1975

Closed-loop theory and the partial recall hypothesis : explanations of the sources of information about knowledge in memory

Thomas Beverly Watkins

CLOSED-LOOP THEORY AND THE PARTIAL RECALL HYPOTHESIS: EXPLANATIONS OF THE SOURCES OF INFORMATION ABOUT KNOWLEDGE IN MEMORY

BY

THOMAS BEVERLY WATKINS

A THESIS
SUBMITTED TO THE GRADUATE FACULTY
OF THE UNIVERSITY OF RICHMOND
IN CANDIDACY
FOR THE DEGREE OF
MASTER OF ARTS
IN PSYCHOLOGY

AUGUST 1975

Thesis Committee

L. James Iannotta _______________ Chairman

Kenneth A Black _______________

Frederick T. Kozub _______________

Thomas B. Watkins _______________ Student

Abstract

Examples of information about knowledge in memory are described, and two conceptualizations of the source of such information--the partial recall hypothesis and the closed-loop theory--are reviewed. Wearing (1970) conducted a study to support the closed-loop theory using 60 CVC pairs in a paired-associate task with a recall measure and confidence ratings. An attempt is made to replicate and extend some of his findings. Some are replicated, except for one finding with which he supported closed-loop theory. With support for closed-loop theory thus reduced, the partial recall hypothesis seems more plausible.

Closed-Loop Theory and
The Partial Recall Hypothesis

Decision-making in learning, like decision-making in any other area, requires valid information. Teachers can use the information provided by pencil and paper tests to help them make decisions concerning the achievement levels of students, but the students seldom have the information from these tests until after they have made important decisions. They may rely on information from other sources to decide when to stop studying, continue or review, and when to change approaches or subjects. (A related area, the response mode issue, has been reviewed by Anderson, 1970, and Tobius, 1973.)

Experience, partial recall, and feelings of familiarity are three possible sources. First, previous experiences with a subject area or similar subject areas enable individuals to estimate the amount or type of study required. The more learning experiences a person has had with a particular area or similar areas, the better he can guide his own learning through that area.

Second, as a person reviews or previews an area, the number of parts or attributes that come to mind automatically provides important information. For instance, when reviewing or previewing Bayesian statistics, a student may encounter the terms "maximum likelihood ratio" and "interval estimation." The number of related pieces of information that come to mind automatically provides a clue to the amount of knowledge of these areas that is already available.

Third, the student uses information coming from general feelings of familiarity produced as he previews or reviews a particular area.

All three sources of information require no overt responding, no recitation of the material, as required by the usual "straw man," stimulus-response (S-R) theory. In spite of the absence of overt responding, people possess varying amounts of information about the knowledge they have in memory. It is possible that, overall for the process of learning, learners rely more heavily on information from informal sources like the three mentioned above than on information from formal sources like paper and pencil tests.

S-R theory would have trouble accounting for this. In fact, Tulving and Madigan (1970) suggest that no theory has incorporated "one of the truly unique characteristics of human memory: its knowledge of its own knowledge." The literature they survey contains at least five examples of such knowledge: tip-of-the-tongue research, feeling-of-knowing research, confidence rating research, and two more described by Adams (1967). I will describe these examples before detailing two ways that S-R theory can be extended to incorporate the source of these examples of information about knowledge in memory.

Information about Knowledge

Everyone has had occasions when he could not recall a name or some other piece of information which he was sure he knew and which he even felt he had on the tip of his tongue. Woodworth and Schlosberg (1954, pp. 719-720) reported studies conducted early in this century and studies by Woodworth (1934) himself, in which tip-of-the-tongue instances were collected from everyday experiences. In the laboratory, Brown and McNeill (1966) produced an experimental demonstration of the validity of tip-of-the-tongue experiences. They read to their subjects the definitions of words of Lorge-Thorndike frequency low enough that many words could not be recalled but high enough that many of them produced a tip-of-the-tongue experience. Of the tip-of-the-tongue experienced words, the subjects had some knowledge about number of syllables, stress positions, and some of the letters and their positions in the words. They also were able to recall words of meaning and sound similar to the tip-of-the tongue experienced words.

The second example of information about knowledge in memory is provided by Hart (1965) and Freedman and Landauer (1966). They conducted research with a design similar to Estes' (1960) miniature experiment. Estes gave a sequence of test trials following one reinforced trial, supporting his one-trial-learning arguments with the conditional probabilities of the outcome of a test trial given the outcome of an earlier test trial. Landauer (1962) placed a matching test between the two test trials for a comparison with Estes.

Hart's design consisted of a basic sequence of a recall test and a recognition test with a judgment between them. His subjects either made ratings on a 6-point scale (1965, Exp. 2; 1967a, Exp. 2) or made a binary (yes/no) judgment (1965, Exp. 1; 1966; 1967a, Exp. 2; 1967b; 1968) about whether or not they felt they could recognize an item on a 4-alternative multiple choice test, given that they could not recall the item. Hart used general knowledge questions except for two experiments in which he paired words with consonant trigram syllables in a paired-associate learning task (1967a). Hart reported significant differences (by either sign or _t_ test) between the number of items which were not correctly recalled but were correctly recognized given a feeling-of-knowing following the recall test, and the number of items which were not correctly recalled but were correctly recognized given a feeling-of-*not*-knowing following the recall test. His subjects had a feeling-of-knowing on the majority of the items they had failed to recall but later recognized.

Freedman and Landauer (1966) used 150 general knowledge questions and a design consisting of an uncued recall test, a confidence rating, a cued recall test, and finally a recognition test. They reported a significant F value for an analysis of variance on the proportions of unrecalled items later recognized across the four confidence rating categories. They suggested the existence of a direct relationship between the degree of accuracy of recognition and the degree to which the subject is confident that he has learned the response even though he cannot recall it at that time.

This study by Freedman and Landauer demonstrated the similarities between feeling-of-knowing ratings and the third example, confidence ratings. Both consisted of either ratings or binary yes/no judgments. In the former the subjects estimated the accuracy of their potential or future responses, and in the latter they estimated the accuracy of past or actual responses.

Jersild (1929) provided one early example of research on confidence ratings. More recently, Pollack and Decker (1958, also in Swets, 1964) and Clarke (1960, also in Swets, 1964) used confidence ratings in psychophysical research (see also Green & Swets, 1966). Murdock (1974, pp. 117-121) suggested that the use of confidence ratings leads to the theoretical separation of memory and decision processes. Most evidence certainly points to a direct relationship between confidence ratings and response accuracy, reflecting the subjects' ability to discriminate between correct and incorrect responses.

Adams mentioned the final two examples, omission and error rejection behaviors. He stated that the first of these occurs when a subject withholds a covert correct response or rejects an overt one because he incorrectly believes it to be wrong, and the second occurs when the subject makes an incorrect response and, as he gives it, realizes that it is wrong and rejects it.

Two extensions of S-R theory can account for these five examples of information about knowledge in memory. (See Greenwald, 1970, for other theories.) The partial recall hypothesis is presented first because it is somewhat simpler than the closed-loop theory.

Partial Recall Hypothesis

The attribute theories (Bower, 1967; Horowitz & Prytulak, 1969; Underwood, 1969, 1972) have made the partial recall hypothesis seem plausible. If a subject can recall parts and attributes of the correct response even though he cannot recall the whole response, he can use these parts and attributes as information for various decisions. He will say he has it on the tip of his tongue if he is close to recalling the whole response, and he will say he feels he knows it if he can recall a certain type or number of parts and attributes. He knows he has a better probability of recognizing items he cannot recall if he can recall some parts and attributes. He can use the parts and attributes that he can recall to help distinguish the correct response from incorrect responses on a multiple choice test.

Blake (1973) used a variation of the feeling-of-knowing design to test the partial recall hypothesis. He used a short-term memory paradigm to reduce inter-item interference. Also, on the recognition test, he reduced the advantage from being able to recall some of the letters of consonant trigram syllables. Those letters of the syllables which the subject could not recall correctly on the recall test were exactly the letters by which the alternatives differed on the multiple choice test. Thus the subjects could not use the letters they could recall to help discriminate between the correct alternative and the incorrect alternatives on the multiple choice test.

This reduction of the partial recall advantage reduced feeling-of-knowing accuracy but did not eliminate it. The advantage produced by other parts and attributes which had also helped give the subject a feeling-of-knowing still were not eliminated on the multiple choice test, and these helped subjects select the correct response.

One problem with the partial recall hypothesis is the question of how the subjects know which recalled parts and attributes or whole responses are correct. Perhaps another attribute--ease or automaticity of recall--helps explain the ability of subjects to distinguish between correct and incorrect. Incorrectly recalled items or parts and attributes usually are slower or more difficult to produce. Subjects use their perceptions of the speed or ease of recall to decide how confident they should be of the recalled parts, attributes or whole responses.

In summary, the attribute of automaticity of recall and the recall of parts and attributes can be used to explain the source of information about knowledge in memory. As will be seen, this extension of S-R theory requires less elaboration than closed-loop theory.

Closed-Loop Theory

Adams (1967, 1968) combines aspects of both sign significate (S-S) theory and S-R theory in a more elaborate conceptualization of the source of information about knowledge in memory. He suggests that during learning two types of traces are formed between the stimulus and the response: the memory and perceptual traces. The memory traces produce the response either covertly or overtly when cued by the stimulus. Covert responses are produced by the thinking process and become overt when spoken or written.

Adams' notion of the perceptual traces derives from S-S theory, which he extends by giving the perceptual trace the ability to indirectly reinforce the memory trace in the following fashion. A memory trace cued by a stimulus produces a covert response. This covert response produces perceptual traces of the stimulus-response association. These covert-response-produced perceptual traces are compared with the original-learning-produced perceptual traces to determine the correctness of the covert response. If the two sets of perceptual traces match, the subject can recognize the covert response as correct. When he gives the response overtly more perceptual traces are produced which can be compared with perceptual traces formed during original learning. If a match again occurs, the memory trace is further strengthened. The strengthening process results from the conscious application of rehearsal strategies such as repetition or mediation by the subject whenever he has recognized a correct response.

This theory is also known as the two-factor feedback model, which is predicated on a definition of reinforcement as knowledge of results. Bilodeau and Bilodeau (1961) have discussed knowledge of results as the perception of any discrepancies between intended and actual behavior. The feedback part of the theory is analogous to theories of proprioception, kinethesis and systems analysis from engineering psychology. One factor, already described as the memory trace, is similar to habit in S-R theories, whereas the other factor, the perceptual trace, is similar to Mowrer's (1960) concept of the conditioned sensation or image.

Adams accepts the two-stage theory of paired-associate learning with response and association learning stages. He states that paired-associate learning requires the ability to recognize both stimuli and responses.

He assumes that recall and recognition depend on different structures. The memory traces are responsible for recalling responses; the perceptual traces are responsible for recognizing stimuli and responses. Perceptual trace strength and therefore recognition performance depend on frequency of exposure to stimuli and responses, whereas response trace strength depends on frequency of reinforcement. Of course, for both recognition and recall, performance depends on interference and trace strength. These in turn depend on pre-experimental trace strength and experimental manipulations which affect trace strength. Montague (1972) provides more discussion of the relationship of Adams' theory to other studies of learning and memory.

A number of studies (Adams & Bray, 1970; Adams, Marshall & Bray, 1971; Adams, McIntyre & Thorsheim, 1969; Wearing, 1971) offer evidence for this theory. Most of these studies involve confidence ratings. Confidence ratings offer support for closed-loop theory in that the subjects are presumed to use their perceptions of the discrepancy between response-produced perceptual traces and learning-produced perceptual traces to determine their confidence ratings. If they perceive a large discrepancy, they give a low rating.

Similarly, tip-of-the-tongue or feeling-of-knowing experiences are presumed to depend on the subjects' perceptions of the same type of discrepancy. Although the memory traces may not be strong enough in tip-of-the-tongue and feeling-of-knowing experiences to recall the whole response, they are strong enough to produce parts and attributes of the correct response. The subjects can perceive the discrepancy between the perceptual traces produced by the recalled parts and attributes and the perceptual traces produced on the learning trials. Through the use of confidence ratings, the study by Wearing provides one of the best examples of support for closed-loop theory.

Wearing's Experiment

Wearing provides some evidence for closed-loop theory by looking at the way confidence rating data vary when memory and perceptual traces are manipulated separately. He uses the following rationale. Because perceptual trace strength is determined by frequency of exposure as represented in measures of familiarity, he chose learning materials which vary in Archer's (1960) association value (AV). Items high on Archer's AV scale should be high in familiarity, since Goss and Nodine (1965) found AV to be highly correlated with familiarity. High AV items should have higher pre-experimental perceptual trace strength than low AV items. Archer determined the AV of each CVC by asking his subjects four questions about it: Is it a word? does it remind you of a word? does it sound like a word? can you use it in a sentence? AV for any particular CVC is the percentage of subjects who could answer at least one question affirmatively.

In Adams' theory, memory trace strength is determined by measuring the level of learning of the associations. Because mediational devices such as natural language mediators (NLM's) are good indicants of high levels of learning (Kiess, 1968; Montague & Wearing, 1967), items with a high potential for mediational devices should have higher pre-experimental memory trace strength than items with a low potential. Likewise, items for which NLM's are reported should have had higher pre-experimental trace strength. Montague and Kiess (1968) provide a scale of the NLM potential of pairs of consonant-vowel-consonant (CVC) syllables sampled across the full range of the AV scale. Both CVC's in a pair have the same value on the AV scale. Montague and

Kiess call their scaling of CVC pairs for NLM potential the associability scale (AS). Wearing assumes that items with high AS should have higher pre-experimental memory trace strength than low AS items. Montague and Kiess determined the AS value of each pair of CVC's by asking their subjects to report any NLM's they could think of for each pair. AS for any pair of CVC's is the percentage of subjects who reported an NLM.

To summarize, Wearing uses AV to vary pre-experimental perceptual trace strength and AS to vary pre-experimental memory trace strength. Readers who still have doubts as to the efficacy of this manipulation can read Wearing (1971) and his sources (Adams & Bray, 1970; Montague & Kiess, 1968; etc.).

Whereas both AV and AS can be considered measures of meaningfulness, AV perhaps may be more dependent on mere frequency of exposure, and AS may be more dependent on the redintegrative power of two CVC's (see Horowitz & Prytulak, 1969), or on the number of transformations necessary to integrate them into a meaningful mediation strategy (see Prytulak, 1971). The first CVC is used as the cue for the second, so that a strong bond between them is necessary.

To show the differences between AV and AS, the one can be controlled and the other allowed to vary. Another way to explore the differences between AV and AS is to look for differences between the effects of AS at low and high AV. When AV is low the effect of increasing AS is simply that of increasing the ability of CVC's to fit into complicated NLM's. When AV is high the effect of increasing AS is that of increasing the ability of CVC's to integrate into simple one-word NLM's (see Montague, 1972, p. 258). Such might be one way to explain an AV-AS interaction.

Unfortunately, complete crossover of AV and AS does not occur to the extent necessary to find--in sufficient numbers for an experiment--items which are high on one scale and low on the other. AV and AS are positively correlated and Montague and Kiess used only a small sample of all the possible pairs of CVC's. It is impossible to compare items with either high or low AS at both high and low AV. It is only possible to compare items with either high or high-medium AS at high AV (HH - HM) and to compare items with either low or low-medium AS at low AV (LL - LM).

Some researchers attempt to overcome the problem of incomplete crossover by performing a third comparison between the two previous comparisons. In short the prediction is that the difference between AS levels at high AV level is significantly larger than the difference between AS levels at low AV level. They assume that this third comparison gives an estimate of the interaction of AS and AV effects.

Instead of this third comparison, Wearing used the reports of NLM use in place of AS, probably assuming that an item for which an NLM is reported has for that subject a higher AS value and therefore higher pre-experimental memory trace strength than an item for which no NLM is reported. Wearing used 60 CVC syllable pairs scaled for AS and AV. Both CVC's in a pair had the same AV. There were 15 high AV, high AS (HH) pairs; 15 high AV, high-medium AS (HM) pairs; 15 low AV, low-medium AS (LM) pairs; and 15 low AV, low AS (LL) pairs. Both LL and LM groups of items had the same average AV, and both HH and HM groups had the same average AV.

Each pair was presented once for 15 seconds. Twenty-four hours later recall was tested by presenting the first member of each pair and asking for the second. Subjects were told that they would receive 2¢ for every correctly recalled item. They also were told to make confidence ratings on a 5-point scale for each response recalled. A second recall test immediately followed the first test so that subjects could report any NLM's they had used in learning any of the pairs.

Wearing found that his subjects recalled 68% more incorrect responses than correct responses and that they were quite capable of distinguishing between the correct and incorrect. The proportion of correct responses increased directly with AV, AS and NLM use.

Wearing used two types of conditional probability, probably as an alternative to receiver operating characteristic (ROC) curves. Both types of conditional probability are derived from a decision matrix of true positives (hits), false positives (false alarms), true negatives (correct rejections), and false negatives (misses). The decision matrix of confidence ratings and recall is shown in Table 1.

Insert Table 1 about here

The first conditional probability is the proportion of all "positives" which were also "true," or in other words the proportion of all responses with high confidence ratings which were also correctly recalled. Because a high confidence rating (a 5) depends on perceptual and memory traces of high strength, this conditional probability should be uninfluenced by AV level or reports of an NLM being used.

The second (or "reverse") conditional probability is the proportion of all "true's" which were also positive, or in other words the proportion of all correctly recalled responses which were also given a high confidence rating. Because a correct recall sometimes can occur with perceptual or memory traces which are weak, but high confidence ratings require strong memory and perceptual traces, this reverse conditional probability is influenced by AV and report of an NLM being used.

According to Wearing, the conditional probability of a response being correct, given a high confidence rating, is high and is "more or less" uninfluenced by AV, AS or NLM use. This conditional probability is the probability of recall conditional on confidence rating. Wearing states that the recognition of a correct response as correct--in other words, a confidence rating of 5--depends on perceptual and memory traces of high strength, too high to be influenced by variations in AV, AS or NLM use.

However, the reverse conditional probability of a high confidence rating given correct recall <u>is</u> influenced by AV, AS and NLM use. (The probability of A conditional on B does not have to equal the probability of B conditional on A.) Wearing states that recall can occur sometimes when either memory or perceptual traces have low strength although a high confidence rating requires both memory and perceptual traces to have high strength. Memory and perceptual trace strength increase with NLM use and AV.

Wearing also found that response latency decreases as subjective certainty increases. He concluded that his results provide support for Adams' dual trace model with the comparison process as the source of information about the information in memory.

His data, however, seem to fit the partial recall hypothesis almost as well as closed-loop theory. Wearing's data on response latency particularly support the automaticity notion. Subjects seem able to base their confidence ratings on the latency of their recall.

Of course Wearing's data on the conditional probabilities do not support this, in that the probability of correct recall given high confidence ratings appears to be uninfluenced by variation in AV, AS or NLM use. The appearances could be misleading, however, for by paying his subjects 2¢ for each correct recall, he lowered their response criterion to the point that they made 68% more errors of commission than correct recalls. With so many responses on which to make confidence ratings, subjects may have given high confidence ratings only to those responses of which they were absolutely certain. For the two levels of AV Wearing found probabilities of 88 and 81 for low AV and 90 and 95 for high AV. For the NLM presence or absence he found conditional probabilities of 88 and 90 for no NLM and 81 and 95 for NLM. These probabilities are so high that any differences among them are likely to be hidden.

If, in replicating Wearing's study without the 2¢ bonus, trends in the conditional probabilities are found, Wearing's support for Adams' theory can be questioned. Such a replication is performed here. In addition to the confidence ratings, feeling-of-knowing ratings are used. Feeling-of-knowing results should duplicate the results of the confidence ratings and provide further evidence for either the partial recall hypothesis or the closed-loop hypothesis.

As already mentioned, feeling-of-knowing ratings in Adams' theory should depend upon a certain minimum perceptual and memory trace strength, whereas recognition requires only perceptual trace strength. Therefore, whereas the results of feeling-of-knowing ratings should depend on both AV and AS, the recognition results should depend only on AV. If AS is found to have an effect on recognition results, then support for Adams' theory can be further questioned.

Method

Subjects. The experiment was administered to undergraduate volunteers from three psychology courses, in five groups ranging in size from twenty-one to three and to two other volunteers from those classes separately. All subjects received the same treatment. A total of 40 subjects participated. Of these, the responses of nine were later dropped: six because of failure to follow instructions, two because of failure to recall any items correctly, and one because of her report that she had stopped concentrating halfway through the procedure because of sleepiness.

Materials. The 48 CVC pairs were divided into four groups of items. The first two groups consisted of 12 high AV (mean=99, range=97-100), high AS (mean=95, range= 91-98) pairs (HH) and 12 high AV (mean=99, range=97-100), high-medium AS (mean=75, range=68-83) pairs (HM). These two groups of 12 pairs each had the same mean AV but different mean AS. The other two groups consisted of 12 low AV (mean=42, range=38-46), low-medium AS (mean=62, range=57-68) pairs (LM) and 12 low AV (mean=42, range=39-46), low AS (mean=33, range=25-39) pairs (LL). Both of these two groups had the same mean AV but different mean AS.

A single order of the 48 items was used for all subjects. To be certain that serial position effects were minimized, the order was divided into four positions. From each group of AV-AS levels, three pairs were randomly assigned to each of four blocks. Within each block the 12 pairs assigned to it were randomly ordered. The four blocks were then randomly assigned to the four order positions. Thus each block consisted of an equal number of pairs randomly selected from each AV-AS group of pairs, the pairs were randomly ordered within each block, and the blocks were randomly ordered. The same order was used on all learning and test trials.

The list was presented by slide projector. Each pair was typed on a single slide for presentation on the learning trial. For the recall test, the stimulus members of the pairs were numbered in the same order as on the learning trial and typed onto slides with the number, one stimulus member to a slide.

A test booklet was developed in which subjects were to record their responses. On the first page were the instructions for the recall test. On pages 2, 3 and 4 were the answer sheets for the recall tests, the ratings and the NLM reports. The answer sheets consisted of lines numbered 1 to 48 (16 lines to a page) on which the appropriate responses could be written. Beside each numbered line were two sets of the numbers 1 to 5 arranged in two columns on each page. Above the first column on each page was printed "confidence ratings." Above the second column of five numbers was printed "feelings of knowing." Beside the two sets of numbers was a longer line on which subjects could record their NLM's.

The recognition test was placed on the back and front of a separate page. It consisted of a multiple choice test format with the first member of each pair followed by four alternatives, one of which was the second member of the pair. The four alternatives all had the same first letter but differed in the last two letters. None of the incorrect alternatives had been paired with another first member in the learning trial. The items on the recognition test were in the same order as on the learning trial and recall trial.

Procedure. Before presenting the pairs, subjects were instructed that following seeing the pairs once, they would be given a recall test in which the first member of the pairs would be presented and they would have to recall the second member with which it had been paired. They were encouraged to try to concentrate on learning the pairs so they could correctly recall as many as possible of the second members. To be correct a recalled second member had to be matched with the first member with which it had been paired.

The pairs were then presented one at a time. Each pair was presented for 10 seconds with no pause between pairs, although in general it required several seconds to remove one pair from presentation and present the next pair.

Following presentation of the list of pairs, the subjects were each given an answer booklet and told to read the directions and listen to a tape recording of the directions being read. Subjects were instructed that the first member of every pair would be presented for 30 seconds and that those first members would be numbered. As a first member was presented the subjects were to try to recall the CVC syllable that had been paired with it. If they could recall it they were to write it down on the blank with the number that matched the number of the first member with which it had been paired. If they were able to recall a CVC syllable but were not entirely certain it was correct they were to write it down anyway. If they could not recall a syllable, they were to put an "X" on the line with the matching number. This step, along with numbering the lines, served to prevent any confusion in scoring.

Subjects were further instructed that if they could recall the response member they were to make a confidence rating by circling the appropriate number in the first column labeled "confidence ratings." The confidence ratings consisted of rating on a scale of 1 to 5 how certain or confident subjects were that the response was correct. The subjects were told that the numbers of this scale represented the following:

<u>Confidence Ratings:</u>

1 - Very confident it is not correct

2 - Fairly confident it is not correct

3 - Do not know one way or another

4 - Fairly confident it is correct

5 - Very confident it is correct

On all first members that were presented, subjects were instructed to make a feeling-of-knowing judgment by circling the appropriate number in the second column labeled "feelings of knowing." These judgments consisted of rating on a scale from 1 to 5 how certain or confident the subjects were that they would be able to recognize the correct response on a multiple choice test from among four alternatives, even though they might not be able to recall the response. Subjects were reminded that it is easier to recognize an item than to recall it. The subjects were told that the numbers of this scale represented the following:

<u>Feelings of Knowing:</u>

1 - Very confident I will not recognize it

2 - Fairly confident I will not recognize it

3 - Do not know one way or another

4 - Fairly confident I will recognize it

5 - Very confident I will recognize it

Subjects were instructed that when they recalled a response they were to write down any associative devices they had used to learn the pair. They were to record these devices on the line on the right hand side of the page beside the feeling-of-knowing rating. They were given an example of each of several types of NLM's.

Finally they were urged to concentrate to avoid careless mistakes and told to open to the second page. After ensuring that all the subjects' questions had been answered the experimenter presented the stimulus members of the pairs one at a time by slide projector. They were presented in the same order as in learning.

After completing the recall test, subjects were given the recognition test. They were instructed to select the alternative which they believed was the second member of the pair and put its letter in the space in front of the stimulus term. If they remembered an NLM which there was any possibility that they had not reported on the recall test, they were to write it down on the recognition test under the four alternatives. They would have 20 minutes to complete the recognition test. When they finished they were to close the test booklet.

Results

Recall, recognition, and NLM use are all directly related to AV-AS level (see Table 2). Because of the incomplete crossover of AS and AV, an assumption is made that a triple comparison can be conducted to obtain some estimate of the interaction of AS and AV that might be found if AV-AS crossover were complete. Earlier studies

---

Insert Table 2 about here

(Wearing, Walker & Montague, 1967; Montague & Kiess, 1968;

Walker, Montague & Wearing, 1970) have also made this assumption.

The triple planned comparison involves testing the difference between HH and HM, the difference between LM and LL, and the difference between the two differences. According to the assumption, unless all tests are significant, no estimate of interaction should be made.

The comparison between the HH and HM groups is significant for mean recall scores, mean recognition scores and mean NLM scores (see Table 3). The comparison between the LM and LL groups is significant for mean recall scores and mean NLM scores but not for mean recognition scores. The comparison of the two differences

Insert Table 3 about here

is significant only for recall. Although the difference between the HH and HL groups is always larger than the difference between LM and LL, AS and AV appear to interact only for recall and not for recognition and NLM use.

The relationships with recall, recognition and NLM use are tested separately for AV and AS. Both AS and AV are directly and significantly related to recall, recognition and NLM reports (see Table 3). The pooled means for high and low AV are also shown in Table 2. All F values reported are calculated with repeated measures on the subjects across all levels of either AV or AS.

Confidence rating level is directly related to correct recall (see Table 4). The χ^2 (1) value for this relationship is 331.44, which is highly significant. Whereas in Wearing's study 68% more errors of commission were given than correct responses, in this study 39% fewer errors of commission are given than correct re-

---

Insert Table 4 about here

---

sponses. Nevertheless, the proportions of both errors of commission and correct responses across the five levels of confidence ratings are close to the proportions found by Wearing.

The relationship between feeling-of-knowing rating level and recognition is significant and direct with a χ^2 (1) value of 227.44 (see Table 4). The relationship of feeling-of-knowing and recognition does not appear to be quite as strong as the relationship of confidence ratings and recall.

The conditional probability of high subjective certainty given correct responding appears to be smaller for high feeling-of-knowing rating conditional on correct recognition than for high confidence rating conditional on correct recall (see Table 5). The trends of the conditional probabilities across AV levels and NLM use for confidence ratings/recall are different from those for feeling-of-knowing ratings/recognition. The trends of the conditional probabilities of confidence ratings/recall are close to the conditional probabilities reported by Wearing, although they are lower. The results for recognition appear different from the results for recall.

Testing for these trends involving 4- and 3-way interactions is difficult because the data are nominally scaled. One possibility is the testing of the four two-by-two contingency tables formed by both the set of the first numbers in the parentheses in Table 5 (see _Note_)

Insert Table 5 about here

---

and the set of the second numbers in the parentheses. For instance, the first two-by-two table consists of the numbers 2, 18, 31, 212. A χ^2 can test the interaction within this contingency table and two binomial tests can test the totals of the rows and columns for main effects.

The only significant χ^2 (97.18, df=1, $p<.01$) is revealed in the contingency table of the number of correctly recognized items. All main effects, however, are significant by binomial probabilities ($p<.01$).

The reverse conditional probabilities of correct responding given high subjective certainty appear to be directly related to AV level and NLM use with one exception (see Table 6). The conditional probability of correct recall/high confidence rating for low AV and no reported NLM is 1.00. The extremely low call frequency of two should cause this conditional probability to be highly unreliable and therefore it should be ignored. With that exception, the trends seem to argue against Wearing's conclusion that this conditional probability is uninfluenced by AV and reported NLM use.

Insert Table 6 about here

Using the four contingency tables formed from the sets of either the first or the second numbers in the parentheses in Table 6 (see Note) provides further evidence for the existence of trends. None of the four contingency tables has a significant χ^2, but all the main effects of AV level and reported NLM use are significant by binomial probabilities ($p<.01$).

One problem in arguing from these conditional probabilities is that they are calculated from slices of a larger pie. They represent only some of the possible relationships to be found in a full analysis of response variable (either recall or recognition) by the subjective certainty variable (either confidence rating or feeling-of-knowing rating) by the reported NLM use variable by the AV level variable. If a larger analysis could be performed, a number of more striking trends could possibly be found. For instance, NLM use is directly and significantly related to both correct recall (χ^2= 908.98, df=2, p<.01) and correct recognition (χ^2= 244.30, df=1, p<.01). The percentages of items for which NLM's were reported are 73% of correctly recalled items and 89% of correctly recognized items. As already mentioned, reported NLM use is also directly related to AS and AV.

Discussion

To summarize the results, (a) correct recall is directly related to confidence ratings, reported NLM use, and AV-AS level. (b) Correct recognition is also directly related to feeling-of-knowing ratings, reported NLM use, and AV-AS level. (c) Reported use of an NLM is directly related to AV-AS level. (d) The conditional probabilities of high confidence ratings conditional on correct recall appear somewhat larger and somewhat different in trend than the conditional probabilities of high feeling-of-knowing ratings conditional on correct recognition. (e) Both of these sets of conditional probabilities are directly related to AV level and reported use of an NLM. (f) The reverse conditional probabilities of correct recall conditional on high confidence ratings are similar in trend to the reverse conditional probabilities of correct recognition conditional on high feeling-of-knowing ratings. (g) Both sets of reverse conditional probabilities are directly related to AV level and reported use of an NLM. (h) Both sets of the reverse conditional probabilities appear larger than both sets of the other conditional probabilities.

This study replicates Wearing's findings with two exceptions which in themselves challenge his support of Adams' closed-loop hypothesis. Wearing states that the conditional probabilities of high confidence ratings conditional on correct recall are directly related to AV level and reported use of an NLM, but that the reverse conditional probabilities of correct recall conditional on high confidence ratings are "more or less the same regardless of AV level and reported use of an NLM." According to closed-loop theory, the reason for this difference between the two types of conditional probabilities is that high confidence ratings require high trace strength for both perceptual and memory traces, but correct recall occurs sometimes when there is a weak memory trace (see Adams & Bray, 1970). By that rationale, a correct recall does not predict a high confidence rating, but a high confidence rating almost always ensures that recall will be found to be correct.

Further, as AV increases and as NLM's are reported, trace strength of both perceptual and memory traces should increase. As trace strength increases the proportion of correctly recalled items which receive a high confidence rating (high confidence rating conditional on correct recall) should also increase, but the proportion of items with a high confidence rating which are also correctly recalled (correct recall conditional on high confidence rating) should not increase because all items with high confidence ratings already have high strength perceptual and memory traces.

The conditional probabilities found in this study do not appear "more or less the same regardless of AV level or whether or not an NLM was reported." This is one exception in the replication of Wearing's findings. It is probably the result of the other exception, which pertains to the proportion of responses which are errors of commission. Wearing reports that subjects gave more errors of commission than correct responses and he reports that this occurred probably because of the 2¢ bonus he gave for each correct response. In the present study no bonus for correct responding was given and the subjects gave fewer errors of commission than correct responses.

In support of closed-loop theory, the reverse conditional probabilities of correct recall conditional on high confidence ratings are still higher than the other conditional probabilities of high confidence ratings conditional on correct recall, but this finding can also be explained by the partial recall hypothesis in the following fashion. If a response occurs easily or automatically it will certainly receive a high confidence rating and will be correctly recalled, but some correctly recalled responses do not occur automatically and will not receive a high confidence rating. If a response receives a high confidence rating because it occurred automatically or easily, it will probably also be correctly recalled, but of course not all of the items with a confidence rating of 5 are correctly recalled. If a response is correctly recalled, it may not have a high confidence rating because it may not have occurred easily or automatically. Of course, correctness and automaticity of recall are highly related, so that many correctly recalled items will have a high confidence rating. The conditional probabilities of high confidence rating conditional on correct recall are lower than the reverse conditional probabilities of correct recall conditional on high confidence ratings.

The results from recognition and feeling-of-knowing ratings have tended to duplicate the results from confidence ratings and recall with two exceptions. First, the conditional probabilities of high feeling-of-knowing rating given correct recognition (in other words, the proportions of the correctly recognized items which had a high feeling-of-knowing rating) were different in trend from the conditional probabilities of high confidence rating conditional on correct recall (in other words, the proportions of the correctly recalled items which were given a high confidence rating). These latter conditional probabilities have a direct relationship between AV level and reported use of an NLM. The former conditional probabilities show an interaction with NLM's, appearing to have different effects at the high AV level than at the low AV level. This reversal of the recognition advantage of NLM's occurs because approximately 25% of items (and probably more) are recognized by chance. In these cases the subject did not know the right answer and he knew he did not know it, but by luck he happened to choose the correct alternative from among the four alternatives. Therefore although fewer of the low AV items were correctly recognized, a greater proportion of them than of the high AV items was recognized correctly because of chance. Items correctly recognized by chance are not likely to be accompanied by an NLM.

The second exception to the duplication of recall results by recognition results occurs in the lack of interaction of AV and AS effects on recognition. This tends to support closed-loop theory because that theory would predict minimal effect for AS on recognition. This finding, too, may have resulted from the higher chance level of recognition than recall, although the reason is not clear. As shown in Figure 1, the main difference between recall and recognition seems to be that AS level has much less effect within AV levels in recognition performance than in recall.

Insert Figure 1 about here

In conclusion, in spite of this recall-recognition difference, the results reduce the support for closed-loop theory and thereby raise the partial recall hypothesis in stature. In addition, the results demonstrate again that confidence ratings and feelings-of-knowing are valid sources of information about knowledge in memory, and that the validity of such information increases directly with AV, AS and reported use of an NLM.

References

Adams, J. A. Human memory. New York: McGraw-Hill, 1967.

Adams, J. A. Response feedback and learning. Psychological Bulletin, 1968, 70, 486-504.

Adams, J. A. & Bray, N. W. A closed-loop theory of paired-associate verbal learning. Psychological Review, 1970, 77, 385-405.

Adams, J. A., Marshall, P. H. & Bray, N. W. Closed-loop theory and long-term retention. Journal of Experimental Psychology, 1971, 90, 242-250.

Adams, J. A., McIntyre, J. S. & Thorsheim, H. I. Response feedback and verbal retention. Journal of Experimental Psychology, 1969, 82, 290-296.

Anderson, R. C. Control of student mediating processes during verbal learning and instruction. Review of Educational Research, 1970, 40, 349-369.

Archer, E. J. A re-evaluation of the meaningfulness of all possible CVC trigrams. Psychological Monographs, 1960, 74, (10, Whole No. 497).

Bilodeau, E. A. & Bilodeau, I. M. Motor skills learning. Annual Review of Psychology, 1961, 12, 243-280.

Blake, M. Prediction of recognition when recall fails: Exploring the feeling-of-knowing phenomenon. Journal of Verbal Behavior and Verbal Learning, 1973, 12, 311-319.

Bower, G. A multicomponent theory of the memory trace. In K. W. Spence & J. T. Spence (Eds.), The psychology of learning and motivation: Advances in research and theory (Vol. 1). New York: Academic Press, 1967.

Brown, R. & McNeill, D. The "tip of the tongue" phenomenon. Journal of Verbal Learning and Verbal Behavior, 1966, 5, 325-337.

Clarke, F. R. Confidence ratings, second-choice responses, and confusion matrices in intelligibility tests. Journal of the Acoustical Society of America, 1960, 32, 35-46. (Also in Swets, 1964.)

Estes, W. K. Learning theory and the new "mental chemistry." Psychological Review, 1960, 67, 207-223.

Freedman, J. L. & Landauer, T. K. Retrieval of long-term memory: "Tip-of-the-tongue" phenomenon. Psychonomic Science, 1966, 4, 309-310.

Goss, A. E. & Nodine, C. F. Paired-associates learning: The role of meaningfulness, similarity, and familiarization. New York: Academic Press, 1965.

Green, D. M. & Swets, J. A. Signal detection theory and psychophysics. New York: Wiley, 1966.

Greenwald, A. G. Sensory feedback mechanisms in performance control: With special reference to the ideo-motor mechanism. Psychological Review, 1970, 77, 73-99.

Hart, J. T. Memory and the feeling-of-knowing experience. Journal of Educational Psychology, 1965, 56, 208-216.

Hart, J. T. Methodological note on feeling-of-knowing experiments. Journal of Educational Psychology, 1966, 57, 347-349.

Hart, J. T. Second-try recall, recognition and the memory-monitoring process. Journal of Educational Psychology, 1967, 58, 193-197. (a)

Hart, J. T. Memory and the memory monitoring process. Journal of Verbal Learning and Verbal Behavior, 1967, 6, 685-691. (b)

Horowitz, L. M. & Prytulak, L. S. Redintegrative memory. Psychological Review, 1969, 76, 519-534.

James, William. The principles of psychology (Vol. 1). New York: Holt, 1890.

Jersild, A. The determinant of confidence. American Journal of Psychology, 1929, 41, 640-642.

Kiess, H. O. Effects of natural language mediators on short-term memory. Journal of Experimental Psychology, 1968, 77, 7-13.

Landauer, T. K. Two states of paired-associate learning. Psychological Reports, 1962, 11, 387-389.

Montague, W. E. Elaborative strategies in verbal learning and memory. In G. H. Bower (Ed.) The psychology _of learning and motivation: Advances in research and theory_ (Vol. 6). New York: Academic Press, 1972.

Montague, W. E. & Kiess, H. O. The associability of CVC pairs. _Journal of Experimental Psychology_, 1968, _78_, (2, Pt. 2).

Montague, W. E. & Wearing, A. J. The complexity of natural language mediators and its relation to paired-associate learning. _Psychonomic Science_, 1967, _7_, 135-136.

Mowrer, O. H. _Learning theory and the symbolic processes_. New York: Wiley, 1960.

Murdock, B. B., Jr. _Human memory: Theory and data_. Potomac, Md.: Lawrence Erlbaum Associates, 1974.

Pollack, I. & Decker, L. R. Confidence ratings, message reception, and the receiver operating characteristic. _Journal of the Acoustical Society of America_, 1958, _30_, 286-292. (Also in Swets, 1964.)

Prytulak, L. S. Natural language mediation. _Cognitive Psychology_, 1972, _2_, 1-56.

Swets, J. A. (Ed.) _Signal detection and recognition by human observers: Contemporary readings_. New York: Wiley, 1964.

Tobius, S. Review of the response mode issue. _Review of Educational Research_, 1973, _43_, 193-204.

Tulving, E. & Madigan, S. A. Memory and verbal learning. Annual Review of Psychology, 1970, 21, 437-484.

Underwood, B. J. Attributes of memory. Psychological Review, 1969, 76, 559-573.

Underwood, B. J. Are we overloading memory? In A. W. Melton & E. Martin (Eds.) Coding processes in human memory. Washington: V. H. Winston, 1972.

Walker, C. B., Montague, W. E. & Wearing, A. J. Natural language associability in paired-associate learning. Journal of Experimental Psychology, 1970, 84, 264-267.

Wearing, A. J. On the Adams-Bray retrieval model. Journal of Experimental Psychology, 1971, 89, 96-101.

Wearing, A. J., Walker, C. B., and Montague, W. E. Syntactic structure and natural language mediation. Psychonomic Science, 1970, 19, 348-349.

Woodworth, R. S. & Schlosberg, H. Experimental psychology. New York: Holt, Rinehart & Winston, 1954.

Woodworth, R. S. Psychology (3rd Ed.). New York: Holt, 1934.

Table 1

Decision Matrix

Subjective Certainty	Memory outcome: Correct	Memory outcome: Incorrect	Total
High	True Positive	False Positive	Positive
Low	False Negative	True Negative	Negative
Total	Correct	Incorrect	

Table 2

Mean Recall, Recognition and NLM Scores for Associability Scale (AS) Levels within Association Value (AV) Levels and Mean Differences between Two AS Levels within each AV Level and Mean Scores for AV Levels pooled over Two AS Levels

Group	Recall	Recognition	NLM
AS Levels within AV levels[a]			
High AV			
H-AS	5.58	10.10	7.77
HM-AS	3.67	9.19	6.10
Difference[b]	1.91	.91	1.67
Low AV			
LM-AS	1.09	6.19	3.39
L-AS	.26	5.84	2.29
Difference[c]	.83	.35	1.10
AV levels only			
High	3.64	9.19	6.10
Low	.68	5.84	2.29

Note. Each AV level is composed of 24 items with a total possible score of 24. Each AS level is one-half of the AV level within which it falls, with a total possible score of 12.

[a]H-AS = high AS, HM-AS = high-medium AS, LM-AS =

Table 2, continued

low-medium AS, L-AS = low AS.

[b]Difference = high AS minus high-medium AS.

[c]Difference = low-medium AS minus low AS.

Table 3

F Values for Analysis of Variance with Repeated Measures on the Relationships between Recall, Recognition and NLM Reports with Association Value Levels (AV) and Associability Scale Levels (AS) Together in a Triple Comparison and Separately

Test	Recall	Recognition	NLM
AV and AS in a Triple Comparison			
HH - HM[a]	32.80**	6.50*	9.29**
LM - LL[b]	21.07**	1.02	8.69**
d_1 - d_2[c]	6.93*	1.39	1.34
Separately for AV and AS			
AV	116.28**	164.44**	112.09**
AS	85.64**	66.63**	63.79**

[a]HH = high AV-high AS, HM = high AV-high-medium AS.

[b]LM = low AV-low-medium AS, LL = low AV-low AS.

[c]d_1 = HH minus HM, d_2 = LM minus LL.

*$p<.05$

**$p<.01$

Table 4

Proportional Distribution of Errors of Commission and Correct Responses of Recall over the Five Categories of Confidence Ratings and of Incorrect and Correct Responses of Recognition over the Five Categories of Feeling-of-Knowing Ratings

	Subjective certainty						Row Totals
Variable	1	2	3	4	5	Sum	
Recall	Confidence ratings						
Errors of Commission	.26	.34	.16	.18	.06	1.00	238
Correct Responses	.03	.01	.06	.11	.80	1.00	330
Recognition	Feeling-of-knowing ratings						
Incorrect Responses	.22	.22	.32	.19	.05	1.00	513
Correct Responses	.10	.10	.20	.21	.39	1.00	966

Note. Response frequencies in the body of this table are expressed as a proportion of their respective row totals.

Table 5

Probability of a Correct Response (Recall or Recognition) Falling into the Highest Category (5) of Subjective Certainty (Confidence or Feeling-of-Knowing Ratings) as a Function of AV Level and the Reported Use of a Natural Language Mediator (NLM)

AV level	Use of	Conditional probability	
Of items	NLM	Recall[a]	Recognition[b]
Low	No	.40 (2/5)	.02 (5/241)
Low	Yes	.47 (18/38)	.38 (50/132)
High	No	.66 (31/47)	.26 (50/193)
High	Yes	.88 (212/240)	.65 (263/405)

Note. Second number in parentheses is total number of correct items. First number is the number of those correct items which were in Category 5.

[a]Probability of high confidence rating conditional on correct recall.

[b]Probability of high feeling-of-knowing rating conditional on correct recognition.

Table 6

Response Correctness (Recall or Recognition) as a Function of Association Value (AV) Level, Reported Use of a Natural Language Mediator (NLM), and High (Category 5) Subjective Certainty (Confidence or Feeling-of-Knowing Ratings)

AV level	Use of	Conditional probability	
Of items	NLM	Recall[a]	Recognition[b]
Low	No	1.00 (2/2)	.36 (5/14)
Low	Yes	.69 (18/26)	.82 (50/61)
High	No	.94 (31/33)	.89 (50/56)
High	Yes	.98 (212/217)	.99 (273/276)

Note. The second number in parentheses is the number of items given a rating of 5. The first number is the number of those with a rating of 5 which were correct.

[a]Probability of correct recall conditional on high confidence rating (Category 5).

[b]Probability of correct recognition conditional on high feeling-of-knowing rating (Category 5).

Figure 1

Proportion of Correct Memory Responses for Recall Compared with Recognition for Each of the Four Groups of Items: Low-Low, Low-Low-Medium, High-High-Medium and High-High