Psychological Review

VOLUME 85 NUMBER 3 MAY 1978

The Trouble with Levels: A Reexamination of Craik and Lockhart's Framework for Memory Research

Alan D. Baddeley
Medical Research Council Applied Psychology Unit
Cambridge, England

The claim is made that despite its current vogue, Craik and Lockhart's levels-of-processing approach to memory is not proving theoretically productive. Attempts to develop the levels-of-processing framework have been frustrated by (a) the absence of an independent measure of processing depth, (b) the failure to identify levels *within* the broad domains of phonemic and semantic coding, and (c) the need to appeal to such further principles as "compatibility" and "elaboration," which in turn cannot be independently measured. Recent evidence suggests that three of the assumptions underlying the levels framework may be false: (a) Rote phonemic rehearsal produces learning and does not simply serve a maintenance function; (b) "shallow" processing may lead to durable memory traces; and (c) the assumption that information is processed through a linear hierarchy of stages, each deeper than the last, is becoming increasingly implausible. It is suggested that the levels assumption be abandoned in favor of a more flexible approach that is less concerned with the search for broad general principles and more prepared to explore specific components of the memory system. This point is illustrated in terms of the concept of an articulatory loop in working memory.

By the late 1960s, there was a good deal of agreement on the need to assume two types of memory: (a) a short-term store of limited storage capacity, capable of holding information over a matter of seconds and relying heavily on phonemically coded information, and (b) a long-term store of much greater capacity and durability, relying heavily on semantic coding. It was also commonly assumed that the short-term store was the means of entering information into and retrieving information from the long-term store. There were many versions of this approach, but most could be said to resemble more or less closely the model presented by Atkinson and Shiffrin (1968). During the early 1970s, it became clear that the situation was considerably more complex than the original Atkinson and Shiffrin model suggested (Baddeley, 1976, pp. 162–165).

In 1972, Craik and Lockhart proposed a revision of the standard approach to memory that seemed to offer a simple but flexible alternative to the more traditional dichotomous view of memory. Their approach retained a dichotomy between primary, or short-term, memory and long-term memory but changed the emphasis from that of separate stores that happened to have different coding processes to one in which the coding itself was the most important variable, with trace durability being a function of the way in which the material was encoded.

The primary memory system is regarded as equivalent to consciousness or an attentional system capable of maintaining information regardless of its mode of coding, provided

Requests for reprints should be sent to Alan D. Baddeley, Medical Research Council Applied Psychology Unit, 15 Chaucer Road, Cambridge, CB2 2EF, England.

the subject continues to pay attention to it. Long-term learning, however, depends on material being recoded at a deeper level. Hence, subjects presented with written words may remember them either very shallowly only in terms of their orthographic features, giving rise to a very short-lived trace, or the subjects may process them into the phonological features represented by the printed word, giving a somewhat more durable trace, or the subjects may go on to process the word in semantic terms. The durability of the trace is assumed to be a function of this processing. Such a view differed from related approaches in the memory literature at that time (e.g., Baddeley & Patterson, 1971; Crowder & Morton, 1969) in two ways. First, the view explicitly suggests that maintaining information at a given level, whether phonological or semantic, will not increase its memorability. Second, the view implies a continuum of depth rather than a series of separate stores or domains of processing.

The levels-of-processing approach has been widely accepted and, apart from its frequent use as a post hoc explanation, has generated a good deal of research. Unfortunately, the bulk of this research has been of a replicative nature, indicating across a wide range of materials and conditions that words coded in terms of their orthographic features are poorly retained; if coded in terms of their phonological characteristics, the words are well retained over a short period of time but less well retained over long periods of time than if they are encoded semantically. This, together with the general recognition that what the subject does with the material is at least as important as the material itself, represents the main theme of a large number of experiments published in the years immediately following Craik and Lockhart's (1972) article. While it could be argued that this generalization was already well known (e.g., Baddeley, 1966a, 1966b; Hyde & Jenkins, 1969; Sachs, 1967), nevertheless, the emphasis was changed to encoding rather than storage as the primary factor. Further, a demonstration of the robustness and generality of the phonemenon has clearly been valuable.

However, the theoretical interpretation of such encoding effects has, on the whole, been somewhat unpromising. The assumption quite commonly appears to be made that levels of processing is an alternative to a dichotomous theory, which it certainly is not (see Craik & Jacoby, 1975); in addition, since many results obtained are consistent with Craik and Lockhart's view, it is often claimed that a dichotomous view of memory must be incorrect. Postman (1975) presents one of the more ingeniously argued versions of this approach. The relation between type of encoding and subsequent retention is not only compatible with a dichotomous model of memory but was indeed an important component of a number of dichotomous views of memory, including my own (Baddeley, 1972).

However, the important issue is not whether levels of processing is best classified as a unitary or dichotomous theory, but whether it is likely to advance our understanding of human memory. I would like to suggest that although the theory represents a useful rule of thumb for predicting the retention of verbal materials, it has not proved theoretically fruitful; as in the case of all useful hypotheses, it went beyond the existing data in suggesting a reconceptualization, but in general, I suggest that such an extrapolation has not proved justified.

I will begin by discussing a number of problems in applying a levels-of-processing approach, will then go on to evidence suggesting that some of its basic assumptions are false, and finally will argue that a more fruitful approach to the problems of memory is offered by the class of information-processing models devised to study such issues as working memory and reading, which aim to explore the processes underlying memory in detail rather than to formulate broad general principles.

Problems of Application and Development

Problem of Measurement

In one of the very few articles that directly questions the value of the levels-of-processing concept, Craik and Tulving (1975) concerned themselves with the problem of attempting to produce an independent measure of depth of processing. Most studies have adopted the procedure of choosing types of processing that it seems intuitively plausible to assume differ in depth. Typically, these involve judgments about the particular letters making up a word at the shallowest level, with phonological or rhyme judgments making up an intermediate level, and semantic processing of one kind or another giving the deepest level. However, if the concept is to progress beyond that of providing a rule of thumb, it is necessary to have some independent method of specifying depth of processing. This is particularly important if one wishes to test the assumption of a continuum of processing rather than a series of quite separate domains such as might be assumed in any dichotomous view of memory.

Craik and Tulving (1975) explored the possibility that processing time might be used as an index of depth. In their initial experiments, this looked quite promising, since judgments of the orthography of the words were indeed made more rapidly than judgments about rhymes, which in turn took less time than semantic judgments. However, as they point out, time per se cannot provide a general indicator of processing depth. It is possible, as they demonstrate, to produce a task that involves only relatively "shallow" processing but that involves a great deal of such processing and therefore takes a long time. In one of their studies, subjects were required to decide on the particular pattern of consonants and vowels in a word, for example, using v to represent a vowel and c a consonant, does the word *rabbit* fit the pattern cvccvc? This takes longer to do than deciding whether or not a rabbit is an animal but, nevertheless, leads to poorer retention.

It is conceivable, however, that *within* a particular type of processing, time could provide an index of processing depth. Craik and Tulving (1975) tested this by taking each type of coding separately and splitting the subject's responses into those that were made more rapidly than average and those that were made more slowly. If processing time is equivalent to processing depth, then decisions made more slowly should result in better retention. No such relation was observed. It appears then that processing time does not afford a suitable indicator of depth of processing. The same conclusion could also be drawn from a study by Carpenter (1974), who observed that sentences involving lexically marked adjectives were processed more slowly than those involving unmarked adjectives but were retained less well. Craik and Tulving came up with a range of further problems for the levels-of-processing approach, which I shall discuss later. For the present purpose, it is sufficient to point out that they failed to develop a satisfactory independent measure of processing depth. This does not, of course, prove that the concept is erroneous, but it does place severe restrictions on its usefulness.

Is There a Continuum of Processing?

One of the attractions of the levels-of-processing approach to memory is in its substitution of a continuum of processing for the simple acoustic–phonemic distinction made in the existing dichotomous models. It is intuitively attractive to think in terms of a semantic continuum, if only because a concept as broad and general as semantic coding seems incapable of doing justice to the richness of the semantic system. One might therefore expect to find a gradation of memory performance, depending on the particular processing task; given such a gradation, one might reasonably hope to use this as a tool for examining the semantic processing itself. Curiously, this aspect of levels of processing does not appear to have been systematically explored. However, in those cases in which more than one type of semantic or phonological processing has been included within the same experiment, the results have proved disappointing. For example, Craik and Tulving (1975) include two levels of semantic and two levels of phonemic processing in their Experiment 1 but fail to observe any reliable difference between the two and, subsequently, confine their attention to one level of semantic and one level of phonemic processing.

Perhaps somewhat surprisingly then, the small amount of evidence available points in the direction of domains of processing rather than a continuum. However, since the evidence is so sparse it would be unwise to draw any stronger conclusion than that the levels-of-processing approach has not yet produced any convincing evidence to suggest a continuum of levels rather than a more limited number of processing domains. That is not of course to say that all semantic processing is equivalent. It seems probable that subjects adopting a strategy of elaborating a visual image are processing information in a different way from those attempting to generate associative links (Paivio, 1971), but it is unclear how one would describe these differences in terms of a levels-of-processing approach.

Compatibility and Elaboration

Craik and Tulving (1975) made two important observations that caused them to modify the original levels-of-processing interpretation of their data. First, they observed that items that had been given a positive response during the initial encoding were better retained than those evoking a negative response. Hence, given the question "Does it rhyme with dog?", the word *log,* evoking a "yes" response, would be more likely to be recalled later than would a negative instance such as *leg*. This phenomenon, which had previously been demonstrated by Schulman (1971), is attributed by Craik and Tulving to a *compatibility* effect: Positive instances are related more closely to the question and hence create a more integrated and retrievable unit.

Craik and Tulving's second basic modification stemmed from the observation that even under conditions of slow presentation (5 sec per word) and when subjects were aware that subsequent recall of the words would be required, a phonemic coding task led to poor retention. As they point out, it seems extremely unlikely that subjects under these circumstances were unaware of the meaning of the words they were judging. Craik and Tulving suggest that simple depth of encoding is less important than degree of coding elaboration. The requirement to make a semantic decision is assumed to lead to a richer and more elaborate semantic code than is involved in simply registering the meaning of a word. Such elaboration is assumed to enhance the memorability of the item in question.

Both these additional assumptions seem intuitively reasonable; indeed, there must be few if any theories other than the original levels-of-processing view that would not predict the better retention of items that are richly encoded and well integrated with their retrieval cues than poorly elaborated items at the same processing level. The introduction of such modifications to the original levels approach, however, does seem likely to make it even harder to test, since the prediction of recall must now be based not only on depth of processing (which we do not now know how to estimate) but also on coding compatibility and degree of elaboration, which again have no obvious means of independent assessment.

It appears, then, that the levels-of-processing approach is becoming more complex. The next section surveys the growing body of evidence suggesting that even in its more complex and less testable form, the levels framework is becoming progressively harder to defend. In particular, recent studies have cast doubt on three of its underlying assumptions, namely (a) that maintenance rehearsal does not lead to learning, (b) that deeper processing leads to better retention, and (c) that information is processed through an ordered hierarchy of levels.

Is the Levels Approach Based on False Assumptions?

Rehearsal and Learning

One of the basic assumptions of the original levels approach is that maintenance rehearsal, that is, mere repetition, does not enhance retention; learning occurs only when material is processed to a deeper level. Evidence favoring this assumption, at least in the case of phonological coding, came from a study by Craik and Watkins (1973). They presented their subjects with word sequences of unknown length. Subjects were given an initial letter and were told that they would be required to report the last word in the sequence beginning with that letter. This required them to identify the word in question and maintain it until it was supplanted by a second word beginning with the same letter. Hence, given letter *B* and the sequence *crocodile, bag, radish, roar, buffalo, tree, barley, radio, sunflower, chair,* followed by the instruction "recall," the subject should report the word *barley.* Subjects almost invariably recalled the last item successfully, suggesting that they were maintaining each target item during the presentation of the irrelevant intervening words. After a number of trials of this type, subjects were unexpectedly asked to recall as many of the presented words as possible. A model assuming that probability of transfer to long-term memory is a function of how long an item is held in the short-term store (e.g., Waugh & Norman, 1965) should predict a correlation between the length of time for which the item is held and probability of final recall. Hence, in the example given, *bag,* which was held during the presentation of two intervening words, should be better retained than *buffalo,* which was replaced after only one such word. No such relation was observed by Craik and Watkins.

Further evidence for the view that simple rote rehearsal does not enhance recall comes from Tulving (1966), who required subjects to read out loud a series of words that they were subsequently required to learn. Prior reading did not enhance subsequent free recall. Other studies that appear to support Craik and Lockhart's view are those of Jacoby and Bartz (1972) and Modigliani and Seamon (1974). Jacoby and Bartz presented sequences of five words and required recall either immediately or after a 15-sec unfilled interval during which subjects presumably rehearsed. A final free-recall test showed no significant difference between the immediate recall conditions and that requiring recall after 15 sec, a result that could be taken to indicate that 15 sec of rehearsal failed to enhance retention. Unfortunately, performance in both these conditions was at such a low level that it is possible that any differences might have been masked by a floor effect.

Such an objection cannot be made to the study by Modigliani and Seamon (1974), who used a Peterson task in which subjects had to recall items after a delay during which they either counted backward, generated images relating the items, or were left free to rehearse. At the end of the experiment, final free recall of the words was requested. Recall probability increased with retention interval for the imagery condition but not for the other two conditions, a result that seems to suggest that maintenance rehearsal does not enhance learning. However, this procedure has a confounding factor, since longer intervals make it less likely that items will be recalled on the initial trial. Since there is considerable evidence to suggest that a recall may function as an additional learning trial and enhance subsequent retention (e.g., Tulving, 1967), this factor clearly needs to be controlled. A subsequent article by Modigliani (1976) did so and found, contrary to the Craik and Lockhart view, that items maintained over longer intervals were more likely to appear in a final free recall, a result that Modigliani went on to show was not attributable to an item-selection artifact.

Two studies have obtained results that are compatible with the Craik and Lockhart position when performance is measured by recall but not when a recognition measure is used (Glenberg, Smith, & Green, 1977; Woodward, Bjork, & Jongeward, 1973). Such results suggest either that maintenance rehearsal does lead to some learning that recall measures are not sensitive enough to detect, an interpretation rejected by Glenberg et al. (1977), or else that the Craik and Lockhart view does not apply to recognition memory. Either of these interpretations would seriously weaken the original Craik and Lockhart position.

A number of studies have produced results that suggest that repetition *does* enhance learning. Mechanic (1964) was one of the first people to use the now popular incidental learning paradigm to control a subject's encoding strategy. He presented his subjects with nonsense syllables and instructed them to read each syllable as many times as possible within a set time interval. Subsequent recall was shown to be a function of number of such rehearsals.

More recently, a number of studies using a more conventional free-recall procedure have also shown a clear relation between number of rehearsals and subsequent recall (e.g., Rundus, 1971). T. O. Nelson (1977) reports three experiments in which subjects are required to process items either once or twice at a phonemic level. In all cases, two repetitions led to better subsequent retention than one. Finally, two studies have shown evidence that increasing the time for which an item must be rehearsed enhances its subsequent recall. Dark and Loftus (1976) used a Brown–Peterson task in which subjects were required to retain sequences of three to five words over intervals ranging from 0–20 sec by overt spoken rehearsal. In order to avoid the confounding effect of initial recall pointed out by Modigliani (1976), they only required recall on half the presentation trials. On a final free-recall test, the crucial items that had not been tested by initial free recall showed clear effects of retention interval, with the overt phonemic rehearsal producing a positive relation between rehearsal interval and final free recall.

A similar result was obtained by Darley and Glass (1975) in a study where final free recall was incidental to the primary task of visual search. Subjects were given lists of words and were instructed to search for a specified target word on each list. The duration of retention of the target word was varied by placing the target either early or late in the list. A subsequent final free-recall test showed better retention of targets that had appeared later in the list and hence had been rehearsed for a longer interval.

How could a levels approach handle these results? Probably the simplest place to start is with T. O. Nelson's (1977) observation that two presentations are better than one. Craik and Tulving (1975) have already conceded that any word that is presented is likely to access its semantic representation in long-term memory, even though the subject is only required to make phonemic judgments. It is not implausible to assume that successive presentations might access a slightly different aspect of the complex network of semantic associations that probably represent the full meaning of any given word to a particular subject. This could be conceived as producing a slightly richer or more elaborate code for that item, which would therefore become easier to recall. From here, one could argue, for example, that Mechanic's (1964) subjects were involuntarily arousing meaningful associations to the nonsense syllables they were reading out; the greater the number of repetitions, the greater the probability of a meaningful association occurring.

The major weakness of such an interpretation lies in its circularity. Unless one has some independent means of assessing such involuntary semantic elaboration, then it is hard to regard it as anything other than a means of explaining away any embarrassing results. Indeed, on such an interpretation, it is not at all clear why pure maintenance rehearsal should ever occur because if such semantic associations occur automatically during rote rehearsal, as the interpretation would seem to suggest, they should presumably always be present, although any given method of testing might not be sufficiently sensitive to detect their influence.

Is Deep Coding Necessary for Durable Traces?

There is no doubt that for the conventional recall of lists of words, semantic coding leads to better performance than coding in terms of the orthography or sound of the constituent words. This point was well known before the Craik and Lockhart (1972) article, which took this relation and attempted to turn it into the general principle that shallow encoding leads to less durable memory traces than deep encoding. In recent years, this assumption has begun to appear less and less plausible. First of all, there is a growing body of evidence to suggest that apparently superficial aspects of a stimulus may be retained over substantial periods of time. This applies to the characteristics of the voice reading out the series of words (Craik & Kirsner, 1974), the visual characteristics of verbal stimuli (Kirsner, 1973), and the visual location of a particular word on a page of text (Rothkopf, 1971).

In a similar vein, D. L. Nelson and his coworkers have shown clear interference effects in the long-term learning of word pairs based on the particular letters comprising the stimulus words, with effects being particularly strong when two or more stimuli share the same initial letter (D. L. Nelson & Borden, 1973; D. L. Nelson & Brooks, 1973); initial letter is presumably a relatively "shallow" feature of the subjects' encoding.

Perhaps the strongest evidence for the retention of apparently superficial aspects of verbal material comes from a series of experiments by Kolers on the reading of orthographically transcribed text (Kolers, 1975a, 1975b, 1976). For example, in one study, Kolers (1976) required subjects to practice reading typographically inverted text and then retested 13–15 months later. Not only had subjects retained the skill to a substantial extent, but memory for specific pages appeared to have been retained, with pages that had been read a year previously taking less time to read than new pages. Whether or not one accepts Kolers's argument that the specific page facilitation does not have a semantic basis, it is nevertheless hard to explain the overall result in terms of deep semantic coding. A similar argument could be presented for motor skills such as tracking, which appear to show virtually no forgetting over intervals of many months (e.g., Fleishman & Parker, 1962). Since retention is much better than typically occurs for verbal materials, should one conclude that tracking involves deeper processing?

The problem of interpreting general statements about depth of processing and trace durability is further emphasized by Morris, Bransford, and Franks (1977), who point out that the question of what mode of processing is optimal depends crucially on the manner in which retention is tested. They carried out a series of experiments in which learning occurred via a task involving either semantic or rhyme processing and which was tested by requiring the subject to pick out words that were either semantically equivalent to the presented words or rhymed with them.

With a semantic test, semantic learning led to better performance; when performance was tested using the rhyming task, the opposite occurred. A similar demonstration of the potential superiority of orthographic coding over semantic is described by Bransford, Franks, Morris, and Stein (in press). Subjects were presented with a series of words, each of which was printed in lowercase except for one letter that was capitalized. Subjects were given either a semantic processing instruction, deciding whether the word fitted into a particular sentence frame, or an orthographic decision as to whether a given letter was or was not capitalized. Retention was then tested by means of a recognition test in which the foils for a particular item were either semantically different words, allowing the decision to be made purely on the basis of meaning, or else were the identical word but with a different letter capitalized. Subjects given the semantic processing instruction performed better on the semantic test, while the reverse occurred for those given the orthographic orienting task during learning.

It is clear then that simple generalizations about the relation between encoding and subsequent retention are unlikely to prove adequate unless the mode of testing is specified. In many of the experiments associated with the levels-of-processing approach, the orienting task has clearly been irrelevant to the subsequent recall procedure; for example, the instruction to decide whether a given word is printed in upper- or lowercase does not require the subject to process the word at all, one letter would be sufficient. Hence, it is hardly surprising that a subject's ability to remember words processed in this way is poor.

Despite the interaction between learning and test procedures observed by Bransford and his colleagues, they do observe an overall tendency for semantic encoding to lead to better retention. However, they attribute this to subjects' greater experience and expertise in semantic processing rather than to any fundamental superiority of semantic coding. They draw an analogy with the chess expert who is much better at perceiving and remembering chess positions than the average player (De Groot, 1966). Such an interpretation would of course be quite compatible with the later Craik and Tulving (1975) acknowledgment of the importance of elaboration. Bransford et al. (in press) go on to demonstrate, however, that a simple concept of elaboration is not adequate without a number of further assumptions concerning both the subject's prior knowledge (Stein, 1977) and the need to differentiate between the items learned (Bransford, McCarrell, Franks, & Nitsch, 1977). Taken as a whole, the work of Bransford and his associates seems to suggest that the concept of levels of processing appears to be rapidly approaching obsolescence.

Is There a Hierarchy of Levels?

A central assumption of Craik and Lockhart's framework is that information is processed by the attentional primary memory system through a hierarchy of levels beginning with physical features and moving, in the case of words, through phonemic features to a semantic level. Such a view has played a central role in theories based on the information-processing view since the late 1950s (e.g., Broadbent, 1958; Sutherland, 1968; Treisman, 1969). However, evidence is beginning to accumulate suggesting that such a hierarchical view is probably erroneous. If so, it seems likely that a levels-of-processing view based on such an assumption will prove inadequate. What then is the evidence against a simple processing hierarchy?

At present, it is suggestive rather than compelling. The power of a heterarchical system that allows a set of processors to interact rather than to function in a hierarchical series is well illustrated by Winograd's (1972) computer program. This uses three component programs—one for analyzing syntax, one for semantics, and a third for reasoning—and shows that by allowing the three to interact rather than operate in series, a more powerful means of language analysis is produced. It seems plausible to assume that a system at least as powerful operates in language comprehension by man, and it seems unlikely that it uses a series of separate hierarchically arranged stages rather than an interactive heterarchical system.

The difficulty that such a view presents for a levels-of-processing model will be illustrated using two recent experimental studies, both concerned with the analysis of reading. In the first of these, Kleiman (1975) investigates the role of articulation in reading using an articulatory suppression technique whereby the subject is required to repeat a stream of random digits at the same time as the subject is required to read material of various kinds. The requirement to articulate the digits is assumed to prevent the subvocalization of the material being read. Articulatory suppression has been shown to eliminate both phonemic similarity effects (Estes, 1973; Murray, 1968) and the word-length effect (Baddeley, Thomson, & Buchanan, 1975) and is presumably a good way of interfering with phonemic coding. A simple hierarchical model of reading might suggest that the printed word is translated into a phonemic code, which then accesses a deeper semantic code. On such a view, articulatory suppression should impair the creation of the phonemic code and, hence, should drastically interfere with reading. Kleiman found no such general interference. Although articulatory suppression slowed down judgments of whether two words rhymed, it did not interfere with semantic judgments based on word meaning, for example, "Is a game mentioned in the sentence *The whole family played Monopoly*?" Only when a judgment of meaning depended on processing several words did suppression impair performance.

Kleiman (1975) concludes from his study that articulation does play a role in adult reading but not at a preliminary shallow stage, as a simple levels-of-processing view might suggest. Rather, it functions as an additional supplement to normal reading, allowing an overloaded processing system to select for temporary storage certain items that may be of current importance. To characterize such a system simply as a form of shallow encoding would clearly fail to do justice to it. An alternative conceptualization is offered by Kleiman (1975) in terms of a working memory hypothesis. I shall return to this in the next section. The results just discussed suggest then that far from being a shallow preliminary to semantic processing, phonemic coding may be a relatively late process.

Another difficulty for the levels approach is raised by evidence that semantic coding may occur under conditions whereby the subject is unaware of the orthographic or phonemic features of a visually presented word. The best instances of this phenomenon are provided by Marcel (Marcel, in press; Marcel & Patterson, in press) in a series of experiments using a pattern masking procedure. In this technique, a tachistoscopically presented word is followed after a brief interval by a pattern mask, a jumble of letters. If the mask follows the word after a sufficiently brief interval, the subject fails to detect the occurrence of the preceding word. Under these conditions, Marcel (in press) noted that subjects would sometimes "guess" a word that was semantically related to the correct word but that was not necessarily visually similar (e.g., *yellow* for *red* and *king* for *queen*). He explored this further in an experiment in which a word was presented, followed by a mask, and the subject was given a series of two-choice questions. The questions required the subject to decide (a) whether or not a word had preceded the mask, (b) which of two words the mask resembled visually (e.g., present *hill*, test *hilt* and *cigarette*), or (c) which of two it resembled in meaning (e.g., present *hill*, test *mountain* and *machine*). The target–mask interval was initially quite long, allowing good performance on all three criteria; during the experimental session it was progressively shortened, and the point at which each type of judgment reached chance level was noted. The first judgment to reach chance was that of whether or not a word had appeared, followed shortly by the judgment of visual similarity, and finally at substantially shorter target–mask intervals by the semantic judgment. Note that this is exactly the opposite to what might be expected on a levels-of-processing view or indeed on commonsensical grounds.

However, the effect was not present for all subjects, since a few subjects (very reasonably) refused to make judgments about words they had reported not seeing. This problem was circumvented in a further experiment using the lexical decision technique. In this procedure, subjects are presented with letter strings and required to decide as quickly as possible whether the string is or is not an English word. Hence, the sequence *dake* would evoke a "no" response, whereas *date* should be categorized "yes." Meyer, Schvaneveldt, and Ruddy (1975) have shown that "yes" responses are faster when preceded by an associated word (e.g., *butter* preceded by *bread*). Marcel (in press) adapted this procedure by masking the first word of a pair to ensure that the subject was quite unaware of its existence. Under these conditions, associative priming occurred and was just as great as when the first word was not masked and the subject was aware of the word. Marcel concludes that information may be analyzed to a semantic level without requiring conscious processing. A similar conclusion is reached by Lewis (1970, 1972) using a dichotic listening task and by Keele (1972) using a variation on the Stroop test.

Further evidence comes from brain-damaged dyslexic patients, who give evidence of semantically processing words they are able to pronounce but are unable to read aloud (Marshall & Newcombe, 1973; Patterson & Marcel, 1977; Shallice & Warrington, 1975). Such evidence is clearly inconsistent with Craik and Lockhart's assumption of a primary memory system associated with consciousness that is responsible for hierarchically transforming orthographic information through a phonological code into a semantic code. The evidence suggests a much richer, more flexible, and more complex system.

Alternative Approaches: General Principles or Specific Processes?

Within current approaches to cognitive psychology, there are substantial differences in strategy. One of the more interesting features of Craik and Lockhart's original framework was the way in which it combined elements of two of these. It sprang from the information-processing short-term memory tradition, with its strategy of formulating specific mechanisms and processes, but it blended this with the more traditional verbal learning approach of seeking general overall principles. The roots of the first strategy are to be found in the work of Broadbent (1958) and Atkinson and Shiffrin (1968) and was reflected largely in the primary memory component of the levels approach. The search for broad overall principles is much more in line with the traditional functionalist approach to memory stemming from such figures as Carr, McGeoch, and Irion (see, e.g., Irion, 1959). Subsequent development of the levels approach has been primarily in this latter tradition, with an increasing tendency to postulate such broad general concepts as elaboration and compatibility, which in turn tend to be based on even more general concepts, such as "distinctiveness" (Jacoby & Craik, in press).

I would like to argue in this final section that such a trend is not fruitful and to press for a return to the detailed analysis of specific subcomponents of human memory. I wish to illustrate this point using the working memory approach to memory (Baddeley & Hitch, 1974). While it would not be accurate to regard this formulation as a rival to levels of processing, since it has largely concerned itself with rather different problems, there is nevertheless sufficient overlap to allow some comparison of the two approaches. For this purpose, I shall choose the most obvious point of overlap: the treatment of the role of phonemic coding in memory.

Craik and Lockhart treat phonemic coding simply as an intermediate level of encoding verbal material, a level that produces better retention than coding in terms of the orthographic features of printed words but leads to poor long-term retention when contrasted with semantic coding. As we have seen, such a concept provides a useful rule of thumb for predicting the outcome of certain types of experiments, but it does not appear to have led to any deeper understanding of the underlying processes.

The working memory approach stemmed from a series of experiments (Baddeley & Hitch, 1974) that aimed to explore the role of short-term memory in a range of information-processing tasks; it tried to answer the question, What is short-term memory for? The initial experiments used a procedure whereby subjects were required to remember sequences of up to six digits while performing other tasks such as verbal reasoning, comprehension, and verbal learning. It was assumed that remembering a sequence of digits would occupy part or all of the short-term storage system; if such a system was necessary for reasoning, comprehension, and learning, performance on these tasks should be impaired by the concurrent digit load. Baddeley and Hitch observed a consistent pattern across the three tasks, with performance being virtually unaffected by loads of up to three items, but with six items producing clear evidence of a decrement in performance. Other experiments explored the role of phonemic coding in reasoning and comprehension and found a small but consistent effect. In order to explain these and other results, we suggested the division of the short-term memory system into at least two components: (a) a central executive, which is responsible for control processes and is possibly also associated with selective attention, and (b) a secondary articulatory loop, which acts as a slave system enabling any material that can be verbalized to be stored in a way that makes minimal demands on the central executive. The system appears to be time based and can hold up to 2 sec of spoken material (Baddeley et al., 1975). The use of the loop can be prevented if material is presented visually, and subjects are required to make irrelevant articulations during presentation, for example, by requiring them to repeat the word *the* or to count repeatedly from one to six (Levy, 1971). The loop appears to utilize a verbal code and is assumed to be responsible for the phonemic similarity effect (Baddeley, 1966b; Conrad, 1964) and the word-length effect (Baddeley et al., 1975), since both effects disappear when visual presentation is used and articulatory suppression is required (Baddeley et al., 1975; Estes, 1973; Murray, 1968).

There are of course clear similarities between the articulatory loop and earlier interpretations of short-term memory, which have also stressed rehearsal as an important component (e.g., Atkinson & Shiffrin, 1968; Broadbent, 1958). The working memory approach, however, differs from earlier views in a number of respects. First it does not regard the articulatory loop as synonymous with short-term storage; the loop is simply a useful peripheral component that can be cut out by articulatory suppression without having a devastating effect on the performance of the rest of the system. A second difference from most earlier interpretations resides in the treatment of recency in free recall, which most earlier approaches attribute to a short-term memory system. Since neither concurrent digit load (Baddeley & Hitch, 1974) nor articulatory suppression (Richardson & Baddeley, 1975) appears to influence the recency effect, it is suggested that recency results from the operation of a particular retrieval strategy rather than the output of a short-term storage system. As Glanzer, Koppenaal, and Nelson (1972) have shown, the recency effect is not influenced by phonemic similarity. Such a result is embarrassing for a number of more traditional views of short-term memory but is quite consistent with the working memory interpretation that assigns phonemic similarity effects to the articulatory loop and recency effects to a retrieval strategy (Baddeley & Hitch, 1977).

A number of results suggest that the articulatory loop may play an important role in reading. Conrad (1972) has pointed out that children start to show an acoustic similarity effect in remembering sequences of pictures having phonemically confusable names at about the time they are first mastering reading. Bakker (1972) has shown that backward readers have difficulty performing tasks that involve the reproduction of order information, tasks that might well be expected to rely normally on the utilization of the articulatory loop. As might be expected with this argument, dyslexic children, who despite normal intelligence have difficulty in learning to read, show impaired forward digit span and even more grossly impaired reversed span (Miles & Wheeler, 1974; Naidoo, 1970). Liberman et al. (1977) present further evidence for such a view in a paper indicating that backward readers tend not to show the normal phonemic similarity effect in remembering visually presented consonant sequences, implying that backward readers do not take advantage of the articulatory loop.

All this evidence points to the importance of the articulatory loop in learning to read. This may be because the articulatory loop allows the reader to store the phonemic representations of each letter read, until sufficient representations have been accumulated to blend into a syllable, which then can itself be stored. It is unlikely that difficulty in utilizing the articulatory loop represents the only cause of dyslexia, but it does appear to provide one potentially important factor.

The role of the articulatory loop in the performance of a skilled reader is probably rather different, allowing the highlighting of certain important aspects of the material read. These can then be stored phonemically and readily retrieved to aid in the comprehension of complex or difficult material. As such, the articulatory loop probably facilitates reading under certain conditions but is almost certainly not essential (Baddeley, in press; Marcel & Patterson, in press).

A major embarrassment for the "modal model" of memory stems from patients of the type studied by Shallice and Warrington (1970), who show grossly impaired digit span but unimpaired long-term learning ability. Such patients show none of the dramatic impairments in performance that might be expected to result from a grossly disrupted short-term store in a model such as that of Atkinson and Shiffrin (1968). However, these results can be explained within the present framework very simply by assuming that such patients have a grossly impaired articulatory loop together with an intact central executive system. An articulatory impairment is consistent with the history of aphasia that is invariably associated with such patients. As would be expected from this interpretation, aphasic patients with impaired articulation typically show grossly impaired memory span even when they are allowed to respond by pointing to the relevant items, which is a mode of responding that places no direct demands on articulation (Albert, 1976). Such a deficit is particularly striking in view of the general robustness of the digit span across a wide range of neurological deficits. This interpretation can be further tested using the word-length effect; if this type of patient is indeed deficient in use of the articulatory loop, then such patients should show no evidence of a word-length effect.

It is perhaps worth pointing out that the articulatory loop is unlikely to be the only source of phonemic coding effects in memory. Clearly, the articulatory loop must itself be fed by some more long-term representation of phonemic information. The articulatory loop represents the way in which such information may be refreshed, circulated, and kept available; but our long-term knowledge of the language is presumably stored elsewhere, since as Kleiman (1975) demonstrated, subjects are quite capable of making phonemic judgments of visually presented words while performing a shadowing task that serves as an articulatory suppressor. It is possible that the motor programs associated with particular words are stored within a lexicon or logogen system (Morton, 1970), which is linked to, but is separate from, the semantic system.

I would therefore like to claim on the basis of the evidence discussed that the concept of an articulatory loop provides not only a simple interpretation of the considerable body of evidence suggesting an association between short-term memory and speech coding, but that the articulatory loop is also capable of handling the apparent exceptions to this relation while at the same time suggesting an interpretation of a wide range of phenomena that would not normally be associated with short-term memory. There is little to suggest that the approach to phonemic coding offered by the levels approach or by any of the general principles approaches that are replacing it will prove similarly productive.

The Trouble with Levels: Conclusion

To recapitulate, the concept of levels appears to have run into a number of problems. The adequate development of the concept requires an independent means of specifying level of processing, but so far, all attempts to measure processing depth appear to have been unsuccessful. Initial hopes that a continuum of levels might be revealed, throwing light in turn on more general aspects of information processing, have proved unfounded. Furthermore, it has proved necessary to introduce other concepts such as compatibility and elaboration, which in turn are poorly defined and whose interaction with depth of processing allows the concept to explain virtually any result in a post hoc manner while having little or no predictive value.

In recent years, there has been an increasing tendency to question the three fundamental assumptions underlying the levels approach. The validity of the concept of maintenance rehearsal has been placed in doubt by a number of recent studies. The suggestion that deeper coding produces more durable memory traces has also come in for criticism; it has been suggested that such a generalization is meaningless, unless the method by which memory is tested is specified. Finally, the concept of a sequential hierarchy of processing stages is becoming increasingly implausible. In the absence of such a hierarchical system, it is doubtful that the concept of depth is either meaningful or useful.

I would suggest that the most fruitful way to extend our understanding of human memory is not to search for broader generalizations and principles but is rather to develop ways of separating out and analyzing more deeply the complex underlying processes.

References

Albert, M. L. Short-term memory and aphasia. *Brain and Language*, 1976, *3*, 28–33.

Atkinson, R. C., & Shiffrin, R. M. Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 2). New York: Academic Press, 1968.

Baddeley, A. D. The influence of acoustic and semantic similarity on long-term memory for word sequences. *Quarterly Journal of Experimental Psychology*, 1966, *18*, 302–309. (a)

Baddeley, A. D. Short-term memory for word sequences as a function of acoustic, semantic and formal similarity. *Quarterly Journal of Experimental Psychology*, 1966, *18*, 362–365. (b)

Baddeley, A. D. Retrieval rules and semantic coding in short-term memory. *Psychological Bulletin*, 1972, *78*, 379–385.

Baddeley, A. D. *The psychology of memory*. New York: Basic Books, 1976.

Baddeley, A. D. Working memory and reading. In P. A. Kolers, M. E. Wrolstad, & H. Bouma (Eds.), *The processing of visible language,* in press.

Baddeley, A. D., & Hitch, G. Working memory. In G. A. Bower (Ed.), *The psychology of learning and motivation* (Vol. 8). New York: Academic Press, 1974.

Baddeley, A. D., & Hitch, G. Recency re-examined. In S. Dornic (Ed.), *Attention and performance* (Vol. 6). Hillsdale, N.J.: Erlbaum, 1977.

Baddeley, A. D., & Patterson, K. R. The relationship between long-term and short-term memory. *British Medical Bulletin,* 1971, *27,* 237–242.

Baddeley, A. D., Thomson, N., & Buchanan, M. Word length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behavior,* 1975, *14,* 575–589.

Bakker, D. J. *Temporal order in disturbed reading.* Rotterdam, The Netherlands: Rotterdam University Press, 1972.

Bransford, J. D., Franks, J. J., Morris, C. D., & Stein, B. S. Some constraints on learning and memory research. In L. S. Cermak & F. I. M. Craik (Eds.), *Rockport Symposium on Levels of Processing.* Hillsdale, N.J.: Erlbaum, in press.

Bransford, J. D., McCarrell, N. S., Franks, J. J., & Nitsch, K. E. Toward unexplaining memory. In R. E. Shaw & J. D. Bransford (Eds.), *Perceiving, acting and knowing: Toward an ecological psychology.* Hillsdale, N.J.: Erlbaum, 1977.

Broadbent, D. E. *Perception and communication.* New York: Pergamon Press, 1958.

Carpenter, P. A. On the comprehension, storage and retrieval of comparative sentences. *Journal of Verbal Learning and Verbal Behavior,* 1974, *13,* 401–411.

Conrad, R. Acoustic confusion in immediate memory. *British Journal of Psychology,* 1964, *55,* 75–84.

Conrad, R. The developmental role of vocalizing in short-term memory. *Journal of Verbal Learning and Verbal Behavior,* 1972, *11,* 521–533.

Craik, F. I. M., & Jacoby, L. L. A process view of short-term retention. In F. Restle (Ed.), *Cognitive theory* (Vol. 1). Hillsdale, N.J.: Erlbaum, 1975.

Craik, F. I. M., & Kirsner, K. The effect of a speaker's voice on word recognition. *Quarterly Journal of Experimental Psychology,* 1974, *26,* 274–284.

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

Craik, F. I. M., & Tulving, E. Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology,* 1975, *104,* 268–294.

Craik, F. I. M., & Watkins, M. J. The role of rehearsal in short-term memory. *Journal of Verbal Learning and Verbal Behavior,* 1973, *12,* 599–607.

Crowder, R. G., & Morton, J. Precategorical acoustic storage (PAS). *Perception & Psychophysics,* 1969, *5,* 365–373.

Dark, V. J., & Loftus, G. R. The role of rehearsal in long-term memory performance. *Journal of Verbal Learning and Verbal Behavior,* 1976, *15,* 479–490.

Darley, C. F., & Glass, A. L. Effects of rehearsal and serial list position on recall. *Journal of Experimental Psychology: Human Learning and Memory,* 1975, *104,* 453–458.

De Groot, A. Perception and memory versus thought: Some old ideas and recent findings. In B. Kleinmuntz (Ed.), *Problem solving.* New York: Wiley, 1966.

Estes, W. K. Phonemic coding and rehearsal in short-term memory. *Journal of Verbal Learning and Verbal Behavior,* 1973, *12,* 360–372.

Fleishman, E. A., & Parker, J. F., Jr. Factors in the retention and relearning of perceptual-motor skill. *Journal of Experimental Psychology,* 1962, *64,* 215–226.

Glanzer, M., Koppenaal, L., & Nelson, R. Effects of relations between words on short-term storage and long-term storage. *Journal of Verbal Learning and Verbal Behavior,* 1972, *11,* 403–416.

Glenberg, A., Smith, S. M., & Green, C. Type I rehearsal: Maintenance and more. *Journal of Verbal Learning and Verbal Behavior,* 1977, *16,* 339–352.

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

Irion, A. L. Rote learning. In S. Koch (Ed.), *Psychology: A study of a science.* New York: McGraw-Hill, 1959.

Jacoby, L. L., & Bartz, W. H. Rehearsal and transfer to LTM. *Journal of Verbal Learning and Verbal Behavior,* 1972, *11,* 561–565.

Jacoby, L. L., & Craik, F. I. M. Effects of elaboration of processing at encoding and retrieval: Trace distinctiveness and recovery of initial context. In L. S. Cermak & F. I. M. Craik (Eds.), *Rockport Symposium on Levels of Processing.* Hillsdale, N.J.: Erlbaum, in press.

Keele, S. W. Attention demands of memory retrieval. *Journal of Experimental Psychology,* 1972, *93,* 245–248.

Kirsner, K. An analysis of the visual component in recognition memory for verbal stimuli. *Memory & Cognition,* 1973, *1,* 449–453.

Kleiman, G. M. Speech recoding in reading. *Journal of Verbal Learning and Verbal Behavior,* 1975, *14,* 323–339.

Kolers, P. A. Memorial consequences of automatized encoding. *Journal of Experimental Psychology: Human Learning and Memory,* 1975, *1,* 687–701. (a)

Kolers, P. A. Specificity of operations in sentence recognition. *Cognitive Psychology,* 1975, *7,* 289–306. (b)

Kolers, P. A. Reading a year later. *Journal of Experimental Psychology: Human Learning and Memory,* 1976, *2,* 554–565.

Levy, B. A. The role of articulation in auditory and visual short-term memory. *Journal of Verbal Learning and Verbal Behavior,* 1971, *10,* 123–132.

Lewis, J. L. Semantic processing of unattended messages using dichotic listening. *Journal of Experimental Psychology*, 1970, *85*, 225–228.

Lewis, J. L. Semantic processing with bisensory stimulation. *Journal of Experimental Psychology*, 1972, *96*, 455–457.

Liberman, I. Y., Shankweiler, D., Liberman, A. M., Fowler, C., & Fischer, F. W. Phonetic segmentation and recoding in the beginning reader. In A. S. Reber & D. Scarborough (Eds.), *Towards a psychology of reading*. Hillsdale, N.J.: Erlbaum, 1977.

Marcel, A. J. Unconscious reading: Experiments on people who do not know that they are reading. *Visible Language*, in press.

Marcel, A. J., & Patterson, K. E. Word recognition and production: Reciprocity in clinical and normal studies. In J. Requin (Ed.), *Attention and performance* (Vol. 7). Hillsdale, N.J.: Erlbaum, in press.

Marshall, J. C., & Newcombe, F. Patterns of paralexia: A psycholinguistic approach. *Journal of Psycholinguistic Research*, 1973, *2*, 175–199.

Mechanic, A. The responses involved in the rote learning of verbal materials. *Journal of Verbal Learning and Verbal Behavior*, 1964, *3*, 30–36.

Meyer, D. E., Schvaneveldt, R. W., & Ruddy, M. G. Loci of contextual effects on visual word-recognition. In P. M. A. Rabbitt & S. Dornic (Eds.), *Attention and performance* (Vol. 5). London: Academic Press, 1975.

Miles, T. R., & Wheeler, T. J. Towards a new theory of dyslexia. *Dyslexia Review*, 1974, *11*, 9–11.

Modigliani, V. Effects of delaying an initial recall on a later recall. *Journal of Experimental Psychology: Human Learning and Memory*, 1976, *2*, 609–622.

Modigliani, V., & Seamon, J. G. Transfer of information from short- to long-term memory. *Journal of Experimental Psychology*, 1974, *102*, 768–772.

Morris, C. D., Bransford, J. D., & Franks, J. J. Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior*, 1977, *16*, 519–533.

Morton, J. A functional model for memory. In D. A. Norman (Ed.), *Models of human memory*. New York: Academic Press, 1970.

Murray, D. J. Articulation and acoustic confusability in short-term memory. *Journal of Experimental Psychology*, 1968, *78*, 679–684.

Naidoo, S. The assessment of dyslexic children. In A. W. Franklin & S. Naidoo (Eds.), *Assessment and teaching of dyslexic children*. London: Invalid Children's Aid Association, 1970.

Nelson, D. L., & Borden, R. C. Interference produced by phonetic similarities: Stimulus recognition, associative retrieval, or both? *Journal of Experimental Psychology*, 1973, *97*, 167–169.

Nelson, D. L., & Brooks, D. H. Independence of phonetic and imageable features. *Journal of Experimental Psychology*, 1973, *97*, 1–7.

Nelson, T. O. Repetition and depth of processing. *Journal of Verbal Learning and Verbal Behavior*, 1977, *16*, 151–172.

Paivio, A. *Imagery and verbal processes*. New York: Holt, Rinehart & Winston, 1971.

Patterson, K. E., & Marcel, A. J. Aphasia, dyslexia and the phonological coding of written words. *Quarterly Journal of Experimental Psychology*, 1977, *29*, 307–318.

Postman, L. Verbal learning and memory. *Annual Review of Psychology*, 1975, *26*, 291–335.

Richardson, J. T. E., & Baddeley, A. D. The effect of articulatory suppression in free recall. *Journal of Verbal Learning and Verbal Behavior*, 1975, *14*, 623–629.

Rothkopf, E. Z. Incidental memory for location of information in text. *Journal of Verbal Learning and Verbal Behavior*, 1971, *10*, 608–613.

Rundus, D. Analysis of rehearsal processes in free recall. *Journal of Experimental Psychology*, 1971, *89*, 63–77.

Sachs, J. S. Recognition memory for syntactic and semantic aspects of connected discourse. *Perception & Psychophysics*, 1967, *2*, 437–442.

Schulman, A. I. Recognition memory for targets from a scanned list. *British Journal of Psychology*, 1971, *62*, 335–346.

Shallice, T., & Warrington, E. K. Independent functioning of verbal memory stores: A neuropsychological study. *Quarterly Journal of Experimental Psychology*, 1970, *22*, 261–273.

Shallice, T., & Warrington, E. K. Word recognition in a phonemic dyslexic patient. *Quarterly Journal of Experimental Psychology*, 1975, *27*, 187–200.

Stein, B. S. The effects of cue-target uniqueness on cued-recall performance. *Memory & Cognition*, 1977, *5*, 319–322.

Sutherland, N. S. Outlines of a theory of visual pattern recognition in animals and man. *Proceedings of the Royal Society (B)*, 1968, *171*, 297–317.

Treisman, A. M. Strategies and models of selective attention. *Psychological Review*, 1969, *76*, 282–299.

Tulving, E. Subjective organization and effects of repetition in multitrial free-recall learning. *Journal of Verbal Learning and Verbal Behavior*, 1966, *5*, 193–197.

Tulving, E. The effects of presentation and recall of materials in free-recall learning. *Journal of Verbal Learning and Verbal Behavior*, 1967, *6*, 175–184.

Waugh, N. C., & Norman, D. A. Primary memory. *Psychological Review*, 1965, *72*, 89–104.

Winograd, T. *Understanding natural language*. Edinburgh, Scotland: Edinburgh University Press, 1972.

Woodward, A. E., Bjork, R. A., & Jongeward, R. H., Jr. Recall and recognition as a function of primary rehearsal. *Journal of Verbal Learning and Verbal Behavior*, 1973, *12*, 608–617.

Received January 11, 1977
Revision received January 5, 1978 ■