pleasant-Unpleasant (P-UP) - see Task 2.
8. Write a few words saying how the two words are similar (SIM)
9. Write an appropriate definition for each word. (DEF)

Procedure. Each subject was presented with a four-page booklet containing the 36 paired comparisons he/she was to make. Comparisons were made with respect to a specific pair of English nouns which was different for each subject. The 12 word pairs were chosen from Paivio, Yuille and Madigan's (1968) norms. The nature of these word pairs is more fully described in Experiment II. Each subject was randomly assigned a word pair and the order of pages containing the ratings was randomized across subjects with the additional requirement that half of the subjects would rate from the bottom to the top of each page and the other half from the top to the bottom. The ordering of tasks on successive pages was randomly determined with the proviso that no task was assigned the same serial position on more than one page.

As part of the general instructions the subjects were given practice with each task by collectively working through the list of nine tasks, keeping the pair 'pliers-crab' in mind. The Experimenter demonstrated the procedure associated with each task. Following this the subjects were told that they would be required to compare each task to every other task in succession and to decide for each particular pair which task involved more conscious effort or attention.

At this juncture the Experimenter elaborated on the nature of depth of processing and its relationship to terms such as "depth" or "spread of encoding" as portrayed in the literature (Craik & Jacoby, 1975; Craik & Tulving, 1975; Lockhart, Craik & Jacoby, 1976). Illustrations using tasks from the non-semantic domains were used to sharpen the concept in the minds of subjects prior to the rating task. Care was taken to emphasize that depth was not necessarily to be equated with time but may likely have more to do with the complexity and qualitative nature of the encoding operation or the amount of conscious effort or attention required to carry out the task. Examples, again drawn from the non-semantic domains, were used to bolster this point. Subjects were then given a word pair to practice on for themselves and encouraged to think about the requirements of each task. The order in which tasks were practised was randomly determined for each subject. Following practice with each type of task (see Appendix B for Practice Sheet) subjects were given a single word pair (which was subsequently to be used in a later experiment) and proceeded to indicate which task of all possible pairs required more conscious effort or attention (see Appendix C for an example of the paired-comparisons task). Although encouraged to be consistent in their ratings the subjects were expressly instructed not to refer back to earlier ratings but rather to re-focus on the criteria for rating if uncertain about a particular pair (see Appendix G).

Results and Discussion

Throughout this paper an effect is considered significant using