Canadian Journal of Psychology, 1990, 44(1), 87-112

Canadian Journal of Psychology Outstanding Contributions Series

Levels of Processing: A Retrospective Commentary on a Framework for Memory Research

Robert S. Lockhart and Fergus I.M. Craik
University of Toronto

ABSTRACT The influence on memory research of levels of processing (Craik & Lockhart, 1972) is reviewed, and a number of conceptual and empirical criticisms are evaluated. Research since 1972 has enabled the original formulation of depth of processing to be refined in various ways, and the concepts of elaboration and distinctiveness of encoding are discussed as examples of this refinement. It is concluded that, despite change and development, many of the original ideas of levels of processing have survived and that as a research framework it has been substantially successful in encouraging the building of a data base that can serve as a foundation for future theory construction.

RÉSUMÉ L'influence des niveaux de traitement sur la recherche touchant la mémoire (Craik & Lockhart, 1972) est passée en revue et un certain nombre de critiques conceptuelles et empiriques sont évaluées. La recherche depuis 1972 a permis de raffiner de maintes façons la formulation originale de la profondeur de traitement et les concepts de l'élaboration et de clarté de codage sont discutés comme exemple de raffinement. L'on arrive à la conclusion que malgré le changement et le développement, nombreuses sont les idées originales quant aux niveaux de traitement qui ont survécu et que, en tant que cadre de recherche, ces idées ont eu un certain succès à encourager l'acquisition de données qui peuvent servir de fondations pour la construction de futures théories.

The *levels of processing* framework proposed by Craik and Lockhart (1972) presented a general approach to memory research that has been widely influential and the subject of intensive critical scrutiny. The purpose of the present article is to offer some retrospective observations on the ideas associated with levels of processing. Why was the article so influential (White, 1983), and has that influence been retained in current theoretical notions, or was it just a transient blip on the path to the real truth about memory? Despite its success, the article gave rise to many misconceptions about what we were trying to say; for example, it is commonly reported that the Craik and Lockhart article was the one that advocated eliminating the

*Preparation of this article was facilitated by grants to both authors from the Natural Sciences and Engineering Research Council of Canada. The authors are grateful to Janine Jennings for helpful library research. Requests for reprints should be addressed to Robert S. Lockhart, Department of Psychology, University of Toronto, Toronto, Ontario, Canada M5S 1A1.

distinction between short-term and long-term memory. That is simply not the case; in fact, we argued for the retention of the distinction, albeit in a somewhat different form (p. 676). Thus, a second purpose of the present paper is to clarify the arguments surrounding such misconceptions.

In such a retrospective analysis it is important to place the ideas expressed in the original paper against the background of the theoretical views prevailing at the time. The general conceptualization of memory that we sought to displace was the idea (a) that memory could be understood in terms of elements ("items") held in structural entities called *memory stores*, (b) that the fate of an item so stored was determined by the properties or parameters of this store, and (c) that a theory of memory consisted of refining our understanding of the number and properties of these stores. We sought to replace this structuralist style of theory with one that was more procedurally oriented.

Our principal objection to the concept of stores was that their major properties — capacity, coding characteristics, and forgetting rates — were quite variable from one experimental paradigm to the next. Some theorists appeared to be dealing with this situation by the simple expedient of postulating new stores for each new set of experimental findings, but such reactions did not seem profitable to us. Instead, we advocated scrapping the whole notion of memory stores and suggested, rather, that theorists should look directly at the relations between different encoding operations and subsequent memory performance. What in essence did we propose?

We endorsed the existing views of some theorists in the areas of attention and perception (e.g., Sutherland, 1968; Treisman, 1964) that the cognitive system is organized hierarchically and that incoming stimuli are processed to different levels of analysis, with the products of early (or shallow) sensory analyses serving as the input to later (or deeper) semantic analyses. Treisman, in particular, had argued that early sensory analyses are carried out in a relatively automatic fashion, whereas deeper analyses depend on a combination of signal strength (d' factors) and contextual probabilities, recent use, and importance (β factors) for their effective execution. In general, deeper analyses require more attentional resources *unless* the stimuli are expected or are very common, like the person's own name. We suggested that the memory trace could be thought of simply as the record of those analyses that had been carried out primarily for the purposes of perception and comprehension and that deeper, more semantic, analyses yielded records that were more durable.

Two further ideas suggested in the Craik and Lockhart (1972) article were first that rehearsal could usefully be broken down into two main types: Type I processing (or maintenance rehearsal) maintained processing at the same level of analysis, whereas Type II processing (or elaborative rehearsal) involved deeper or more extensive processing of the stimulus. If memory performance is a function of the deepest level of analysis achieved, only the second type of rehearsal should lead to an improvement in memory. The second idea was that the concept of primary memory should be retained, but it was seen as continued processing activity, not as a separate mechanism or structure.

THE FOUNDATION OF LEVELS OF PROCESSING

Before proceeding with a discussion of criticisms and later developments of levels of processing, it may be helpful to outline the underlying principles on which it is based. Underpinning our entire argument was the claim that the memory trace should be understood, not as the result of a specialized memory-encoding process, but rather as a by-product or record of normal cognitive processes such as comprehension, categorization, or discrimination. Successful remembering is a function of many variables, most of which have been the subject of intensive experimental study. In the late 1960s the variables receiving greatest attention were such things as presentation rates, scaled properties of stimuli, serial position, or the form of the memory test. The common sense starting point of levels of processing was the observation that when all these traditional and much-studied determinants of memory were held constant, major effects on remembering could be obtained merely by influencing the form of perceptual or conceptual analysis that the subject performed on the material to be remembered, and that, futhermore, such variation in conceptual analysis is precisely what characterizes everyday goal-directed cognition.

There are many cognitive operations and functions, but, we argued, among these various processes there is nothing that corresponds to committing to memory. There is no distinct process of memorizing that can take its place alongside other cognitive operations. We argued that the traditional intentional instructions that exhort subjects to "try to remember" amount to the experimenter's request that subjects engage in whatever processing they think will best lead to successful remembering. Depending on the nature of the materials and their meta-memory skills, subjects may (among other possibilities) repeat the items to themselves verbatim, extract gist, attempt to organize the material into coherent structures, or form visual images. Such various forms of processing will have differential consequences for any subsequent demand to remember the material so processed, but no one of them enjoys any special status as the basic process of committing to memory. We drew a number of immediate implications from this position.

Preference for Incidental Orienting Tasks: We argued that intentional instructions represent a loss of experimental control over processing operations, confounding as they do the subject's choice of processing strategy on the one hand, with the differential consequences (for subsequent tests of memory) of that particular form of processing on the other. Orienting task instructions provide better, if not perfect, experimental control over input processing. Such tasks provide an independent variable that can be operationally defined, described in functional processing terms, and, through monitoring subjects' responses during the orienting task, can provide some check that the designated processing is being performed. Another virtue of incidental instructions lies in their capacity to model the cognitive operations that occur in everyday cognition (Lockhart, 1978). Thus an ecologically valid approach to the study of memory does not demand the abandonment of laboratory experimentation. Rather, it imposes the requirement that laboratory paradigms capture and preserve those features of remembering that are important to everyday adaptive cognitive functioning.

Memory as Retrieval: A second implication of the view that the memory trace is the by-product of perceptual/conceptual analysis, rather than the consequence of a special memory-encoding process, is that the only cognitive operation that can legitimately be referred to as "remembering" is that of retrieval. Hence, rather than downplaying retrieval as some have claimed, levels of processing places it firmly at the centre of memory research; it is retrieval that is the memory area's focal explanandum.

Memory Structures and Systems: The third implication is that memory structures and systems should not be defined in terms of their retention characteristics. The prototypical memory structure (store) so defined is short-term memory. But if the persistence or durability of a memory trace is determined by the form of the initial processing (other conditions being held constant), then retention characteristics may vary over a continuous range; hence discrete, temporally defined, memory systems become unnecessary constructs. The subject of short-term memory is discussed in greater detail below in relation to working memory, which we consider to be a theoretically very different construct.

Memory and General Cognition: The fourth implication is that an adequate theory of memory can be constructed only within the context of a general theory of cognition. If the major determinants of remembering are processes serving the goals of cognition in general, then memory cannot be treated as a self-contained cognitive module. A central requirement of any theory of memory would be to formulate the principles underlying the relationship between remembering and the range of processing that constitutes the entire scope of cognition. Only then would it be possible to explain in a systematic way the differential effects on remembering of various forms of processing. We attempted to capture these principles in the concept of depth of processing; what we meant by this term and how our understanding of it has developed since 1972 is discussed more fully below.

Critical Assessments of Levels of Processing

The levels of processing framework attracted attention for a number of reasons; but, to a large extent, its appeal reflected the extent to which researchers were dissatisfied with the structural limitations of memory stores and were therefore receptive to an alternative conceptualization in terms of cognitive activities and processes. The ideas also very properly attracted criticism (Baddeley, 1978; Eysenck, 1978; T.O. Nelson, 1977). The criticisms fall into two broad groups; the first is concerned with conceptual and methodological issues or with the appropriate ways in which theories should be constructed; the second is concerned more with empirical issues — experimental support (or the lack of it) for hypotheses suggested by the levels of processing framework.

In the first group of criticisms, the danger of circular reasoning was pointed out — the tendency to define depth in terms of the memory outcome. That is, Craik and Lockhart postulated that deeper processing was associated with long-lasting memory traces; but, in the absence of an independent index of depth, there is a tendency to conclude that well-remembered events must therefore have been deeply processed. The second major criticism is incorporated in the first — the lack of an independent index. Third, Baddeley (1978) vigorously questioned the value of seeking to formulate general functional principles as a strategy likely to lead to a fuller understanding of what memory is and how it works; he argued, instead, for specifying mechanisms that can be explored experimentally and ultimately identified with brain structures and processes. Baddeley also questioned the reasonableness of assuming one linear sequence of levels of processing and pointed to the lack of evidence for further levels within the semantic domain. The most general criticism of this type was that the ideas were not falsifiable by experiment and were therefore not properly scientific (T.O. Nelson, 1977). Another general criticism was that our ideas were vague; so vague, indeed, that this same critic suggested (quoting Wittgenstein as support) that we should have remained silent. Clarity, of course, is a noble goal; but in a developing science, it is rarely achieved through silence or by appeals to authority.

In the group of empirical criticisms, both T.O. Nelson (1977) and Baddeley (1978) questioned the evidence for two types of rehearsal; in particular, whether maintenance processing had no effect in strengthening memory for the rehearsed material. Baddeley also pointed out that whereas the Craik and Lockhart position was that sensory information should be lost rapidly, there was growing evidence for long-lasting sensory traces under certain circumstances. Finally, evidence from neuropsychological studies seems often at odds with the levels of processing ideas. For example, amnesic patients can comprehend conversations and other everyday happenings perfectly well, that is, they can clearly process deeply, yet their memory for such information is demonstrably poor or nonexistent.

We do not pretend to have answers to all of these criticisms. In the almost 20 years since the original article was written, we have conceded some points and changed our own views as the relevant evidence has accrued. We have also commented previously on at least some of the arguments cited (Lockhart & Craik, 1978; Lockhart, Craik, & Jacoby, 1976). However, in this retrospective assessment we felt that it might be of interest to provide brief commentaries on some of these critical points.

CONCEPTUAL AND METHODOLOGICAL ISSUES

Circularity and the Need for an Independent Index of Depth

Perhaps the most frequent and persistent criticisms of levels of processing has been the lack of an independent index of depth. The usual claim is that without an independent index the logic is circular, in that depth must be defined in terms of its alleged effects. It has been claimed that such circularity makes levels of processing unfalsifiable and thus scientifically meaningless (T.O. Nelson, 1977).

This criticism supposes that an ideal state of affairs would be one in which the concept of depth could be operationally defined so that hypotheses relating depth to memory performance could be tested and potentially falsified. We discussed this matter some time ago (Lockhart & Craik, 1978), but it is worth updating the discussion since the criticism is still alive. It raises a number of issues that are important, not only in the narrow context of levels of processing, but for the development of theories of memory generally.

The first thing to notice is that the accusation of circularity is at most only partially justified. The distinction of qualitatively different domains of processing, such as semantic versus phonemic, can be made quite independently of any effects such processing might have on memory performance; the hypothesis that processing to a semantic level yields better free recall than processing to a phonemic level is quite falsifiable. Craik and Tulving (1975), for example, specified the level of processing of their orienting tasks quite independently of the subsequently observed memory performance. There are many other examples of experiments in which the level of processing of an orienting task has been specified by an a priori analysis of the processes involved in performing the task, and in these cases there can be no question of circularity. Rather, the accusation of circularity arises in relation to within-domain differences such as those obtained between items that warrant "yes" and "no" answers to orienting questions, both of which obviously entail semantic-level analysis, or to differences between qualitatively distinct orienting tasks, say between evaluating synonymity of adjectives and evaluating those same adjectives as self-descriptors. In these cases, there appears to be no index of depth other than through the post hoc observation of their differential effects on memory performance.

What would an independent index look like? One possibility is that measures such as processing time, processing effort, or some neurophysiological measure might provide such an index. But how might this index be validated? Surely not through establishing that it has a monotonic relation to memory performance; apart from being doomed to almost certain failure, such a simplistic strategy would certainly embody the circularity of which we have been accused. In fact, it was never envisaged that depth could be thought of as a unidimensional continuum, simply measured, with a monotonic relation to memory performance – a kind of updated synonym for strength. It is unfortunate in many ways that the literal interpretation of our spatial metaphor lends itself so readily to the idea of simple unidimensional measurement. However, an independent index, were it to exist, would be an integral part of a general theory of cognitive functioning. A measure of the depth of processing involved in performing some task would be the numerical output from some function applied to an explicit model of the processes involved in performing that task. For example, a numerical index of the depth of processing involved in deciding that the word *dog* refers to an animal could only be obtained from an explicit model of the operations on semantic memory needed to answer questions of category membership. Hence a complete definition and quantification of terms such as depth or level is appropriately thought of as the end product of research, not the starting point. In this sense, our original use of these terms was a tentative label to characterize processes that require further explication and specification through ongoing research.

There are many examples of this ongoing development and refinement, some of which will be discussed below. For present purposes the case of self-reference orienting tasks provides a good example of the sense in which depth of processing is a general concept needing to be explicated, rather than a well-defined explanatory construct. The self-reference phenomenon first reported by Rogers, Kuiper, and Kirker (1977) is interesting for several reasons. In the first place the high level of performance produced by this incidental orienting task, relative either to intentional instructions or to other semantic orienting tasks, illustrates the point that the relation between memory performance and various forms of semantic-level processing is not a simple one. It clearly represents little advancement in explanatory understanding to state that the superiority of the self-reference orienting task is a consequence of it requiring deeper levels of processing; it is scarcely better than saying that it yields a stronger trace or that it increases the probability of the item entering a particular memory store. At best the appeal to depth as an explanatory construct is a weak explanation of why the self-reference orienting task yields better performance than, say, a rhyming orienting task. However, contrary to some critical reviews of levels of processing, we did not consider such an appeal to the concept of depth to be the end of the matter but the correct starting point in the quest for an adequate explanation.

Moreover, notice that it was levels of processing as a research framework that did much to stimulate the collection of such data in the first place. The self-reference phenomenon comprises precisely the kind of data needed to build a comprehensive theory of memory encoding; and it was the levels framework that stressed that its explanation was to be found in an analysis of the particular meaning-extracting operations involved in performing the orienting task. Obviously such an analysis, if it is to isolate those principles that are important for remembering, may require considerable research into the structure of the self-concept, as well as into how such judgements are made; but, when completed, the achievement will have been to provide an explication of what was initially meant by depth. Such an analysis is well under way (e.g., Bellezza, 1984; S.B. Klein & Kihlstrom, 1986). It may well be that self-reference is effective because performing the comparison judgement effectively structures the elements into an organized set, as S.B. Klein and Kihlstrom have argued; but as a research framework, levels of processing is entirely neutral on the matter. Finally, notice how such a research programme provides yet another example of the ongoing integration of memory research into the broad context of cognition.

Principles and Mechanisms

In his critical assessment of levels of processing, Baddeley (1978) suggests that memory researchers should abandon the search for broad general principles and concentrate instead on explaining specific components of the memory system. As an example of the latter approach, Baddeley cites his own work on the articulatory loop concept within the notion of working memory. The argument is that such specific components are more amenable to experimental explanation and are also perhaps more readily identifiable in neuropsychological and even in neuroanatomical terms. We discuss the concept of working memory below. Our concern in this section lies more in the domain of philosophy of science: whether it is more profitable to induce broad general principles and then use experimentation as a sort of dialogue with nature to refine the concepts in question, or more profitable to postulate smaller-scale mechanisms and structures whose reality can be confirmed through various types of experimentation.

This is not an either/or matter, since obviously we need both principles and mechanisms. Mechanisms must be located within a larger cognitive architecture that embodies general principles; a general theory of memory cannot be constructed merely by assembling mechanisms. As we have already pointed out, one of the major shortcomings of early models of short-term memory was their isolation from general cognitive processes. In many models the major mechanism for establishing a memory trace was rehearsal. Rehearsal was regarded as a mechanism that (with a certain probability) moved items from one (short-term) store to another (long-term) store and did so without reference to other aspects of the cognitive architecture, such as long-term knowledge structures or skilled procedures. The broad architecture of later models, such as Anderson's ACT* (1983), stands in marked contrast to these early modular theories.

Another example of the danger of overemphasizing isolated mechanisms is the enormous body of research that was launched in an effort to understand the mechanism of scanning in the Sternberg paradigm. Granted that the basic phenomenon (Sternberg, 1966) is interesting and important, the same cannot be said for much of the research it stimulated. Most of this research is now forgotten for the simple reason that the scanning mechanism and its associated task became objects of study in their own right. Rather than seeking to understand how such a mechanism might relate to general principles of cognitive functioning, many researchers regarded the task and its explanation as an end in itself.

Having made this point, it must be acknowledged that general principles without mechanisms do not constitute a theory. A major function of general principles is to guide the data gathering process by posing interesting questions, suggesting testable hypotheses, and distinguishing important phenomena from laboratory trivia. But general principles do not constitute an adequate explanation of the data they generate. It should not be thought, for example, that the general principle of depth of processing would provide a complete explanation of why various orienting tasks have the effects they do, even if we possessed an independent index of depth; such a general principle does not constitute a mechanism, and so a concept such as deep processing should not be thought of as a mechanism that directly influences memory. Judging that the word *dog* refers to a domestic animal may lead to its being better recalled than judging that it rhymes with *fog*, but the mechanisms through which this difference is effected remain to be established. Standing between the functional description of the orienting task and the subsequent retrieval are mechanisms that must be identified and included as part of any complete theory of remembering. However, it will be obvious that from the standpoint of levels of processing, these mechanisms are appropriately formulated in procedural rather than structural terms. Of course there must be structures at the neurological level sustaining the procedures; but, in terms of hypothetical mechanisms at the cognitive level of analysis, we see no need to posit structural mechanisms intervening between procedures and their neurological substrates.

Levels: A Fixed Series of Stages?

Baddeley (1978) questioned the reasonableness of our apparent assumption of a single linear sequence of levels of processing. Insofar as our original statement can be interpreted as assuming that the processing of incoming stimuli proceeds in a strict temporally ordered sequence of stages from shallow sensory levels to deeper semantic levels, then Baddeley's criticism is well-founded. It must be admitted that our original statement did little to discourage such a simplistic interpretation, but it is one that we ourselves rapidly sought to correct (Lockhart et al., 1976). The problems associated with experimentally evaluating claims about the sequencing and temporal relationships among levels of processing have been analyzed thoroughly by Treisman (1979) and will not be reviewed here. However, it is important to realize that the validity of our basic framework has never depended on any specific model of how the cognitive system achieves a particular level of processing. It is likely that an adequate model will comprise complex interactions between top-down and bottom-up processes, and that processing at different levels will be temporally parallel or partially overlapping. What *is* important to our position is not the sequencing of different stages of processing but the pattern of analyses finally achieved.

EMPIRICAL ISSUES

Types of Rehearsal

In our original article we suggested that the concept of rehearsal could be usefully broken down into the maintenance of a level of processing already achieved and further elaboration of the stimulus. We further argued that only the second type of rehearsal would be associated with an increase in subsequent retention. This formulation was at odds with the claims of Atkinson and Shiffrin (1968) whose model predicted that mere residence in the rehearsal buffer served to transfer information across to the long-term store. The implausible suggestion that maintenance processing does not lead to an improvement in memory received empirical support from studies by Craik and Watkins (1973) and by Woodward, Bjork, and Jongeward (1973). However, the latter study showed that maintenance rehearsal was associated with an increase in subsequent recognition of the rehearsed items. A further complication was pointed out by T.O. Nelson (1977) who demonstrated that when subjects were required to process twice at a phonemic level, recall was superior to performance following one phonemically processed presentation.

In his thoughtful review of work in the area, Greene (1987) concluded that pure maintenance rehearsal *does* lead to small increases in subsequent recall, although the data are inconsistent. The evidence for an increase in recognition memory is clearcut. One possibility here (discussed by Greene, 1987, and by Naveh-Benjamin & Jonides, 1984) is that recall depends primarily on interitem elaboration, whereas recognition is incremented by intraitem integration (Mandler, 1979); if maintenance rehearsal leads to little interitem processing, the observed result would be expected. Naveh-Benjamin and Jonides make the further interesting suggestion that maintenance rehearsal is a two-stage activity. The first stage consists of assembling the items and setting up an articulatory programme; the second stage requires comparatively little effort and involves repetitive execution of the rehearsal programme. The suggestion is that memory is enhanced only at the first stage.

In addition, the more recent work surveyed by Greene (1987) makes it clear that a distinction must be drawn between repetition of items which does lead to a clear increase in memory (T.O. Nelson, 1977) and rote maintenance which is associated with much weaker changes. It is also reasonably clear (and clearly reasonable!) that maintenance rehearsal of linguistic items is associated with an enhancement of the acoustic/articulatory properties of the material (Elmes & Bjork, 1975; Geiselman & Bjork, 1980). The Geiselman and Bjork study also demonstrated that the effectiveness of particular retrieval information depended on the qualitative type of rehearsal activity initially performed.

In summary, the situation regarding the relations between types of rehearsal and subsequent memory performance are considerably more complex than envisaged by Craik and Lockhart in 1972. Nonetheless, we would argue that our original suggestions were along the right lines, stressing as they did qualitative types of rehearsal. The fact that our original proposals were too simplistic is much less important than the gains in understanding that have arisen from 15 years of research in rehearsal processes. Rather than abandon the problem because the approach is based on "false assumptions" as Baddeley (1978) seems to advocate, it has been quite fruitful to pursue it and attempt to resolve empirical inconsistencies. One interesting footnote in this connection is that the second stage of maintenance rehearsal as suggested by Naveh-Benjamin and Jonides (1984) – the execution of a relatively automatic articulatory programme — bears a very close resemblance to the articulatory loop subsystem in Baddeley's working memory model. As we understand Baddeley's model, there is little reason for expecting learning to occur as a function of maintenance in the articulatory loop.

Domains and Levels Within Domains

One of Baddeley's (1978) criticisms of the levels of processing framework was that it was not at all clear how levels within the broad domains of phonemic and semantic coding might be specified. This criticism seems reasonable to us. In fact, Lockhart et al. (1976) had suggested that the original notion of depth might refer to two rather different concepts: the first pertaining to qualitatively different domains, with analysis in shallow sensory domains typically preceding analysis in semantic domains; the second pertaining to further analyses within a domain, and here we conceded that elaboration might capture the essence of the argument better than notions of depth. However, as in the other empirical areas surveyed in this article, many useful studies on the relations between types of processing and subsequent retention have been carried out in the last 15–20 years, and again we would argue that our framework provided the theoretical motivation for much of this research.

In the domain of semantic processing, there is general agreement that further analysis of the stimulus is associated with higher levels of retention. Thus, greater enrichment or elaboration of encoding is associated with enhanced memory performance (Anderson & Reder, 1979; Craik & Tulving, 1975). Greater amounts of processing in this sense have been specified operationally in terms of the number of semantic decisions made in connection with a word (e.g., Johnson-Laird, Gibbs, & de Mowbray, 1978; Ross, 1981). A related idea is the amount of effort involved in the encoding process (Tyler, Hertel, McCallum, & Ellis, 1979), although we would argue that memory performance depends on the qualitative type of encoding achieved, not time or effort as such. Some types of elaboration are likely to be more beneficial than others; for example, Hashtroudi (1983) has shown that nouns encoded in the context of core-meaning adjectives were recalled better than nouns encoded in the context of adjectives whose meanings were peripheral to the central meaning of the nouns.

A second line of argument relating to further semantic processing suggests that distinctiveness of encoding is the key concept associated with good retention. For example, Frase and Kamman (1974) showed that words categorized in terms of their general properties (e.g., foods) were less well recalled than words categorized in more specific ways (e.g., vegetables). In a similar demonstration, K. Klein and

Saltz (1976) suggested that distinctiveness might provide the mechanism underlying the depth-of-processing effects. Also, Moscovitch and Craik (1976) showed that uniqueness of encoding conferred a potential benefit to later retrieval and suggested that the compatibility of retrieval information determined the extent to which the potential was realized. The notion of distinctiveness has been developed further by Bransford, Stein, and their collaborators in terms of precision of encoding, the relationship of encoding to existing knowledge, and the availability of appropriate cues at the time of retrieval (Bransford, Franks, Morris, & Stein, 1979; Stein, 1978). Finally, D.L. Nelson (1979) again stressed the concepts of distinctiveness and interactiveness (the compatibility of encoded events with each other, and with preexisting knowledge), observing that depth was an unnecessary construct once these other two concepts were taken into account.

So, in this instance, we tend to agree with Baddeley that the original notion of depth is too simple a formulation to provide an adequate analysis of situations involving further processing within a given domain, except perhaps in the sense that such further, more elaborate analyses often require more effort and more processing resources. Nonetheless, the many studies relating differences in types of encoding to subsequent retention that were stimulated (or provoked perhaps!) by the levels of processing framework have undoubtedly added to the richness and precision of our knowledge of memory processes.

Long-Term Retention of Sensory Features

One central idea of the original levels of processing formulation was that shallow (typically sensory) processing is associated with the formation of very transitory memory traces. This concept was, of course, in line with the concept of sensory registers or stores, whose contents were rapidly lost through passive decay or overwriting by further stimuli in the same modality. Since 1972, however, a number of results have been reported in which sensory information persists for hours, minutes, and even months. Kirsner and Dunn (1985) summarize a number of such findings and suggest that input modality may be retained in several forms, varying from a literal perceptual record to a more abstract representation of surface qualities.

It seems clear at this point that our original suggestions were again too simple. Surface information is often lost rapidly, but there are also cases in which a record of surface form clearly persists to affect later performance over long retention intervals. We can offer a few comments from the perspective of levels of processing. First, some of the cases involving the dramatically long-term retention of surface form do not involve the explicit retrieval of surface information, but involve the implicit use of such information to facilitate current performance. Kolers's (1979) demonstrations of long-lasting information regarding transformed typographics fall into this category. In general, it now seems that many perceptual memory (priming) tasks are data-driven and are sensitive to modality-specific information (e.g., Roediger, Weldon, & Challis, 1989). However, even if such modality-specific information cannot be explicitly retrieved, it must be represented in some manner to affect performance on implicit memory tasks, and such findings are incompatible with our original statement.

A second point is that recall of input modality is recall in a rather different sense from recall of a scene, or of a conversation, or of a word list. In these latter cases, the subject must bring to mind highly specific information about the original event; however, when recalling modality, the subject often simply chooses between visual or auditory presentation modes, and the information recalled is of a much more general type. Third, in some cases at least, the surface form may modulate its accompanying verbal information so that information about modality is preserved in the abstract gist of the message; for example, an utterance by Speaker A may have quite different implications from the same utterance spoken by Speaker B. A related case is one in which surface information is transformed into deep information through the gaining of expertise; for example, wine-tasting to the novice may consist simply of the 2 × 2 classification red/white × sweet/dry, whereas to the expert the same sensory data may convey a wealth of meaningful information and evoke rich associations. The term deep was never meant to be restricted to linguistic meaningfulness.

Picture memory is clearly a case in which the stimulus can be processed rapidly to a deep level and so be well retained (Craik & Lockhart, 1972, p. 676). Less easily explained are the findings of Intraub and Nicklos (1985) who reported better memory for pictures following physical than following semantic orienting questions. They account for their results in terms of enhanced distinctiveness, a concept that Conway and Gathercole (1987) also invoke to describe their findings of long-term retention of input modality.

A final interesting situation in which long-term retention of surface information is obtained involves excellent verbatim memory for emotional utterances (Keenan, MacWhinney, & Mayhew, 1977). A satisfactory theoretical account still has to be developed, but two suggestions are, first, that such emotional events may function as weaker versions of "flashbulb memories" (Brown & Kulik, 1977), and second, the involvement of the self in emotional utterances may evoke self-referential processing which, as noted previously, typically yields excellent long-term retention.

In any event, it is clear that sensory or surface aspects of stimuli are not always lost rapidly as we claimed in 1972. Some contrary results may reflect the operation of a somewhat different procedural memory system, and some may reflect deep nonlinguistic processing. It seems clear in 1990 that more surface information is retained than most people believed in 1972, but a satisfactory account of all the new findings has still to emerge.

FURTHER CONCEPTUAL DEVELOPMENTS

Depth, Elaboration, and Distinctiveness

The concept of depth of processing was an initial attempt to sketch a framework for how qualitatively different forms of processing might be related to memory performance. The fundamental claim was that, in terms of their impact on remembering, these perceptual/conceptual analyses could be organized at least to a first approximation in terms of the degree to which they entailed the extraction of meaning. It was never intended that the notion of depth should be thought of as some fixed and well-defined concept ready to take its place in a precise formalism. It was, as noted above and as critics have pointed out, a largely common sense statement; however, it captured an insight that we thought needed emphasis, and its obviousness is more apparent in retrospect than it was in 1972. The fundamental insight was well expressed by William James: "The art of remembering is the art of thinking . . . our conscious effort should not be so much to *impress* or *retain* (knowledge) as to *connect* it with something already there" (James, 1983, p. 87).

Early experimental work within the framework of levels of processing was concerned to document the way in which different orienting tasks produced large and predictable effects on subsequent memory performance. These orienting tasks were chosen to serve as reference points that would mark in an obvious way different levels of processing. Apart from this demonstrative function, they enjoyed no special status. There has been a tendency for these particular orienting tasks (for example, judging category membership, rhyme, or case) to acquire a privileged status as defining different levels of processing and for levels to be thought of as an independent variable with two or three values. Such a categorical orientation leads to criticisms that are rather beside the point, such as the claim made by D.L. Nelson (1979) that particular orienting tasks are impure representations of their putative levels of processing. The claim is no doubt true. The Stroop effect would suggest, for example, that in reading common words there is inevitably some semantic processing, even in a nominally phonemic-level task, and, as Treisman (1979) points out, it should not be assumed that processing can optionally be stopped at any given level of processing. But such a criticism would be relevant only if one were interested in the precise influence of a particular cognitive operation (judging rhyme, say) for its own sake.

These comments should not be taken as criticism of the use of these marker orienting tasks in experiments designed to examine the relative sensitivity of different forms of remembering to a wide range of levels of processing. It is, for example, of great interest and a challenge to a comprehensive theory of memory that some forms of implicit memory are unaffected by different orienting tasks, even though the tasks vary over a wide range of levels of processing. In addressing research questions such as these, the use of orienting tasks merely to represent a wide range of processing levels is perfectly justified. Rather, our criticism is of the tendency to treat levels as a small number of discrete (albeit ordered) classes and to regard the processing within each class as equivalent. But as we saw in the discussion of the alleged circularity of levels of processing, there is no such simple characterization of levels, and research aimed at explicating the concept of depth must not stop with these demonstration orienting tasks.

An example of such explication is to be found in attempts to exploit the terms elaboration and distinctiveness. The concept of elaboration was introduced by Craik and Tulving (1975) in their interpretation of results from a series of experiments carried out to illustrate levels of processing ideas. Their basic paradigm was one in which a series of words was presented; each word was preceded by a question that related to the case that the word was printed in, to the rhyming characteristics of the word, or to its semantic nature. The answer to the question could be either "yes" or "no." So, for example, the word BRUSH might be presented in capital letters. Examples of possible preceding questions (and answers) would thus be:

"Is the word in small print?" (case-no); "Does the word rhyme with cotton?" (rhyme-no); "Is the word something used for cleaning?" (semantic-yes). The preceding questions acted as orienting tasks that induced either shallow processing of the visual characteristics of each word (case questions), deeper phonemic processing (rhyme questions), or even deeper semantic processing. In line with our predictions, subsequent recall and recognition of the words was profoundly affected by the nature of the initial question, with semantic processing being associated with later performance levels that were two to six times higher than the levels associated with surface processing.

However, one unexpected result was that questions leading to a "yes" response were associated with higher memory performance levels than questions leading to a "no" response. Craik and Tulving's (1975) suggestion was that the compatible (positive) questions served to enrich or elaborate the encoded trace of the target word in a way that the incompatible (negative) questions could not. Thus, even the simple phrase "something used for cleaning" can specify and enrich the subjectively encoded concept of BRUSH in a way that the incompatible phrase "something found in the jungle" cannot. The further suggestion, then, was that richly elaborated traces supported better recall and recognition.

The ideas of encoding richness or elaboration thus suggest the metaphor *spread of encoding*, and some writers have suggested that extensiveness of processing or spread of encoding are better descriptions than depth of processing (e.g., Kolers, 1979; Murdock, 1974). Another version of this argument is that deeper levels of processing (i.e., greater degrees of meaningfulness) simply give greater scope for elaborative processes, so that elaboration is the only concept necessary (Anderson & Reder, 1979). We have taken the view that both depth and elaboration are necessary terms, with depth referring to qualitatively different types of processing and elaboration referring to rich or impoverished processing within any one qualitative domain (Lockhart et al., 1976).

Does elaboration have a greater effect at deeper levels? This question may not be answerable because of great differences in the qualitative nature of the encoding processes in question. It is like asking whether an increase in the brilliance of a diamond has a greater effect on its value than an increase in its size. The answer is that both factors affect value, but it is not possible to compare them directly.

However, it may be sensible to suggest that greater degrees of knowledge and expertise afford richer and more elaborate analyses and interpretations. To a non-Arabic speaker, a written sentence in Arabic can only be interpreted as a visual pattern; the nonword GLARFON yields no clear interpretation, but it can be pronounced and may be given some idiosyncratic meaning in terms of similar-sounding words; finally, a nonword can become meaningful through usage (e.g., GLASNOST) and thus yield rich images and associations. Is elaboration all that is required then? We argue that both notions of depth and elaboration are necessary: depth to talk about qualitative types of analysis and to capture the notion that some types of processing (typically sensory analyses) precede others and require few attentional resources to achieve them; elaboration to refer to richness or extensiveness of processing within each qualitative type. The last point may be illustrated by contrasting proof-reading and reading for gist; the former involves elaborate processing of visual and lexical features with little involvement of meaning, whereas the latter involves relatively elaborate semantic processing with correspondingly slight analysis of spelling patterns and the like.

This example may provoke the comment "Well, the former type of processing may be better for later memory of type font or of specific wording, whereas the latter type of processing may be better for later memory of meaning, but who is to say that one type is generally superior to the other?" In essence, this was the influential argument put forward by Morris, Bransford, and Franks (1977) under the rubric of transfer appropriate processing. We strongly endorse the idea of transfer appropriate processing (and the very similar notions of encoding specificity and repetition of operations); in many cases, the argument may involve comparison of apples and oranges, and we have no quarrel with the conclusion that for good apple-memory one should do apple-processing and for good orange-memory do orange-processing. But in many other cases there *is* some common measure by which different types of processing can be legitimately compared — how well a specific word or event can be recalled or recognized, for example — and in such cases it is surely reasonable to compare the relative effectiveness of different types of initial processing. It is worth noting, for example, that in the Morris et al. experiments whereas initial rhyme processing was indeed superior to semantic processing for a subsequent rhyming recognition test and semantic processing was superior to rhyme processing for a standard (semantic?) recognition test, the two compatible conditions were by no means equivalent. That is, the mean value for "rhyme-yes" processing followed by a rhyming test was .40 (averaged over Exp. 1 and 2), whereas the mean value for "semantic-yes" processing followed by a standard recognition in the same two experiments was .68. When memory (in the ordinary sense of recall or recognition of an event) is considered, some types of initial processing are superior to others, and these types involve deeper, semantic types of analysis. We will return to the concept of transfer appropriate processing in our discussion of retrieval.

The concept of elaboration has also been tied to the concept of distinctiveness of the memory trace (Craik, 1977; Jacoby & Craik, 1979; Moscovitch & Craik, 1976). The distinctiveness of traces as a determinant of their memorability has been discussed by several previous writers (e.g., Murdock, 1960), and the distinctiveness or specificity of connections between cues and traces is the central explanatory notion lying behind the phenomenon known variously as A-B, A-C interference, cue overload, the fan effect, and the category size and list length effects. In all of these cases, a cue word or concept is more effective at retrieval when relatively few items were associated with it, or nested under it, at the time of encoding. The benefit to recognition and recall conferred by distinctiveness of the memory trace points up the similarities between remembering and perceiving. Just as a distinctive stimulus stands out and is therefore readily identified against a background of different stimuli, so a distinctive memory trace stands out and is therefore readily retrievable, especially when the wanted trace can be contrasted to other dissimilar encoded events as part of the retrieval process (Jacoby & Craik, 1979).

What is the mechanism that ties greater degrees of encoding elaboration to good memory performance? The account favoured by Anderson (1976) is that encoding an event in terms of some rich knowledge base allows the system to generate further elaborative productions; in turn, these productions constitute a redundant network of encoded information, thereby enhancing the chances of successful retrieval. An alternative account is that depth and elaboration of processing together allow the production of distinctive encodings, and that it is this distinctiveness against the background of well-learned meaningful knowledge that is the crucial feature (Ausubel, 1962). Why is the elaboration of surface features not so helpful for later memory as is the elaboration of deeper features? After all, an elaborate though meaningless colourful pattern can be just as perceptually distinctive as an elaborate picture of a scene. The answer seems to lie in the fact that the latter example forms one coherent image, whereas the former pattern is composed of unrelated elements. Meaningful schematic knowledge allows both for the interpretation of an incoming stimulus array as a unitary configuration and also for its reconstruction at the time of retrieval.

A related point that several researchers have made (e.g., T.O. Nelson, 1977) is that we should talk simply about different types of processing and not about levels or depth. That is, what evidence is there that the cognitive system is organized hierarchically, with some types of processing requiring more time or effort to achieve? Alternatively stated, is it really necessary to go through all shallow levels to reach deeper levels? Clearly some analysis of physical features is necessary before meaning can be analyzed; also, it is often the case that incoming stimulus patterns are comprehended in terms of language before more abstract analyses are carried out or action plans formulated. When language is involved, it again seems reasonable to assume that some analysis of phonemes or graphemes is necessary before words and meaning are identified. But not all patterns are analyzed in linguistic terms — the analysis of a visual scene surely proceeds in a very different way from the analysis of auditorily perceived speech. Acknowledging these points, Lockhart et al. (1976) put forward the alternative notion that analysis proceeded in domains of knowledge, tied to some extent to specific sensory modalities for shallow (or early) analyses, and proceeding to deeper levels of analysis within that specific domain. In some cases, such deeper knowledge-informed analyses remain restricted to their specific domain (expertise in wine-tasting, for example, or in judging the authenticity of a painting), whereas in other cases (notably language) deeper analysis in one cognitive domain (e.g., reading text) has rich interconnections with other domains (e.g., spoken language). Such rich and extensive cross-modal connections have been suggested by many other theorists, of course, and they are the central concept of associationism. Another proponent of this idea is Paivio (1971, 1986) with his suggestion that language and imagery processes must be extensively interconnected.

Thus it seems to us that we can say something more than different types of processing exist. Related types of processing can be organized into domains with sensory and abstract/conceptual aspects. Deeper processing is still a sensible notion, where deeper refers to the greater involvement of processes associated with interpretation and implication within the relevant domain. Greater knowledge and expertise enable deeper processing to be accomplished, and such processing is associated with superior memory performance. To reemphasize one point: The concept of depth of processing is *not* restricted to the linguistic domain as some critics appear to have assumed (e.g., Kolers & Roediger, 1984); the growth of expertise affords deeper processing in any domain of knowledge. Within a domain, analyses may typically proceed in similar ways from one occasion to the next (that is, from predominantly sensory analyses to predominantly conceptual analyses), but, as acknowledged previously (Lockhart & Craik, 1978; Lockhart et al., 1976), it is undoubtedly too simplistic to argue that all stimuli are analyzed by one common cognitive hierarchy or that all patterns are processed through the same sequence of levels of analysis.

Levels of Processing, Short-Term, and Working Memory

Our criticism of short-term memory was directed towards its structural incarnation in "box and arrow" models such as that of Atkinson and Shiffrin (1968). Such models reify memory structures in a way that leads to unprofitable questions (because they are unanswerable), such as the capacity of, or the rate of decay of items from, short-term memory. We will consider these two examples in greater detail.

The Capacity of Short-Term Memory: Capacity is a strictly quantitative concept, and so the first problem for any attempt to measure the capacity of short-term memory is to specify the appropriate units of measurement. Attempts to use metrics based on some function of the amount of information remembered (number of items, chunks, bits, etc.) have never succeeded in establishing the universal constant sought after. The reason for this failure is not difficult to find. *Capacity*, as measured in terms of units of remembered material, is dependent on the processing skill or resources of the subject interacting with properties of the material to be remembered (cf. Jenkins, 1979). For this reason, measured capacity can vary over an enormous range, either within the individual across content domains, or across individuals for a single domain.

Perhaps the most dramatic example of this point comes from recent research in the learning of memory skills. In one of the best known examples Chase and Ericsson (1981, 1982) report one subject who acquired an immediate digit span of 80 by virtue of extended practice of recoding digits into meaningful sequences. On the other hand, such extended practice left the same subject's letter span unchanged. In order to describe the capacity of short-term memory in terms of number of items that is constant across content domains, it becomes necessary to redefine *item* as a function of the processing operations applied to the material to be remembered. In the case of Chase and Ericsson's subject, it would become necessary to define higher order units (chunks) in terms of the knowledge base being used to structure the incoming digits; but precisely how this is done to yield a measure of item capacity is anything but clear. Equally unclear is the theoretical value of such an exercise, even if it could be successfully completed, since it would be nothing more than a measure derived from more basic theoretical constructs. Our argument is that to claim that information processing is constrained by the structural properties of a limited capacity short-term memory is to confuse cause with effect and is a bad way of stating the case. The apparent limited capacity of short-term memory is a consequence of limited processing resources.

Posed in these terms, questions of capacity can be seen to confuse structure and process; limited capacity turns out not to be an invariant structural property of a memory system, but a limitation imposed by processing. Capacity, measured in terms of quantity of material remembered, will be highly variable depending on the readily modifiable skills of the rememberer in relation to the nature of the material to be remembered. That is, if, as we have argued, memory is a function of the form of processing, then insofar as it is necessary to invoke a concept of limited capacity, the appropriate units of measurement are not units of memory information but of processing resource. Thus, rather than say that Chase and Ericsson's subject increased short-term memory capacity in any structural sense, it makes much more sense to say that he has acquired processing skills that enable him to process deeply long sequences of numbers with extraordinary rapidity. The advantage of this approach is that it integrates memory encoding with theories of skilled performance (Chi, Glaser, & Farr, 1988), with theories of attention, especially that of Kahneman (1973), and, as we discuss in more detail below, Baddeley's working memory. In brief, our advocated approach is an example of a putative structure of memory encoding being integrated into general cognitive theory.

Rate of Decay or Loss From Short-Term Memory: A similar analysis can be made of attempts to define short-term memory in terms of rate of decay. With little sense of tautology, it was commonly claimed that a defining characteristic of short-term memory was the rapid decay or expulsion to which its contents were subject.

This parametric approach to memory structures has a long tradition. Starting with Ebbinghaus, there has been a quest for universal constants of forgetting. The modern-day counterpart was the theoretical use made of data from the Brown-Peterson paradigm as if the form of processing allowed by a particular combination of item exposure times, items of a particular kind, subjects with particular processing skills, and a particular form of processing instructions, would yield forgetting data of universal significance. From the viewpoint of levels of processing, the claim is that, with particular combinations of these various factors, it is possible to produce virtually any forgetting function one could care to nominate and that there is nothing of special theoretical significance with the particular conditions commonly employed either by Ebbinghaus or in the Brown-Peterson paradigm. Another way of stating this position is that if memory structures are to be defined in terms of their forgetting characteristics, then, so defined, there is a continuum of memory systems. Such a conclusion suggests that either the notion of memory systems should be discarded or some other basis for defining them should be found. Such an alternative basis is exemplified in Baddeley's working memory.

Working memory. Working memory is the set of processes and/or structures concerned with the temporary holding, manipulation, and integration of various types of information. Such a concept is thus largely orthogonal to the problems addressed by levels of processing – the relations between encoding processes and long-term retention. Hence, we have no quarrel with the concept of working memory. Indeed we believe that it is a necessary concept in some form and that it can coexist very comfortably with levels of processing notions. Whereas it is widely believed that one of the main purposes of our 1972 paper was to abolish the distinction between short-term and long-term memory, such a claim is simply untrue, as a rereading of the paper will show. We argued against the structural notion of memory stores, and short-term store was the most obvious target. The bulk of the paper is concerned with the relations between initial encoding operations and subsequent retention, but we go on to say:

> However, superimposed on this basic memory system there is a second way in which stimuli can be retained — by recirculating information at one level of processing. In our view, such descriptions as "continued attention to certain aspects of the stimulus," "keeping the items in consciousness," "holding the items in the rehearsal buffer," and "retention of the items in primary memory" all refer to the same concept of maintaining information at one level of processing. To preserve some measure of continuity with existing terminology, we will use the term primary memory (PM) to refer to this operation, although it should be noted that our usage is more restricted than the usual one. ... The essential feature of PM retention is that aspects of the material are still being processed or attended to. Our notion of PM is, thus, synonymous with that of James (1890) in that PM items are still in consciousness. (Craik & Lockhart, 1972, p. 676)

The essential differences between this characterization of primary memory and the concept of a short-term store are, first, that primary memory is a processing activity, not a structure, and that the phenomenon of limited capacity is therefore a processing limitation, a consequence of limited processing resources as discussed above, not a limitation of "space." The second major difference is that primary memory is not one "place" in the cognitive system, but, rather, represents the deployment of processing resources to various parts of the system, thereby giving rise to the phenomenological experience of paying attention to various types of information — phonological, pictorial, semantic, or whatever. By this view the puzzle of what constitutes *the* short-term code (acoustic? articulatory? visual? semantic?), which occupied many researchers in the 1960s and 1970s, simply dissolves. The qualitative contents of primary memory are given by the particular processes that are active at the time. A very similar set of views was put forward by Shiffrin (1976) who argued that short-term store did not precede long-term store (as it did in the "modal model"), but was an active subset of long-term store. Further, "Short-term store is assumed to recapitulate the structure of LTS, in which it is embedded" (p. 217).

Is the concept of working memory as proposed by Baddeley and Hitch (1974) and developed by Baddeley (1986) analogous or identical to this view of primary memory as an active processor? The concepts clearly share some features: Both primary memory and working memory are processors as opposed to passive memory stores, and both deal with various types of information. But whereas the Craik and Lockhart concept of primary memory is a set of processing activities carried out in various parts of the cognitive system, Baddeley's working memory appears to be more structural, located in one place, and receiving, integrating, and managing different types of information. The core of Baddeley's working memory is the *Central Executive* whose function is to co-ordinate the activities of such peripheral slave systems as the articulatory loop and the visuospatial sketchpad. But what is the nature of the Central Executive? Can its capacity or its managerial prowess be independently measured? In the absence of satisfactory answers to such questions, it is difficult to see how the concept escapes Baddeley's (1978) own criticisms of levels of processing that the ideas are poorly defined, too general, and not easily amenable to measurement.

The concept of a flexible primary memory within the levels of processing framework has much more in common with the domain-specific view of working memory put forward by Monsell (1984) and by Daneman and Tardif (1987). In their version, working memory is not one thing, but is the collection of computational abilities associated with many types of information. According to Monsell, working memory "is no more (or less) than a heterogeneous array of independent storage capacities intrinsic to various subsystems specialized for processing in specific domains" (p. 344). By this view then, working memory is an umbrella term for a number of highly specific processing abilities. This characterization leaves the important question of how the distributed system of processors is controlled and directed. Monsell suggests that some higher order processors are simply specialized to monitor the activities of other processors and direct the flow of information among them. The Craik and Lockhart version equates primary memory with attention paid to a particular type of information; there is no reason why attention (or processing resources) should not be directed at two or more types simultaneously, although such a notion is admittedly vague. The executive control of working memory, however conceptualized, remains something of a puzzle for all theorists at the present time.

To summarize our thoughts on short-term memory functioning: First, Craik and Lockhart (1972) did not deny the usefulness of some form of distinction between long-term and short-term memory; in fact, we endorsed the distinction, although primary memory was seen as the flexible allocation of processing resources within the cognitive system, not as a separate structure. Second, capacity limitations are seen as processing limitations, not as structural limitations. Third, the concept of working memory is perfectly compatible with the levels of processing framework, although the notion of working memory as a collection of domain-specific processors (Daneman & Tardif, 1987; Monsell, 1984) is more congenial to our way of thinking than is Baddeley's (1986) model.

Levels of Processing and Retrieval

Another frequent criticism of levels of processing has been its neglect of retrieval. Insofar as the criticism is one of incompleteness, it has some justification, although the issue of retrieval was addressed by Lockhart et al. (1976). However, if the criticism is that we considered retrieval processes unimportant or that most of the variance in memory performance is to be accounted for by encoding factors alone, then it is quite unfounded. As pointed out above, retrieval is the only process that can be meaningfully labelled a distinctly memory process. Thus, rather than assigning retrieval processes a secondary status, levels of processing assigns them a dominant role in a theory of remembering.

However, there are a number of issues that cannot be sidestepped by this general apology. The most important of these is the possibility that the concept of memory performance being a function of depth of processing is wrong, even as a first approximation, and that the typically observed superiority of semantic processing is an artifact of retrieval conditions which, so it is claimed, are typically better matched to semantic level processing. The argument is that sensory processing can be as well remembered as semantically processed material, provided the retrieval conditions are appropriate for that form of processing. An example of this kind of argument is transfer appropriate processing (Morris et al., 1977).

The first point to note is that the general principle that memory is determined by the relationship or interaction between the form of processing at acquisition and at retrieval is undoubtedly true. Apart from the evidence supporting transfer appropriate processing, the entire body of research surrounding Tulving's encoding specificity principle (Tulving, 1983) makes this claim incontrovertible. However, the critical issue is not the existence of such encoding/retrieval interactions, but whether there are also main effects determined by the form of encoding, Granted the relevance of transfer appropriate processing, are there levels effects even when the transfer appropriateness of the initial processing is held constant? As pointed out in the earlier discussion of elaboration, such a main effect does exist. However, in terms of retrieval processes there are two possible ways that this effect might operate.

One is that relative to shallow encoding, deep processing decontextualizes retrieval. That is, granted that retrieval is never totally independent of the retrieval context, deep processing sustains retrieval over a wider range of retrieval conditions and contexts: Retrieval becomes (relatively) robust against changes in the context and form of the retrieval. If this is true, then an interesting research question is to establish the degree to which various encoding operations render retrieval contextually robust. Certain operations may make the retrieval highly dependent on the recapitulation of the context, others may make retrieval independent of context. Such differences in contextual robustness may exist quite apart from the level of performance: For some encoding operations, retrieval may be uniformly low, regardless of the cuing conditions for retrieval, or it may be uniformly high; for other operations, it may be high if certain retrieval conditions hold, but poor otherwise.

The second possibility is that deep processing boosts retrieval, but such boosting is highly dependent on retrieval context so that levels effects are greatly reduced if retrieval conditions differ too markedly from what is appropriate relative to the encoding. Put differently, transfer appropriate processing is correct; but if the form of processing is deep, the increment in performance when it is also appropriate is greater than the increment for shallow processing, even when it too is paired with appropriate retrieval conditions. Fisher and Craik's (1977) finding that semantic processing tested with semantic cues yields higher recall than rhyme processing tested with rhyme cues supports this second possibility, although as suggested above, some deep encoding operations may serve to make retrieval relatively context independent.

CONCLUDING COMMENTS

The Complementary and Interactive Roles of Data and Theory

Much of the debate surrounding levels of processing might have been more productive had our original formulation made a more explicit statement about the relationship between data and theory. A common misunderstanding of our paper was to interpret it in a narrow hypothetico-deductive tradition — a theory that could immediately be subjected to crucial experimental test. This tradition of theory construction is still strongly entrenched in many areas of psychology, bringing with it the view that the only role of theory is to generate potentially falsifiable hypotheses, and the only role of data is to evaluate those hypotheses. This view is probably mistaken in any area of science, but in the cognitive sciences, where theory development is so rudimentary, it can have devastating consequences. A narrow focus on theory testing often generates data that have little archival value — data that have no significance apart from the theory they were designed to test. Once the theory is discarded, the data, rather than serving as stepping stones to a more refined theory, are also discarded. Rather than building a cumulative data base, theories often lead to experimental findings that are as ephemeral as the theory that inspired them. In our view, at this stage in the development of our understanding of memory, the major role of theoretical formulations is to guide the data gathering process. Data serve to evaluate theories and thereby determine their lifespan; but equally, theoretical ideas influence the useful lifespan of data in the sense that they help determine whether or not those data will take their place in a cumulative data base that, in turn, will guide and constrain subsequent theory construction. Data are rarely gathered in a theoretical vacuum; but granted that in the area of memory all current theories are far from the ultimate truth, the one thing that must be demanded of our theories is that they leave in their trail data that will serve as stepping stones to a better theory.

The Influence of the Levels of Processing Framework

In his influential critical article, Baddeley (1978) comments that "the concept of levels of processing appears to be rapidly approaching obsolescence" (p. 146). On the other hand, White (1983), surveying publications in cognitive psychology and their citation rates, comments that the Craik and Lockhart article "undoubtedly has had the greatest influence of any single contribution published in the 1970's" (p. 426). But numbers of citations do not tell the whole story. They could reflect the diligent activity of a small coterie of zealots working and reworking the same narrow set of issues. In fact, the citations reflect quite the opposite trend, one that can be best described as a spread of effect. Scanning the literature reveals that the impact of the ideas has moved beyond the narrow confines of traditional memory research. Applications of the framework are to be found over an amazingly wide range of topics. There are pharmacological studies (e.g., Malone, Kershner, & Siegel, 1988), psychophysiological studies (e.g., Sanquist, Rohrbaugh, Syndulko, & Lindsley, 1980), studies of hemispheric function (e.g., Leong, Wong, Wong, & Hiscock, 1985), studies of selective attention (e.g., Tsuzuki, 1986) and of cognitive deficits (e.g., Boyd & Ellis, 1986). There are also applications to reading and prose comprehension (e.g., Friederici, 1985), to educational psychology (e.g., Watkins, 1983), child development (e.g., Owings & Baumeister, 1979), and time perception (e.g., Arlin, 1986). We will not attempt to review these many applications, but the above list of areas (which is by no means exhaustive) gives some idea of how far our ideas have spread from their original highly circumscribed context and how wide has been their influence. If our primary goal was to provide a heuristic framework that would stimulate the gathering and interpretation of new data, then we certainly seem to have succeeded.

A wide influence is one thing; whether or not it is thought to be a good thing is quite another matter. We will leave that global evaluation to future and less biased commentators, but a few concluding observations should be made. It will be clear from this review that many of our ideas have changed since 1972. Our undertstanding of what constitutes a deep level of processing and of the components of such processing have undergone considerable change from our original simplistic formulation. Similarly, research has led us to change our ideas about the retention of sensory features. It would surely be alarming if after some 17 years and several thousand published experiments, our views were unaltered. Not only would such obdurateness be alarming, it would stand in absolute contradiction to the basic spirit of our 1972 paper. The function of a heuristic framework for research is not to provide the foundation for a career spent defending a fixed set of ideas, but to stimulate change and development through experimentation and refinement of theory.

Despite this change and development, many of our original ideas have survived with their value confirmed by subsequent research. It is now generally accepted that memory performance is directly and strongly linked to the nature of the processing underlying the original experience and that an adequate theory of memory will have to include an analysis of those processing operations. The earlier efforts to understand memory in terms of the properties of static structures have been largely abandoned, and few would disagree that to have any claim to completeness a theory of memory must be integrated into a general theory of cognitive functioning. All current theories of memory are imperfect and incomplete. The major goal of levels of processing as a research framework was to promote a climate of empirical research and theory development in which specific theories, when revised or abandoned, would leave behind them a cumulative data base that would serve as a foundation for a better theory. It is our view that in serving this function, levels of processing can claim substantial success.

References

Anderson, J.R. (1976.) *Language, memory and thought*. Hillsdale, NJ: Erlbaum.

Anderson, J.R. (1983). *The architecture of cognition*. Cambridge, MA: Harvard University Press.

Anderson, J.R., & Reder, L.M. (1979). An elaborative processing explanation of levels of processing. In L.S. Cermak & F.I.M. Craik (Eds.), *Levels of processing in human memory* (pp. 385-404). Hillsdale, NJ: Erlbaum.

Arlin, M. (1986). The effects of quantity and depth of processing on children's time perception. *Journal of Experimental Child Psychology, 42*, 84-98.

Atkinson, R.C., & Shiffrin, R.M. (1968). 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, pp. 89-125). New York: Academic Press.

Ausubel, D.P. (1962). A subsumption theory of meaningful verbal learning and retention. *Journal of General Psychology, 66*, 213-244.

Baddeley, A.D. (1978). The trouble with levels: A re-examination of Craik and Lockhart's framework for memory research. *Psychological Review, 85*, 139-152.

Baddeley, A. D. (1986). *Working memory*. Oxford: Oxford University Press.

Baddeley, A.D., & Hitch, G.J. (1974). Working memory. In G.H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 8, pp. 47-90). New York: Academic Press.

Bellezza, F.S. (1984). The self as a mnemonic device: The role of internal cues. *Journal of Personality and Social Psychology, 47*, 506-517.

Boyd, B.D., & Ellis, N.R. (1986). Levels of processing and memory in mentally retarded and nonretarded persons. *Intelligence, 10*, 1-8.

Bransford, J.D., Franks, J.J., Morris, C.D., & Stein, B.S. (1979). Some general constraints on learning and memory research. In L.S. Cermak & F.I.M. Craik (Eds.), *Levels of processing in human memory* (pp. 331-354). Hillsdale, NJ: Erlbaum.

Brown, R., & Kulik, J. (1977). Flashbulb memories. *Cognition, 5*, 73-99.

Chase, W.G., & Ericsson, K. A. (1981). Skilled memory. In J.R. Anderson (Ed.), *Cognitive skills and their acquisition*. Hillsdale, NJ: Erlbaum.

Chase, W.G., & Ericsson, K.A. (1982). Skill and working memory. In G.H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 16). New York: Academic Press.

Chi, M.T.H., Glaser, R., & Farr, M.J. (Eds.). (1988). *The nature of expertise*. Hillsdale, NJ: Erlbaum.

Conway, M.A., & Gathercole, S.E. (1987). Modality and long-term memory. *Journal of Memory and Language*, *26*, 341–361.

Craik, F.I.M. (1977). Depth of processing in recall and recognition. In S. Dornic (Ed.), *Attention and performance VI* (pp. 679–697). Hillsdale, NJ: Erlbaum.

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

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

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

Daneman, M., & Tardif, T. (1987). Working memory and reading skill re-examined. In M. Coltheart (Ed.) *Attention and performance XII*. Hillsdale, NJ: Erlbaum.

Elmes, D.G., & Bjork, R.A. (1975). The interaction of encoding and rehearsal processes in the recall of repeated and nonrepeated items. *Journal of Verbal Learning and Verbal Behavior*, *14*, 30–42.

Eysenck, M.W. (1978). Levels of processing: A critique. *British Journal of Psychology*, *68*, 157–169.

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

Frase, L.T., & Kamman, R. (1974). Effects of search criterion upon unanticipated free recall of categorically related words. *Memory and Cognition*, *2*, 181–184.

Friederici, A.D. (1985). Levels of processing and vocabulary types: Evidence from on-line comprehension in normals and agrammatics. *Cognition*, *19*, 133–166.

Geiselman, R.E., & Bjork, R.A. (1980). Primary versus secondary rehearsal in imagined voices: Differential effects on recognition. *Cognitive Psychology*, *12*, 188–205.

Greene, R.L. (1987). Effects of maintenance rehearsal on human memory. *Psychological Bulletin*, *102*, 403–413.

Hashtroudi, S. (1983). Type of semantic elaboration and recall. *Memory and Cognition*, *11*, 476–484.

Intraub, H., & Nicklos, S. (1985). Levels of processing and picture memory: The physical superiority effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *11*, 284–298.

Jacoby, L.L., & Craik, F.I.M. (1979). 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.), *Levels of processing in human memory* (pp. 1–22). Hillsdale, NJ: Erlbaum.

James, W. (1983). *Talks to teachers on psychology and to students on some of life's ideals*. Cambridge, MA: Harvard University Press.

Jenkins, J.J. (1979). Four points to remember: A tetrahedral model of memory experiments. In L.S. Cermak & F.I.M. Craik (Eds.), *Levels of processing in human memory* (pp. 429–446). Hillsdale, NJ: Erlbaum.

Johnson-Laird, P.N., Gibbs, G., & de Mowbray, J. (1978). Meaning, amount of processing, and memory for words. *Memory and Cognition*, *6*, 372—375.

Kahneman, D. (1973). *Attention and effort*. Englewood Cliffs, NJ: Prentice Hall.

Keenan, J.M., MacWhinney, B., & Mayhew, D. (1977). Pragmatics in memory: A study of natural conversation. *Journal of Verbal Learning and Verbal Behavior*, *16*, 549–560.

Kirsner, K., & Dunn, J. (1985). The perceptual record: A common factor in repetition priming and attribute retention? In M. Posner & O. Marin (Eds.), *Attention and performance XI*. Hillsdale, NJ: Erlbaum.

Klein, K., & Saltz, E. (1976). Specifying the mechanisms in a levels-of-processing approach to memory. *Journal of Experimental Psychology: Human Learning and Memory*, *2*, 671–679.

Klein, S.B., & Kihlstrom, J.F. (1986). Elaboration, organization, and the self-reference effect in memory. *Journal of Experimental Psychology: General*, *115*, 26–38.

Kolers, P.A. (1979). A pattern analyzing basis of recognition. In L.S. Cermak & F.I.M. Craik (Eds.), *Levels of processing in human memory* (pp. 363–384). Hillsdale, NJ: Erlbaum.

Kolers, P.A., & Roediger, H.L., III. (1984). Procedures of mind. *Journal of Verbal Learning and Verbal Behavior*, *23*, 425–449.

Leong, C.K., Wong, S., Wong, A., & Hiscock, M. (1985). Differential cerebral involvement in perceiving Chinese characters: Levels of processing approach. *Brain and Language*, *26*, 131–145.

Lockhart, R.S. (1978). Method and theory in the study of human memory. In J.P. Sutcliffe (Ed.), *Conceptual analysis and method in psychology*. Sydney: Sydney University Press.

Lockhart, R.S., & Craik, F.I.M. (1978). Levels of processing: A reply to Eysenck. *British Journal of Psychology*, *69*, 171–175.

Lockhart, R.S., Craik, F.I.M., & Jacoby, L. (1976). Depth of processing, recognition and recall. In J. Brown (Ed.), *Recall and recognition*. New York: Wiley.

Malone, M.A., Kershner, J.R., & Siegel, L. (1988). The effects of methylphenidate on levels of processing and laterality in children with attention deficit disorder. *Journal of Abnormal Child Psychology*, *16*, 379–395.

Mandler, G. (1979). Organization and repetition: An extension of organizational principles with special reference to rote learning. In L.-G. Nilsson (Ed.), *Perspectives on memory research*. Hillsdale, NJ: Erlbaum.

Monsell, S. (1984). Components of working memory underlying verbal skills: A "distributed capacities" view. In H. Bouma & D.G. Bouwhuis (Eds.), *Attention and performance X* (pp. 327–350). Hillsdale, NJ: Erlbaum.

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

Moscovitch, M., & Craik, F.I.M. (1976). Depth of processing, retrieval cues, and uniqueness of encoding as factors in recall. *Journal of Verbal Learning and Verbal Behavior*, *15*, 447–458.

Murdock, B.B., Jr. (1960). The distinctiveness of stimuli. *Psychological Review*, *67*, 16–31.

Murdock, B.B., Jr. (1974). *Human memory: Theory and data*. Potomac, MD: Erlbaum.

Naveh Benjamin, M., & Jonides, J. (1984). Cognitive load and maintenance rehearsal. *Journal of Verbal Learning and Verbal Behavior*, *23*, 494–507.

Nelson, D.L. (1979). Remembering pictures and words: Appearance, significance, and name. In L.S. Cermak & F.I.M. Craik (Eds.), *Levels of processing in human memory* (pp. 45–76). Hillsdale, NJ: Erlbaum.

Nelson, T.O. (1977). Repetition and levels of processing. *Journal of Verbal Learning and Verbal Behavior*, *16*, 151–171.

Owings, R.A., & Baumeister, A.A. (1979). Levels of processing, encoding strategies and memory development. *Journal of Experimental Child Psychology*, *28*, 100–118.

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

Paivio, A. (1986). *Mental representations: A dual coding approach*. New York: Oxford University Press.

Roediger, H.L., III, Weldon, M.S., & Challis, B.H. (1989). Exploring dissociations between implicit and explicit measures of retention: A processing account. In H.L. Roediger, III & F.I.M. Craik (Eds.), *Varieties of memory and consciousness*. Hillsdale, NJ: Erlbaum.

Rogers, T.B., Kuiper, N.A., & Kirker, W.S. (1977). Self reference and the encoding of personal information. *Journal of Personality and Social Psychology*, *35*, 677–688.

Ross, B.H. (1981). The more, the better? Number of decisions as a determinant of memorability. *Memory and Cognition*, *9*, 23–33.

Sanquist, T.F., Rohrbaugh, J.W., Syndulko, K., & Lindsley, D.B. (1980). Electrocortical signs of levels of processing: Perceptual analysis and recognition memory. *Psychophysiology*, *17*, 568–576.

Shiffrin, R.M. (1976). Capacity limitations in information processing, attention, and memory. In W.K. Estes (Ed.), *Handbook of learning and cognitive processes*. Hillsdale, NJ: Erlbaum.

Stein, B.S. (1978). Depth of processing reexamined: The effects of precision of encoding and test appropriateness. *Journal of Verbal Learning and Verbal Behavior*, *17*, 165–174.

Sternberg, S. (1966). High speed scanning in human memory. *Science*, *153*, 652–654.

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

Treisman, A. (1964). Selective attention in man. *British Medical Bulletin*, *20*, 12–16.

Treisman, A. (1979). The psychological reality of levels of processing. In L.S. Cermak & F.I.M. Craik (Eds.), *Levels of processing in human memory* (pp. 301-330). Hillsdale, NJ: Erlbaum.

Tsuzuki, T. (1986). Effects of levels of-processing on retention of target and nontarget words in a selective attention task. *Japanese Journal of Psychology*, *56*, 328–334.

Tulving, E. (1983). *Elements of episodic memory*. New York: Oxford University Press.

Tyler, S.W., Hertel, P.T., McCallum, M.C., & Ellis, H.C. (1979). Cognitive effort and memory. *Journal of Experimental Psychology: Human Learning and Memory*, *5*, 607–617.

Watkins, D. (1983). Depth of processing and the quality of learning outcomes. *Instructional Science*, *12*, 49–58.

White, M.J. (1983). Prominent publications in cognitive psychology. *Memory and Cognition*, *11*, 423–427.

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