Relativity of Remembering: Why the Laws of Memory Vanished

Henry L. Roediger, III

Department of Psychology, Washington University in St. Louis, Missouri 63130;
email: roediger@wustl.edu

Annu. Rev. Psychol. 2008. 59:225–54

First published online as a Review in Advance on September 12, 2007

The *Annual Review of Psychology* is online at http://psych.annualreviews.org

This article's doi:
10.1146/annurev.psych.57.102904.190139

0066-4308/08/0203-0225$20.00

Key Words

learning, recall, recognition, retention, transfer

Abstract

For 120 years, cognitive psychologists have sought general laws of learning and memory. In this review I conclude that none has stood the test of time. No empirical law withstands manipulation across the four sets of factors that Jenkins (1979) identified as critical to memory experiments: types of subjects, kinds of events to be remembered, manipulation of encoding conditions, and variations in test conditions. Another factor affecting many phenomena is whether a manipulation of conditions occurs in randomized, within-subjects designs rather than between-subjects (or within-subject, blocked) designs. The fact that simple laws do not hold reveals the complex, interactive nature of memory phenomena. Nonetheless, the science of memory is robust, with most findings easily replicated under the same conditions as originally used, but when other variables are manipulated, effects may disappear or reverse. These same points are probably true of psychological research in most, if not all, domains.

Contents

INTRODUCTION

The *Annual Review of Psychology* is 58 years old. The intellectual heritage of this review on human memory began with Arthur W. Melton's article on *Learning* in the first volume in 1950. The topic of learning covered a wide band of research, and memory research was considered a subfield. Depending on the author of the learning chapters in the early volumes of the *Annual Review of Psychology*, human memory received either considerable attention (e.g., when B.J. Underwood was the author) or practically none (e.g., when H.F. Harlow was the author). The topic of learning continued to appear for the first nine years, with a review each year, but then became more specialized (e.g., perceptual learning). In 1968, the word "memory" appeared for the first time in a chapter title, when G. Keppel wrote a review on "Verbal Learning and Memory" (Keppel 1968). Two years later, E. Tulving & S.A. Madigan (1970) turned the title around for emphasis on memory, but they began their "Memory and Verbal Learning" review by commenting, "The domain of psychological research known today under the bifurcated title of verbal learning and memory has suffered through a long and dull history" (p. 437).

During the 1960s through much of the 1980s, authors took it upon themselves to cover the whole field, albeit selectively, in their reviews. Often they reported the exact period of months being covered. However, at some point along the way, that strategy became hopeless due to the explosion in research, and writers of these reviews (at least in the field of human memory) wisely concentrated on one or a few topics [e.g., Richardson-Klavehn & Bjork (1988) wrote on the study of implicit or indirect tests, in a heavily cited review, and ignored the rest of the field]. I have decided to follow in this tradition, and so my review does not pretend to cover the field or even to cover much recent research. Even making the bold assumption that it was once possible to knowledgeably survey the whole field of learning and memory, those days are long past. In some ways, my review would be more appropriate for a Centennial Review of Psychology.

The focus of this review is on relativity of remembering, using "remembering" in its generic sense of performance on a memory test rather than in its specialized sense developed by Tulving (1985a,b). The tradition of research considered here is that of the experimental/cognitive psychologist who, starting with Ebbinghaus (1885/1964), believes that incisive experimentation and judicious (but not expansive) theorizing is a powerful road to seeking truth about human memory. Of course, many other approaches to the topic of memory are perfectly valid in their own realm: cellular and molecular neurobiology studies with emphasis on synaptic change, long-term potentiation, and many other topics; approaches from animal learning and behavior that emphasize conditioning; ethological studies of foraging and homing, among other topics; neuropsychological studies of patients with various defects in learning and memory; behavioral neuroscience approaches studying animal models of memory; cognitive neuroscience approaches using various imaging techniques; social psychological approaches concerned with social remembering; and broad approaches from history and sociology on the study of collective memory and how it shapes personal identity (Wertsch 2002). All these fields have their own traditions for the study of learning and memory. We may someday hope for a unified science of memory, but that day is not yet at hand (see Roediger et al. 2007 for a start in this direction).

For purposes of this review, I follow in the tradition started by Ebbinghaus in assuming that scientists can wrest hard-won truths about memory from Mother Nature through careful, thorough (perhaps even compulsive) experimentation on adult human subjects. My review focuses further on long-term memory and not so much on short-term or working memory, a topic covered well in the review by Jonides et al. (2008).

LAWS OF MEMORY

The theme of this review—the relativity of remembering—contests a major assumption that shaped beliefs of the pioneers of our field, i.e., that there are general laws of learning and memory. Early researchers pronounced several laws of memory, and other generalizations and regularities have been proposed over the years. In *Animal Intelligence: Experimental Studies* (1911), E.L. Thorndike wrote, with breathtaking authority, "Two laws explain all learning." (They were the law of effect and the law of exercise.)

Dated from 1885, the experimental psychology tradition of the study of learning and memory is 123 years old. We have learned many fascinating facts about memory in this time, ones that would astonish Ebbinghaus. Yet the thesis of this review is that one central lesson to be gained from thousands of experimental studies is that no general laws of memory exist. All statements about memory must be qualified.

By a "law," I do not necessarily mean anything too grand, either, like Newton's three laws. Rather, I use the term "law" simply to mean an empirical regularity, an established functional relation, one that holds widely (ideally, universally) across manipulation of other variables. In an excellent article on "One Hundred Years of Laws in Psychology," Teigen (2002) proposed five criteria for laws in science: validity (the law should be a well-established regularity, with deterministic laws tolerating no exception and probabilistic laws having few); universality (the law should be independent of place and time); priority (laws take precedence over observations, such that when observations seem at odds with the law we tend to doubt the observations); explanatory power (the law is connected to other general principles); and autonomy (the law should be self-contained, able to be encapsulated in a brief description, preferably mathematical).

Many early researchers hoped that psychology would be like physics and produce general laws of behavior, to rival (say) Kepler's laws of planetary motion. Kepler was a devotee of numerology, and sought simple laws that would unite the whole solar system; he even sought a relation between musical harmonies and planetary motion. It took him 17 years of hard work to produce his famous third law. Commenting on Kepler's quest, Holton & Brush (1985) stated, "This conviction that a simple rule exists, so strong that it seems to us like an obsession, was partly a remnant of his earlier numerological preoccupations, and partly the good instinct of genius for the right thing to work on. But it also indicates a deep undercurrent running through the whole history of science: the belief in the simplicity and uniformity of nature" (p. 44). The same tendency pervades any scientific field, but often the facts discomfort the belief, and calls for parsimony may be misguided if the parsimonious law is only about a circumscribed set of behaviors (Battig 1962).

Teigen (2002) examined 1.4 million abstracts in the PsychLit database from 1900 to 1999 for the occurrence of the word "law." His results reveal a striking regularity in that the term "law" has become much less frequent over this time, dropping from 266 mentions per 10,000 entries in 1900–1909 to a mere 10 from 1990 to 1999. "In other words, today one must read about 1000 journal articles before encountering a single law" (p. 108). Teigen's article was concerned with all of scientific psychology, but the same trends seem to be true in learning and memory research. The laws of memory have vanished from the scene.

Memory researchers have announced a number of laws of memory over the years. For example, some half dozen laws are provided in McGeoch's (1942) great textbook. Jost's (1897) two laws are still known today (and are considered below), but other "laws" of the day are not even recognizable 65 years later by those on the contemporary scene. Who today can even define the Kjerstad-Robinson law or the Müller-Schumann law of associative inhibition? (The first term refers to the form of the learning curve relating number of responses to its shape; the second essentially refers to the phenomenon of conditioned inhibition—see McGeoch 1942, p. 48 and p. 402, respectively.) As McGeoch (1942) reported in footnotes, the first law has rather immediate boundary conditions, so even as he wrote, the generality (and hence validity) of the law was called into question.

Besides a lack of generality of formal laws, even more commonsensical generalizations, ones that "everybody knows" (for example, that repetition improves memory), are either invalid or at least need to be qualified. No principles emerge that hold across various types of memory test, subject populations, retention intervals, instructional strategies, and so on. Although the relativity of remembering presents an uncomfortable fact for textbook writers and those of us wishing to communicate general principles to the lay public, the great truth of the first 120 years of the empirical study of human memory is captured in the phrase "it depends." Does repetition improve memory? Are spaced presentations better than massed presentations? Does deeper, more meaningful processing during encoding enhance retention relative to less meaningful, superficial analyses? Does generation (or active involvement) with learning materials improve retention relative to passive reading? Does the passage of time lead to forgetting? Do retrieval cues improve retention? If we cast our net of inquiry broadly across the field to consider these questions, the answer (as we shall see) is always, "it depends."

JENKINS' TETRAHEDRAL MODEL OF MEMORY EXPERIMENTS

In 1979, James J. Jenkins proposed a model of memory experiments, shown here (somewhat modified) in **Figure 1**. The ideas were presented in a book chapter, and the chapter by Jenkins was a commentary on other chapters and, indirectly, the whole volume. Furthermore, the chapter was tucked into the back of a very long edited book (Cermak & Craik 1979). The topic of the book was the levels-of-processing framework to learning and memory, which dominated the field in the 1970s (e.g., Craik & Lockhart 1972, Craik & Tulving 1975). Of course, many book chapters get lost in the shuffle, and the chapter by Jenkins was the twentieth of 21 chapters. Commentary types of chapters may get even shorter shrift than standard book chapters. It is perhaps for these reasons that the contribution from Jenkins has not (in my opinion) influenced the field as much as it should have. Although others have voiced somewhat similar ideas, Jenkins' contribution captures truths about memory simply, powerfully, and in just a few pages. His ideas were explicitly not intended as a theory of memory, but rather as a theory of memory experiments and how to

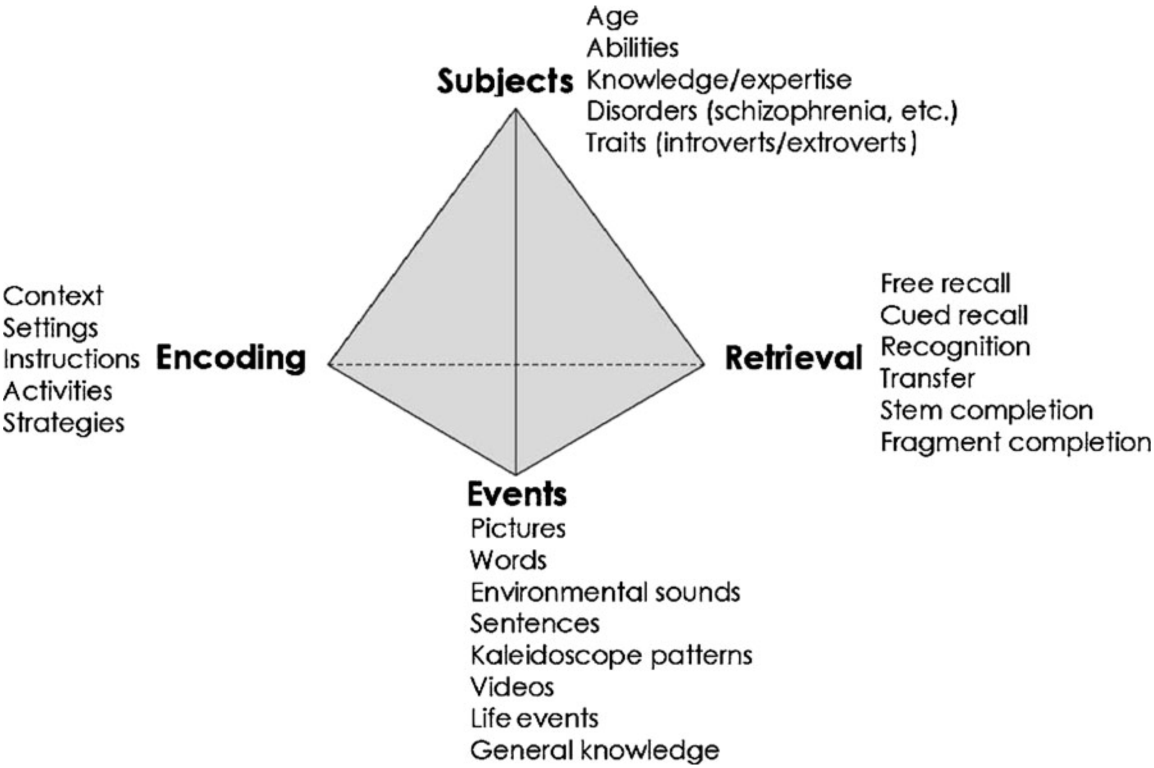

Figure 1

Jenkins' tetrahedral model of memory experiments. Memory experiments can be considered a combination of four factors: subjects, encoding activities, events, and retrieval conditions. In a typical experiment, variations are made in one or two factors and others are randomized or held constant. Jenkins' point was that the effects obtained by manipulation of the independent variable on the dependent variable often depends on the levels of the control variables that are held constant. Adapted from Jenkins (1979).

interpret results from them. Yet his ideas can help lead to an appropriate theory. The general term for the theory might be contextualism, although that term has quite different meanings in various realms of social science and not all of them align with the view of Jenkins (see too his 1974 paper in which similar thoughts were expressed).

Jenkins' (1979) main point was that findings in any experiment about memory are *context sensitive* (italics are his, p. 431) and depend on the level of other variables that were not manipulated. That is, the control variables, those held constant, greatly influence the outcome of the experiment, but researchers usually remain blissfully ignorant of their influence. Of course, it is natural at first to ignore variables that are not manipulated, but researchers should remain mindful of the fact that the particular outcomes in the research may be restricted in that they may only occur with particular settings of the control variables. Additional research in which these variables are manipulated is necessary to determine whether findings are robust (lawful) across a wide range of conditions, but often researchers repeatedly use a paradigm under the same conditions as in the original studies. Jenkins made his point with regard to memory experiments, but it actually holds across empirical research in all fields of psychology. In fact, Battig (1978) proposed broadening Jenkins' model of memory experiments to all psychological experiments.

Figure 1 reveals that experiments performed by cognitive psychologists are composed of a constellation of four basic factors. I have taken the liberty of updating the terminology used by Jenkins, or, less charitably, replacing a few terms with my own. One factor is the people involved in the research: college students (psychology's *Drosophila*), children, older adults, expert bridge players, depressed people, and so on. A second corner refers to events manipulated (or not) during encoding (context and setting, instructions given to subjects, activities or strategies used for learning, and more). For example, subjects can be told before receiving material that there will be a test (intentional learning instructions), or this fact may be omitted (incidental learning instructions) and the later test will be unexpected. A third corner refers to events to be remembered. These can be materials presented in a laboratory setting (word lists, stories, pictures, sentences, a crime scenario), general knowledge questions (What is the capital of Australia?), and events from one's life, among others. The fourth set of factors has to do with retrieval—the way retention is measured—and this factor has been much studied since the Jenkins (1979) chapter was published. A huge number of criterial task have been used to measure retention, from classic tasks like free recall, serial recall, paired associate learning, and various recognition procedures, to newer ones such as primed completion of word fragments or answering general knowledge questions. As with encoding, retrieval can be either intentional (when subjects are asked to remember events) or incidental (when the impact of prior experience is assessed through transfer or priming; see Jacoby 1984). This intentional/incidental contrast during retrieval corresponds to the distinction between explicit and implicit tests of memory (Schacter 1987).

Jenkins called **Figure 1** the "Problem Pyramid" or the "Theorist's Tetrahedron," and he noted, "The memory phenomena that we see depend on what kinds of subjects we study, what kinds of acquisition conditions we provide, what kinds of materials we choose to work with, and what kinds of criterial measures we obtain. Furthermore, the dependences themselves are complex; the variables interact vigorously with one another" (p. 431). Jenkins pointed to many interactions (second order and higher order) in his chapter, and the situation is surely more complex today than it was nearly 30 years ago. Even the separation of the factors above is not clear. For example, "instructions" is listed under encoding, but of course, the effect of the instructions will depend on the type of subjects receiving them and the knowledge the instructions activate.

Jenkins' model of memory experiments points up the possible complexity of the subject matter and the ways in which factors may interact. However, even it is incomplete. In later sections of this review, I consider cases in which all factors in Jenkins' pyramid are held constant and yet manipulation of another factor eliminates or even reverses experimental effects. This fact suggests that more faces might be added to the figure.

A CASE STUDY: THE LEVELS-OF-PROCESSING EFFECT

Following the spirit of the chapter by Jenkins and the volume in which it appeared, I apply Jenkins' scheme to the levels-of-processing effect (Craik & Lockhart 1972), the finding that semantic processing of materials (usually words in a list) leads to better retention on recall and recognition tests than do other types of processing, which channel attention to less meaningful aspects of the materials (e.g., phonemic or orthographic analyses). The levels-of-processing effect is enshrined in virtually all introductory, cognitive, and human memory textbooks, often without much qualification. Many in the field know of some limitations of the effect, but the power of semantic processing is often taken as first principle in the study of memory.

Let us take the famous experiments of Craik & Tulving (1975) as a reference point, because they were so powerful. Holding many classic variables of verbal learning constant (type of material, study time, general instructions to subjects, etc.), Craik and Tulving manipulated only the orienting task and the split-second judgment that subjects made about material during encoding. This orienting-task (or levels-of-processing) manipulation greatly affected the later recognition test, taking performance from nearly chance in one condition to close to perfect in another condition, as discussed below. In the basic procedure (similar in spirit to earlier experiments by Jenkins himself, e.g., Hyde & Jenkins 1969, 1973), students made judgments about 60 words (e.g., BEAR) on three different dimensions: Was the word printed in capital letters? Did it rhyme with chair? Did it refer to an animal? These tasks engender graphemic (case), phonemic (rhyme), and semantic (category) types (or levels) of processing, respectively. In the experiment, the answer to half of the 60 questions was yes, and for half the answer was no. In most of the experiments, the dependent measure was recognition memory.

Table 1 provides results from Experiment 9 (averaging across two replications) in the paper by Craik & Tulving (1975). Two basic effects are seen in **Table 1**, which were replicated several times in their experiments: Semantic processing of words provided better recognition than phonemic processing, which in turn led to better recognition than graphemic processing. In addition, items that required a "yes" answer were generally better recognized than were those given a "no" answer (with two of the three orienting tasks). However, as is also apparent in **Table 1**, the two factors interacted such that the levels-of-processing effect was greater for "yes" than "no" answers.

Table 1 Proportion of words recognized after encoding words in relation to three types of questions (case, rhyme, and category) and their answers (yes or no). Adapted from Craik & Tulving (1975, Table 6)

	Yes	No
Case	0.33	0.33
Rhyme	0.62	0.42
Category	0.86	0.64

Craik & Tulving (1975) reported ten or so experiments (some quite briefly), and their basic results have been replicated countless times when similar conditions have been used. As noted above, the typical levels-of-processing effects are huge, with recognition going from chance (33% under their test situation) in the case/graphemic conditions to 86% in the category/semantic condition when the task required subjects to provide a "yes" response. This experiment and ones like it using free recall measures (e.g., Tresselt & Mayzner 1960, Hyde & Jenkins 1969) are among the most powerful in the experimental study of memory. We still do not have a good theory of why the levels-of-processing effect (as it is called) occurs—Craik & Tulving's (1975) research showed that the original levels-of-processing ideas were wrong—but the fact that meaningful orienting tasks produce better retention than do ones that focus subjects' attention on phonemic or visual features is touted in practically all textbooks today. Roediger & Gallo (2002) discuss some of the mysteries left unexplained by the original levels-of-processing framework.

Surely the powerful effect of orienting tasks on later retention seems a good candidate for a general law of memory—meaningful encoding tasks produce better retention than phonemic or orthographic encoding tasks. In the 1970s, most researchers followed Craik and Tulving's procedures by using words, college student subjects, and recognition or recall tests. As noted, Craik & Tulving (1975) and others (Hyde & Jenkins 1969) did show generality over such variables as presentation rate and orienting instructions (intentional versus incidental learning). However, most variations were rather small ones, in retrospect.

In applying Jenkins' ideas in **Figure 1**, we can also ask if the effect of orienting task (or the levels-of-processing effect) holds across subject groups, materials, and a variety of types of memory test. Considering subject groups first, in research to date, the effect seems quite secure when the standard procedure is used. Besides young adults, the effect has been shown to hold in preschool children (Murphy & Brown 1975), older adults (Craik 1977), and even patients with Korsakoff amnesia (Cermak & Reale 1978). Although exceptions may turn up in the future, so far all subject groups tested show the effect.

Turning to materials, the levels-of-processing effect holds widely across verbal materials (see Lockhart & Craik 1990 for a review). Relatively few studies have used nonverbal materials, but the effect has been obtained with faces (Smith & Winograd 1978) and chess positions (Lane & Robertson 1979). However, in a series of experiments, Intraub & Nicklos (1985) reported an exception: subjects studied pictures and answered questions about physical appearance (Is this horizontal or vertical?) or meaning (Is this edible or inedible?). Subjects were asked to recall the pictures using one- or two-sentence phrases, enough to indicate to the experimenter what picture was being recalled. Surprisingly, later recall was greater following physical encoding than following meaningful encoding. This advantage of physical encoding to meaningful encoding was replicated across six experiments but remains unexplained. In fairness, this is a rather isolated exception, and persistent experimental attention has not been given to it. Still, it points to a possible lack of generality of the levels-of-processing effect across types of material.

Turning to type of criterial test, the evidence is quite mixed. On the positive side, for standard recall and recognition tests, the levels-of-processing effect holds quite well. On the other hand, when testing conditions are broadened, the effect disappears or even reverses. Type of criterial test is today widely acknowledged as a limiting condition of the effect. This is true of both implicit memory tests that are perceptual in nature (e.g., word identification, stem completion, fragment completion) and explicit memory tests that require subjects to access phonemic or orthographic information. Each case is considered in turn.

Jacoby & Dallas (1981) performed essentially the same type of experiment as Craik & Tulving (1975) except that they used two different measures of retention. That is, they manipulated question types to instantiate three levels of processing (case, rhyme, and category) and then gave subjects either a standard recognition test or what would today be called a perceptual implicit test. On the standard recognition test, they replicated Craik and Tulving's results, as can be seen on the left side of **Table 2**. Orienting task greatly affected level of recognition, as did whether the answer to the question was yes or no. However, on their second test, the outcome was quite different. The test they used is perceptual recognition (or word identification). In this test, subjects saw exactly the same sequence of items as on the standard recognition

Table 2 Proportion of words recognized (hit rates) in an explicit test (left side) and priming in identifying words on an implicit word identification test (right side). Priming is the difference in identifying studied relative to nonstudied words. Adapted from Jacoby & Dallas (1981, Experiment 1)

	Proportion recognized		Priming	
	Yes	No	Yes	No
Case	0.51	0.49	0.13	0.16
Rhyme	0.72	0.54	0.17	0.15
Category	0.95	0.78	0.15	0.18

test, but the words were presented very rapidly (about 35 msec per word, on average), and the subjects' task was to attempt to name the words as they whizzed by. The words were presented too fast to permit accurate recognition, so the measure of interest was how well subjects could name previously studied words relative to new words. Previous work had shown that prior study of words increased (or primed) the ability to name them on a word identification test, so the test can be used as an indirect (or implicit) measure of retention. The question of interest is whether orienting tasks would affect priming in word identification as they did in standard recognition and recall tests.

The answer was a resounding no. Shown on the right side of **Table 2** are priming scores derived from Jacoby and Dallas's results in their Experiment 1. The priming score is defined as the probability of correct identification of studied items minus identification of nonstudied items, which was relatively high in this experiment (0.65). Remarkably, the data on the right of **Table 1** show that neither the variable of orienting task nor the answer to the orienting task question (yes or no) systematically affected performance. Priming was about the same for all conditions in word identification under exactly the encoding conditions that had produced gigantic effects in episodic recognition (and, in other experiments, in recall). Yet priming definitely showed the effect of prior study (and hence measured retention), because the priming scores are positive—prior study increased identification. Jacoby & Dallas's (1981) results and many others since then argue that very different processes underlie certain types of explicit and implicit tests.

Graf & Mandler (1984) showed the same generally null results from the same type of encoding manipulation with word stem completion as the criterial test, and Roediger and colleagues (1992) replicated this result and showed the same null effects in primed word fragment completion. The conclusion from these (and many other) studies is that manipulation of orienting tasks produces the levels-of-processing effects on some tests and not on others. In fact, Roediger et al. (1992) reported a situation in which the levels-of-processing effect was reversed, because priming was greater on verbal implicit tests following a graphemic encoding condition (subjects had to imagine the names of pictures and count the ascending and descending letters in the word) than following meaningful processing.

Perhaps this disparity in results between the left and right panels in **Table 2** arose because some researchers used explicit tests and others used implicit tests. This is not the case. Blaxton (1989) and Srinivas & Roediger (1990), among others, obtained the levels-of-processing effect using implicit memory tests (albeit ones conceptual rather than perceptual in nature). The explicit or implicit nature of the test does not determine whether the levels-of-processing effect occurs.

Even before Jacoby & Dallas's (1981) research, Morris et al. (1977) used a test that measured phonemic knowledge (Was a word that rhymed with care on the list?) and showed that prior phonemic encoding led to better performance on this type of criterial test than did prior meaningful encoding. That is, they showed that a reverse levels-of-processing effect can be obtained on explicit tests that ask subjects to consult the type of physical information encoded on "shallow" tasks. Morris et al. (1977) advanced the concept of transfer-appropriate processing, a contextualist idea in line with the arguments of Jenkins, and proposed that types of processing are not inherently deep or shallow (or good or bad) for later retention. Rather, whether types of encoding will enhance later retention depends on the properties of the test and whether information accessed during encoding will transfer to performance on the test. Processing must be appropriate for use on the test for positive transfer to occur. Other researchers reported either the same "reversed" levels-of-processing effect or no effect of orienting tasks on other explicit memory tests (Fisher

& Craik 1977, McDaniel et al. 1978), with the critical variable being the nature of retrieval cues and instructions given during the criterial task (see Jenkins 1979 for further examples).

In short, despite the powerful effects of orienting task judgments (semantic > phonemic > graphemic) observed by Craik & Tulving (1975) and many others, the effect does not constitute a general law because it depends on the nature of the test used (and perhaps other variables as well, such as type of material). If tests that draw on the perceptual record of experiences are used, no effect of orienting task generally occurs because perceptual characteristics of the stimuli are encoded in all the orienting tasks (Roediger et al. 1989b). If the test requires knowledge of the phonemic or graphemic characteristics that were encoded, a phonemic encoding task can produce greater performance than a semantic encoding task (Morris et al. 1977) and so can even a graphemic encoding task (McCabe & Jenkins 1978, as cited in Jenkins 1979, Stein 1978).

Challis et al. 1996 conducted the most systematic study illuminating the relativity of remembering in the levels-of-processing paradigm. In their ambitious experiment, five encoding conditions (manipulated within subjects) were crossed with six memory tests (examined between subjects), with performance in all conditions measured relative to a nonstudied baseline. The to-be-remembered items were words. Subjects simply learned them (intentional learning instructions), made judgments of whether the word could be related to the person (self-judgment), judged whether it referred to a living thing (living judgment), counted the number of syllables (count syllables), or counted the number of letters of a certain type (count letters). The tests were yes/no recognition, free recall, cued recall with semantic cues, cued recall with graphemic cues, or, finally, the two implicit memory tests of answering general knowledge questions and completing word fragments.

The results are presented in **Figure 2** in units of least significant differences relative to baseline performance (to put all tasks on

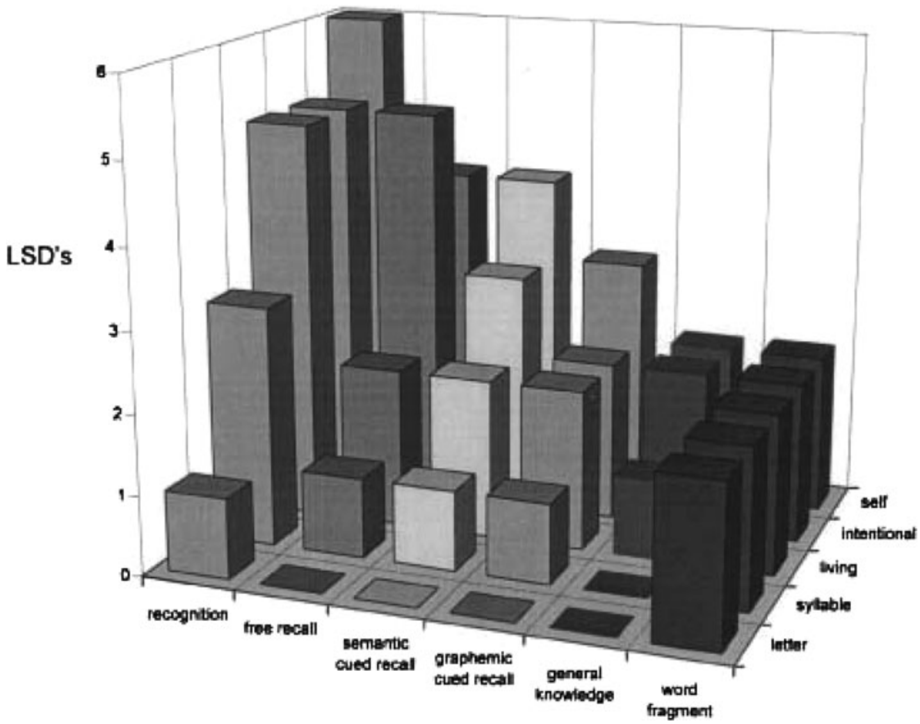

Figure 2

An experiment in which encoding conditions were manipulated within subjects and memory tests were administered between subjects. The results portray a complex interaction between encoding and retrieval conditions, showing that various measures of memory reflect different aspects of performance. Adapted from Challis et al. 1996.

a common scale and so that any bars of different height are significantly different). Just a quick look at the figure shows that the results are complex—some encoding tasks produced better performance on some tests than others; tests differed markedly in revealing effects of past experience. All tests produced equivalent priming on the perceptual implicit memory test of word fragment completion, but the letter-encoding task produced significantly better performance than baseline only on this test and on recognition. The other tests picked up no effect of prior study in this encoding condition.

Consider just the nine bars in the upper left of **Figure 2**, comprising a 3 × 3 encoding/retrieval design. All the tests considered (recognition, free recall, and semantic cued recall) are based primarily on meaning (according to criteria spelled out by Roediger et al. 1989b), yet three different patterns of effect that occurred across encoding conditions (self-referential processing, intentional learning, and living/nonliving judgments) were observed. Recognition was best after self-encoding and equivalent after the other two manipulations. Free recall, on the other hand, was best after intentional learning, next best after self-encoding, and least good (among these three encoding conditions) after making living/nonliving judgments. Finally, for semantic cued recall, yet a different pattern emerged, with self > intentional > living. Thus, even though the three tests are similar in many ways (at least compared with, say, word-fragment completion), three different patterns of results emerged as a function of manipulation of encoding tasks (which themselves were rather similar, at least relative to rhyme or letter encoding tasks).

The point here (luckily) is not to explain the results in **Figure 2** (see Challis et al. 1996 for a valiant attempt), but rather to use the results to point to the relativity of remembering. In most experiments in which a variety of encoding manipulations are compared and contrasted across a variety of tests, the outcome is often just like that in **Figure 2**—a complex interaction (see Kolers & Roediger 1984 and Roediger et al. 1989a for additional examples). When asking if Condition A provides better retention than Condition B, the answer is always "it depends" (on the type of test, or the materials, or the retention interval, and so on).

MEMORY TESTS

No comprehensive history of experimental studies of memory has ever been written (and we can pause briefly to wonder why, given how many researchers have devoted their lives to this field). However, if one were to be written, two dramatic changes that would be noted since 1970 would lie in the proliferation of different methods of testing memory (see Schacter 1987) and the explosion of different kinds of memory that have been postulated (e.g., Tulving 1972, 2007). The two changes are not unrelated, of course.

Here I focus on tests of memory. The field began with researchers measuring recall of presented items in order, as though serial recall were the natural starting place. Ebbinghaus (1885/1964) required himself to recall nonsense syllables in order, although the primary measure he used was trials to criterion (the number of study and test trials required to reach a single perfect recitation). He then measured the savings in the number or percentage of trials to repeat this feat at a later time. Ebbinghaus preferred the savings method because he regarded it as objective and much preferable to the "introspective methods" of recall or recognition. After all, in these latter measures, how could one know if the subject were performing optimally, or perhaps merely constructing plausible answers, or even blindly guessing? Savings methods overcame these problems to permit a more certain measure of retention.

Working at about the same time, Nipher (1876, 1878) presented digits and also measured serial recall, but he scored the number or percentage of correct responses in remembering digits in order, as in a telephone number (not that telephones were in use at that point; Nipher's materials were the mantissas of logarithms). He first reported the serial position effect, among other discoveries (Stigler 1978).

A bit later, Calkins (1894) developed the paired-associate learning technique, and other methods to measure retention quickly followed, primarily free recall (Kirkpatrick 1894) and recognition memory (e.g., Strong 1912). After these early developments, researchers generally stuck to these measures and used them and straightforward variations on them (such as measuring latencies) to ask many questions. Hall's (1971) *Verbal Learning and Retention*, a thorough text from the heyday of the verbal learning tradition, focused almost entirely on serial learning and paired-associate learning, although small sections were also devoted to free recall, recognition, and transfer procedures.

As far as I can tell, little was said in the early days about selection of dependent measures for analysis. The implicit assumption seems to have been that "memory" was a single entity and that most any measure would do, although some might be more convenient and more sensitive than others depending on the question being asked. For example, paired associate learning was appropriate if interest was in learning of single associations, whereas serial learning was more appropriate for learning a chain of associations. Recognition was often thought to be more sensitive than recall because it could detect memory traces of less strength (e.g., see Kausler 1974, p. 8). Free recall measures came to the fore later for intensive study. Deese (1957) introduced the technique of single-trial free recall and began the study of its serial position curve, which fueled much research in the 1960s and 1970s (and even today). Studies of multitrial free recall were introduced to study organizational processes, either subjects' tendencies to adopt organization inherent in materials (category clustering; Bousfield 1953) or organization imposed by the subject on materials selected to have no obvious organization (Mandler 1967, Tulving 1962). On the other hand, single-item (yes/no or free choice) recognition procedures were more appropriate for assessing item-specific knowledge. However, the overarching assumption seemed to be that all these methods were intended to study the same entity, memory. The procedures varied in their ability to elicit different aspects of retention, but of seemingly the same kind of memory.

Even as late as 1979, Jenkins noted that only "a few workers are interested in criterial tasks" (p. 432). He meant "few" in relation to the huge number of researchers interested in the other three prongs shown in **Figure 1** having to do with events (materials), encoding manipulations, and subject populations. However, since 1979, the number of memory measures that are in common use has exploded, fueled in large part by research on implicit or indirect measures of memory but also with a general broadening of scope of the field (e.g., autobiographical memory, prospective memory, among other topics).

The seeds for the cataclysmic changes in memory research had been sewn in the late 1960s, in papers by Warrington & Weiskrantz (1968, 1970), but no one knew it at the time. In their famous experiments, amnesic patients and control subjects studied lists of words and then were given tests of free recall, recognition and two novel tests. One of these tests involved completing words when given their initial letters, a task now called word stem completion. The other involved giving subjects fragmented words, words with parts missing. In both cases, the subjects' task was to complete the stem or fragment with a word. The finding from the experiment was that amnesic patients performed worse than control subjects on the recall and recognition tests, but the two groups performed about equally on the word stem and word fragment tests. Amnesic patients were not amnesic on all tests of memory. Warrington & Weiskrantz (1968) concluded, " . . . retention by partial information is a particularly favourable retrieval method for amnesics but not for controls" (pp. 629–630). In other words, they thought that the stem and fragment tests were simply powerful cues that were accessing the same type of memory as the other tests, the implicit assumption being there was only one form of memory to be accessed. Of course, from this standpoint, one could wonder why recognition performance was not also equivalent: How could a stem or fragment of a word serve as a more powerful cue than the whole word, when the target was the whole word? Warrington & Weiskrantz (1970) attempted to explain this paradox in terms of proactive interference being expressed differently on the different kinds of test (see p. 630).

Some controversy erupted about the replicability of these findings in the 1970s when other researchers did not find performance of amnesic patients equivalent to that of controls on stem and fragment tests (e.g., Squire et al. 1978). However, later research shows that the issue hinges on instructions subjects are given on these stem cued tests. If subjects are told to try to use each stem or fragment as a cue to retrieve a word from the list (in today's parlance, if they are given explicit memory instructions), amnesic patients perform worse than do normal subjects. On the other hand, if they are told to respond with the first word that comes to mind when they see the cue (implicit memory instructions), then equivalent priming is observed between groups (e.g., Graf et al. 1984). In general, on many tasks, when implicit memory instructions are used, researchers find equivalent priming between amnesic patients and control subjects (Shimamura 1986; see too Moscovitch et al. 1993). In their original paper, Warrington & Weiskrantz (1968, p. 974) pointed out, "... in addition to the rapidity and uniformity in learning this task [naming picture fragments], patients find it a much less exacting test than more conventional ones. They treat it more as a 'guessing game' than a formal test of memory."

Jenkins (1979) pointed out that the experimental set during encoding matters, but the literature shows that instructional set during testing greatly matters, too. In explicit memory tests, instructions to subjects require them to enter what Tulving (1983) called the retrieval mode, which is generally a necessary precursor for retrieval from episodic memory. By definition, instructions on implicit tests do not cause subjects to enter the retrieval mode, and often the instructions are designed to discourage such a mental set. Schacter et al. (1989) suggested that one strong form of evidence for distinguishing explicit and implicit forms of retention can be obtained by varying only instructions to subjects and holding all other aspects of the situation constant, a procedure they dubbed the retrieval intentionality criterion. If a researcher obtains different patterns of performance on explicit and implicit forms of a test holding everything else constant, then intentional and incidental (implicit) forms of retention are more secure. Of course, even on implicit tests subjects may sometimes produce an item that reminds them of the retrieval episode, leading to the phenomenon that has been labeled involuntary conscious memory and studied in its own right (e.g., Richardson-Klavehn et al. 1999).

Tulving (1972) proposed the distinction between episodic and semantic memory, and, for a time, the dissociations between performance on explicit memory tests and implicit memory tests were attributed to broad differences between episodic and semantic memory (Tulving 1983, 1985a) or between declarative and procedural memory in other classification schemes (e.g., Squire 1987). However, this broad heuristic did not survive empirical scrutiny, because it could be shown that tasks tapping the same putative system could be dissociated (e.g., Blaxton 1989, Srinivas & Roediger 1990; see Roediger et al. 1989a for a review). Researchers explained the different pattern of performance on certain implicit tests by proposing that performance on word stem or word fragment completion was caused by data-driven or perceptual processing (Roediger & Blaxton 1987), or reflected the perceptual record of experience (Kirsner

& Dunn 1985), or depended upon perceptual memory systems (Tulving & Schacter 1990).

Propelled by the flurry of research on implicit memory in the 1980s and 1990s, the number of tests used and the varieties of memory proposed has grown over the years. Toth (2000) listed 40 different tasks that have been (or in some cases, might be) used to study priming on implicit memory tests (see his Table 16.1 on p. 251). In a recent review, Tulving (2007) asked, semiseriously, "Are there 256 kinds of memory?" The large number of implicit and explicit memory tests tapping many varieties of memory point to a general problem for the field if one seeks general laws of learning and memory. Is it possible to find common laws across these many manifestations of learning and memory? The answer is clearly no.

Of course, one can wonder if the term "memory" has become too broad, encompassing all types of improvements with experience that might better be labeled with other terms (perceptual plasticity, priming, enhanced sensitivity, or whatever). Tulving (1983) has indeed argued that the term "memory" is too broad and is akin to similarly synoptic terms like locomotion. No one seeks general laws of locomotion and, similarly, perhaps no one should seek general laws of memory. Tulving (1985b) wrote, "[N]o profound generalizations can be made about memory as a whole, but general statements about particular kinds of memory may be perfectly possible" (p. 385). This statement is certainly true in principle, but even if a researcher looked for laws in only certain classes of test (e.g., explicit or implicit memory tests, or even free recall and recognition), the answer would be that there are none. No independent variable has even a monotonic effect on a variety of dependent measures of memory; exceptions to any broad generalizations exist. In the next section, I consider some of the most likely candidate variables for putative laws of memory and review why they do not survive close scrutiny.

CANDIDATE VARIABLES FOR LAWFUL RELATIONS

Many variables have been proposed as candidates for possible lawful relations. I consider here repetition, study time, spacing, generation, the mirror effect, imagery and the picture superiority effect, testing, and forgetting as the pre-eminent candidates. In all these cases attention is focused on exceptions to what has been proposed to be a general rule or a law.

Repetition

Perhaps the oldest generalization about learning is that it improves with repetition. Ebbinghaus (1885/1964) pointed to the gradual nature of learning and the effect of repetition in a passage in which he employed a metaphor for memory first used by Aristotle. In commenting on the relation between repetition and savings, he wrote:

> These relations can be described figuratively by speaking of the series as being more or less deeply engraved on some mental substratum. To carry out this figure: as the number of repetitions increases, the series are engraved more and more deeply and indelibly; if the number of repetitions is small, the inscription is but surface deep and only fleeting glimpses of the tracery can be caught; with a somewhat greater number the inscription can, for a time at least, be read at will; as the number of repetitions is still further increased, the deeply cut picture of the series fades out only after longer intervals. (Ebbinghaus 1885/1964, pp. 52–53)

Ebbinghaus provided a clear account, and one can find many confirmations of the general point that repetition on a task provides general improvement. However, many exceptions also exist in which repetitions do not improve performance at all, much less in accord with a particular function. We consider just a few exceptions here. Tulving (1966) exposed subjects to words under incidental learning conditions for six times before he placed the items in lists and required subjects to learn them. A control group performed the same initial task, but the words seen in the first list did not overlap with those later encountered in the to-be-learned list. The prior repetitions of the relevant words did not improve learning a whit, contrary to any account that strength accrues simply as a function of repetition.

Mandler & Pearlstone (1966) had one group of subjects sort 52 unrelated words into 2 to 7 idiosyncratic groups of their own choosing, with the requirement that subjects be able to make two consecutive sorts that were nearly identical. A yoked control group was required to do the same task, except that they had to discover the organization used by one of the subjects in the other group. Not surprisingly, the number of repetitions between the free choice and the yoked groups varied widely—an average of only 3.5 sorting trials in the free choice group compared to 7.5 trials in the yoked group. Nonetheless, when Mandler & Pearlstone (1966) later required both groups to recall the words, they did so equally well. Despite a doubling in the number of repetitions, the yoked group performed no better than the free choice group. However, for both groups the number of categories was strongly related to recall. As Crowder (1976) put it in summarizing these results, "the effect of repetition on recall was not direct; instead, repetition provided the occasion for organization to occur and organization was what supported good recall" (p. 340).

Many other examples could be provided showing that repetition does not always improve performance. Craik & Watkins (1973) reported two experiments that varied the amount of rehearsal (covert repetition) that subjects gave to words in lists. Despite huge variations in the amount of rehearsal across various experimental conditions, the number of rehearsals was unrelated to later recall. Glenberg et al. (1977) confirmed that rote repetition did not affect recall, but they showed that it did have an effect on recognition, an early indication that these measures differ in important ways. Challis & Sidhu (1993) examined priming on implicit memory tests and manipulated the number of presentations before the test from 1 to 16. The amount of priming on a word fragment completion test was about as great from 16 presentations as from 1. In short, repetition does not always affect memory performance; there is no law of repetition.

Repetition also has to do with practice effects, and Newell & Rosenbloom (1981) argued for a "law of practice" that followed a power function. The law of practice is like Thorndike's law of exercise—as people repeatedly practice a task, they get better at it; they become more accurate and faster (Anderson 1982). The practice law is usually considered a law of learning, but of course, traditional concepts of learning and memory are hopelessly intermixed in traditional learning experiments with repeated practice. Unless previous practice sessions are retained, improvements on future trials will not occur. Although the power law of practice does hold over a wide variety of tasks, debate exists as to whether it is a general law. The research reviewed above shows that repetitions can have little impact on learning, contrary to the power law (albeit from different kinds of experiments than those usually considered). In addition, Heathcote et al. (2000) have argued that the power law is misnamed and that an exponential equation fits the data better than a power function. More importantly, Rickard (1997) reported an exception to the power law and argued that the type of function obtained depends on the types of tasks that are practiced.

Study Time

Bugelski (1962) and Cooper & Pantle (1967) offered the total time law of learning and memory. This law stated that, within limits, the probability of recall of an event is a direct result of the amount of study time afforded the event. The kind of evidence adduced for the law was from list learning experiments in which subjects studied items for two 10-second periods, four 5-second periods, or ten 2-second periods. Some results worked out well within this framework, showing equivalent recall with equal amounts of total study time. However many problems exist for the total time law. For one thing, all the results reviewed in the prior section indicating that amount of repetition is not always related to retention also show that the total time processing material is not always a relevant factor. Results reviewed in the next section about distribution and spacing of repetitions also undercut the total time law, because two study periods distributed in time produce better recall than the same amount of time spent in massed study. Thus, evidence accumulated since the 1960s shows that total time spent in study seems to play little role in retention except in circumscribed situations. Writing in 1970, Melton commented that "it seems clear that the Total-Time Law is in deep trouble as an empirical law . . . " (p. 601). Evidence since 1970 further undercuts the "law," which indicates that the claim of a total time law of memory represents another case where the term law was inappropriately applied.

Distribution and Spacing

Ebbinghaus (1885/1964, p. 89) discovered, more or less by accident, that presentations of material that were distributed in time were retained better than presentations presented close together in time. Crowder (1976, pp. 275–276) usefully distinguished among repetition, distribution, and lag (or spacing) effects. Repetition effects refer to any situation in which a repeated event is better retained than an event presented once; distribution effects refer to the case when events distributed in time are better recalled than ones presented back-to-back, or massed; and finally, spacing or lag effects refer to cases in which a systematic increase in retention occurs with the amount of lag or spacing between two events. We have seen that repetition effects do not qualify for laws of memory, but might distribution of repetitions qualify? The answer seems to be no. Although spaced repetitions do lead to greater retention than massed repetitions under many circumstances, numerous experiments have failed to find such an effect.

Melton (1970) introduced one of the most widely used paradigms to study spacing effects, single-trial free recall, and many studies have shown that the more widely spaced two presentations of a word are in a list, the better is recall. However, exceptions to this rule are many in number. For example, in the earlier mentioned studies by Tulving (1966) and Mandler & Pearlstone (1966) that failed to find any evidence of repetition, the presentations of words were widely spaced. Even in Melton's (1970) paper, he spent considerable effort toward explaining why the spacing or lag effect did not always occur. For example, if testing occurs shortly after the second presentation, massed presentation produces greater recall than does spaced presentation in a paired associate paradigm (Balota et al. 1989, Peterson et al. 1963). Once again, although spacing of presentations may aid retention under some conditions, the spacing effect hardly qualifies as a general law of memory.

The experiments cited above generally used single items in long lists as the unit of repetition. In the 1950s, Underwood examined distributed practice effects using lists of material (generally paired associates) as the relevant unit of repetition, and he reviewed this work in "Ten Years of Massed Practice on Distributed Practice" (Underwood 1961). At the outset of the review, he commented, "The primary empirical goal at the time this series of studies was initiated was a straightforward one, namely, to determine the range of conditions and material within which distributed practice facilitated learning or retention. The fact that 10 years have passed since this goal was established indicates that it has proven to be an elusive objective to obtain. Indeed, no implication should be drawn from the present paper that the goal has now been reached; the pursuit continues" (Underwood 1961, p. 229).

Underwood's (1961) review showed that distributed practice effects occurred only under rather narrow conditions in his experiments and others like them, especially ones involving response interference. It is probably for this reason that most researchers have examined repetition, distribution, and spacing effects of single items in long lists rather than with larger units of material as the unit of repetition. Still, Underwood's (1961) research undercuts the idea of spacing effects generalizing across tasks.

Jost's (1897) two laws announced more than 100 years ago have to do with repetition and distribution effects. As stated by Alin (1997), the laws are, "1) Given two associations of the same strength, but of different ages, the older one will get a greater value on a new repetition; 2) Given two associations of the same strength, but of different ages, the older one will fall off less rapidly in a given length of time" (p. 2). The first law represents another statement of distribution and spacing effects, and Jost used Ebbinghaus's results to support the law (and he replicated those results). However, to the extent that distribution effects sometimes disappear (e.g., Underwood 1961) or even reverse at short retention intervals (Balota et al. 1989), exceptions to the law exist. Similarly, the second law depends on forgetting having a negatively accelerated function. To the extent that exceptions are obtained (in consolidation or hypermnesia experiments, as reviewed below), that law is limited, too.

The material cited in this section is not intended to discredit the generalization that distribution and spacing often obtain and have many general properties when they are observed. Rather, the point is that many experiments have failed to find such distribution and spacing effects and the precise boundary conditions for spacing effects have yet to be determined. Cepeda et al. (2006) provided a large-scale review that shows considerable consistency in certain types of experiments. However, they excluded studies in which the unit of analysis was greater than the single repeated item in a larger set of items to be recalled, which often do not show distribution effects (Underwood 1961).

Generation Effects

Active learning seems to be better than more passive learning in many situations. The generation effect refers to the fact that when people have to generate information they retain it better than if they read it. Jacoby (1978) showed this effect using simple materials such as "foot–sh_," where subjects in the generate condition had to name the related item (shoe, in this example), whereas subjects in the read condition were required to read it ("foot–shoe"). Recall of the target items such as "shoe" was generally greater following generation than reading. Slamecka & Graf (1978) showed this effect with other types of materials, such as telling subjects to generate opposites and then giving them items such as "hot – ????" Generating "cold" led to better retention than when subjects had previously read a pair such as "hot – cold."

Generation effects are often large and robust in within-subjects designs, but they can also be fragile. For example, when whole lists of items are read or generated in either between-subjects or within-subjects/between-lists designs, the generation effect is eliminated and often reversed; that is, under these conditions, read items are better retained than are generated items (Nairne et al. 1991, Schmidt & Cherry 1989). (I return to the issue of type of design in a section below.) Furthermore, even in conditions in which a robust generation effect is obtained on standard recall and recognition tests, the effect is reversed on perceptual implicit memory tests such as identification of words from brief flashes (perceptual identification; Jacoby 1983) or in word fragment completion (Blaxton 1989, Srinivas & Roediger 1990). Once again, although active learning and generation aid performance in some situations and on some types of memory tests, many exceptions exist.

The Mirror Effect

Glanzer & Adams (1985) noted a regularity in recognition memory they dubbed the mirror effect. When high- and low-frequency words are studied and tested on a yes/no recognition test, hit rates are higher for low- than high-frequency words, but false alarm rates show the opposite pattern (more false alarms occur for high- than low-frequency words). The effect is easily replicated under their conditions. The mirror pattern is puzzling, because it seems incompatible with signal detection theories attributing positive responses on recognition tests to familiarity or strength of representations. If low-frequency words have higher familiarity for studied items (hits), why should they have lower familiarity for nonstudied items (false alarms)?

Theorists have taken the task of explaining the mirror effect seriously and created and modified models of recognition memory to do so (e.g., Murdock 2003). However, Greene (2007) has recently provided a review of the literature that shows the mirror effect is not very general. Many situations exist in which hits and false alarms are positively correlated for types of material. Perhaps most tellingly, when subjects study pseudowords (nonwords that can be easily pronounced, like "flirp") and words, the mirror effect disappears on a recognition test. That is, pseudowords show higher hit rates and higher false alarm rates than do words (e.g., Greene 2003, Hintzman & Curran 1997), rather than showing the mirror pattern. As Greene (2007) comments, "Ironically, the prototypical case of the mirror effect involved comparing high-frequency with low-frequency words. However, if extremely low frequency words are used instead, the mirror pattern is violated" (p. 61).

Imagery and the Picture Superiority Effect

The ancient Greeks discovered that imagery can aid memory, and the Romans taught imagery mnemonics to aid rhetoricians faced with making long speeches (Yates 1966). Many studies have shown that pictures are recalled and recognized better than words, as well as showing that when the referents of concrete words are imagined, they are better remembered than when coded only in verbal form (e.g., Bower 1972, Paivio 1969). In addition, the advice to use bizarre images rather than common images to enable retrieval also has generated empirical support (Webber & Marshall 1978, among others). However, against this backdrop of positive findings, limitations and boundary conditions appear. For example, some imagery findings hinge on the type of experimental design used. When within-subjects (mixed-list) designs are used, for example, robust bizarre imagery effects are obtained (e.g., McDaniel & Einstein 1986). However, when between subjects or (within-subjects, between-list designs) are used, bizarre images often produce no better retention than common images (Collyer et al. 1972, Hauck et al. 1972). As noted in the section on generation effects, the same design issue occurs in that domain: When item types (generate or read) are mixed in within-subjects designs, strong effects are usually seen, but when the same variable is manipulated in blocks or between subjects, the effects disappear.

I noted in discussing the Jenkins tetrahedral model of memory experiments that one could hold constant all variables posited there and still eliminate or reverse experimental effects by manipulating another variable. The variable is type of design. Cognitive psychologists often prefer to study a variable by manipulating it within subjects, for reasons of economy. However, as we have seen with generation and imagery, manipulating the same variable between subjects (or within subjects, but with blocked lists of materials) can cause the effect to disappear or even reverse (see Schmidt 2007 for many examples). This fact suggests that the Jenkins (1979) model might need another face, although perhaps "type of design" might be considered under encoding factors. In terms of generality of effects to conditions outside the lab, people will probably believe that one particular strategy is better than another (say, forming images relative to not forming them) and so would use that technique exclusively rather than intermixing the techniques. However, under such blocked conditions, the technique may not be effective. Curiously, the fact that numerous manipulations at encoding change depending on the nature of the design has received little direct attention in the literature, at least in human memory (see Poulton 1982 for consideration of perceptual and attentional phenomena that vary with design changes). Nairne et al. (1991) proposed an account for the vicissitudes of the generation effect with design changes, and recently McDaniel & Bugg (2007) have broadened this approach to provide a general account to explain why design changes affect performance in many tasks. The main point here is that the effects of seemingly powerful variables such as generation and imagery do not generalize across design changes.

Pictures are better remembered than words (the picture superiority effect). This outcome occurs in between-subjects designs (e.g., Erdelyi & Becker 1974) as well as within-subjects designs. However, the positive effects of imagery are not found on all tests. In verbal implicit tests such as word fragment and word stem completion, words produce much more priming than do pictures (e.g., Rajaram & Roediger 1993, Weldon & Roediger 1987). Although effects of pictures and imagery are often powerful when they occur, they are hardly ubiquitous across types of test.

Testing

Generation effects refer to active processing during learning relative to more passive processing. Testing effects refer to the advantage often conferred on retention if subjects actively retrieve information (e.g., Carrier & Pashler 1992, Gates 1917; see Roediger & Karpicke 2006a for a review). The testing effect can be powerful; retrieving only some of the information and not receiving feedback usually produces better retention than does restudying the whole set of material, although obtaining this effect depends on the delays used in the first test. [If the first test is delayed so long that performance on it is quite poor, then the testing effect will not be obtained (Spitzer 1939).] Although the testing effect can be strong under the right conditions, it is not ubiquitous. For example, Roediger & Karpicke (2006b) had subjects study a prose passage twice or study it once and take a test on it. A final criterial test was given after five minutes, two days, or one week to independent groups. On the nearly immediate test, performance was better following massed studying rather than studying and testing, so in this limited context of immediate testing, cramming (repeated reading) produced better performance. However, after two days or a week, the study-plus-test condition outperformed the repeated-study condition, showing that taking a test (even without feedback) is better for long-term retention than an equivalent time spent in repeated study.

Testing for smaller units of information, such as in paired associates, can also lead to impressive testing effects (e.g., Carrier & Pashler 1992). However, testing of a pair is ineffective in promoting retention on a later test when the first test is given quite soon after studying the pair (Jacoby 1978, Karpicke & Roediger 2007). Testing probably does not help much when retrieval occurs from working memory but only when it occurs with some difficulty from secondary memory.

Testing can have a large influence on performance, as with the other variables considered in this section, but no general law exists in this case, either. As usual, if one were to ask if testing helps promote later retention, the answer is "under certain conditions."

Forgetting

Woodworth (1929) wrote, "The machinery developed in the process of learning is subject to the wasting effects of time" (p. 93). McGeoch (1932) argued that interference, not decay, is the cause of forgetting. The first forgetting curve was produced by Ebbinghaus (1885/1964), and Wixted & Carpenter (2007) argued that he even nailed the equation relating time since learning to performance, a particular logarithmic function that is for all practical purposes equivalent to a power function. Forgetting functions would seem to be universal in the study of learning and memory, so do they perhaps represent a law of memory? They can even be fit precisely by equations, as noted above, although some debate occurs over which function fits best (Rubin & Wenzel 1996, Wixted & Carpenter 2007). However, for our purposes we can note that the contending functions all account for around 98% or better of the variance.

The question then is, are exceptions found to this seemingly universal law of forgetting? The answer is yes. One perhaps trivial exception has to do with continued processing—if people are permitted to rehearse or otherwise continually process small amounts of information, the information can be retrieved nearly perfectly over time. Assuming we discount this situation, but look at short-term retention after distraction, forgetting over time still depends heavily on other variables. Using the Brown-Peterson paradigm, Keppel & Underwood (1962) showed that on the first trial there was hardly any forgetting even over fairly long periods of distraction, at least long in the context of short-term memory experiments (see too Reitman 1974). The amount of proactive interference heavily determines the amount of forgetting over time; the same generalization holds true in long-term retention (Underwood 1957).

One might object, quite rightly, that the forgetting function might still be the same, a power function, but that the intercept is raised and lowered by other variables. That is true, but the situation is worse than that: Sometimes performance actually improves with increasing amounts of time since learning. The phenomenon of hypermnesia (e.g., Erdelyi & Becker 1974) represents one such case. In hypermnesia experiments, subjects study a set of pictures and then take repeated tests over time, usually using free or forced recall. The finding that is often obtained (see Payne 1987 for a review) is improved performance over time. Hypermnesia depends on repeated testing of the same information (Roediger & Payne 1982), whereas in standard forgetting experiments, different subjects (or different items) are tested at various intervals to avoid the "confound" of prior testing. Once again, the difference is in between- versus within-subject manipulations of tests. One might argue that using within-subjects testing to measure changes in performance over time is the wrong way to proceed, due to the influence of earlier tests on later tests (the testing effect). However, in normal life we often find ourselves attempting to recall the same events repeatedly, so ruling out hypermnesia experiments as not representing the proper design for studying forgetting seems rather arbitrary. A commonplace occurrence outside the lab is to fail to retrieve information (a name, a fact) at one point in time only to recover it a bit later. Such reminiscence or memory recoveries have been studied for nearly a century (Ballard 1913). Just as some encoding variables lead to different patterns of effect when manipulated within subjects or between subjects, so might some testing variables.

Roediger & Payne (1982) reported an experiment on hypermnesia that produced a puzzling result that did not depend on repeated testing. They had three separate groups of subjects study 60 pictures and required them to recall the names of the pictures later. After the study phase, subjects were kept busy with instructions and (in two cases, reading an article) for 2, 9, or 16 minutes before commencing with a series of three tests. Of course, performance on the very first test is uncontaminated by any prior testing and so might be expected to show the usual power function for forgetting, because the reading diverted subjects' attention. However, recall of the pictures was 43%, 42%, and 43%, respectively, across the three delays. No forgetting at all occurred. Of course, after studying pictorial material, a verbal task might not be expected to create much interference relative to other tasks, but keep in mind that subjects knew they would have to recall the names of the pictures and so doubtless used verbal encoding strategies during study. Luckily for Roediger & Payne (1982), the experiment was about repeated testing, and the editor did not ask the authors to explain why no forgetting occurred between 2 and 16 minutes on the first test given to the different groups.

Even if hypermnesia experiments are excluded, another huge class of experiments—those demonstrating consolidation—sometimes show improved recall over time with no intervening practice (Dudai 2004, McGaugh 2000). Wixted (2004) noted that researchers in the tradition of human experimental (i.e., cognitive) psychology have rarely incorporated (or even discussed) the issue of consolidation. The reason is that it does not fit with the notion, implicit in most accounts, that encoding happens quickly, immediately after perception. After that, the processes responsible for forgetting (usually assumed to be various types of interference) exert their inexorable influence. However, in consolidation experiments with animals and humans, performance can sometimes be shown to improve over time for relatively long periods (without repeated testing during that time). In the past ten years, much research has been directed at the issue of whether consolidation of certain types of memories occurs during sleep (see Payne et al. 2007 for a review). Several different types of task—visual and auditory discrimination, video games, various motor memory tasks, and even certain episodic memory paradigms—show actual improvements over time during sleep (relative to waking), an effect attributed to enhanced consolidation during sleep. In a typical study, subjects practice some relatively novel skill (e.g., the video game Tetris) either at night or in the morning. They are then tested 8–12 hours later after either a night of sleep or a day of waking activities. The typical pattern is that subjects will show improved performance on the video game after a night of sleep (showing absolute increases in performance), whereas after the equivalent amount of time spent in waking activities, subjects will show no improvement or a drop in performance (Stickgold et al. 2001). Apparently, both REM sleep and slow-wave sleep contribute to these consolidation effects (Stickgold 2005).

Dudai (2004) argued for two types of consolidation, synaptic consolidation (which is relatively quick) and system consolidation (which occurs more gradually). Consolidation is usually measured by less interruption from application of some amnestic agent with time since a learning experience rather than with any absolute increase in performance over time, as in the sleep and learning experiments cited above. Still, the latter type of phenomenon argues that there is no general law that says forgetting always occurs in the time since presentation of information. Some consolidation experiments, as well as hypermnesia experiments, undercut the idea that a universal law of forgetting exists. Once again, whether retention declines or improves over time depends on other factors.

QUALIFICATIONS AND PROVISOS

The purview of this review is studies of long-term memory in the tradition of experimental/cognitive psychologists. However, as noted at the outset, this treatment does not consider many other valid studies of memory (and learning) from other traditions. I mention some possible laws here from other traditions, but am not knowledgeable enough to comment on their current status.

Other traditions within cognitive psychology, such as short-term/working memory and motor learning and memory, have received short shrift here. Some discoveries in these arenas suggest a general principle, at least. For example, is there a fixed capacity to short-term memory at about 3–4 items, as Cowan (2005) maintains? Although not a law, exactly, Cowan's proposal at least suggests an interesting invariance. I must leave this discussion to others.

Teigen's (2002) review noted that nearly all putative laws of psychology date back many years. These include classic laws of learning, too, such as the laws of effect and exercise (Thorndike 1898, 1911). In my review, I rather casually lumped together studies traditionally considered those of learning and those of memory, because separating this boundary is difficult. However, from the tradition of animal learning and behavior studies, we can consider such entries as Premack's law (or principle) of the relativity of reinforcement (Premack 1962) and Herrnstein's (1961) matching law, as well as Thorndike's laws of effect and exercise (although the latter has been widely discredited by some of the results discussed above, among others). Further, in classical conditioning (and the learning of associations more generally), we may consider the principle arising from the Rescorla & Wagner (1972) model stating that a mismatch in the current state of knowledge and new information provided to the system drives learning (Rescorla 1988). (This idea is akin to discussions in cognitive psychology of distinctiveness and novelty in promoting memory.) As I say, I am not competent to judge whether the ideas cited above represent general laws and must leave the task for other people. Seligman (1970) reviewed "laws" then believed about classical conditioning and showed them invalid, but of course, we can all hope that researchers in all these areas will find general laws in the future. The fact that past laws have been undermined does not mean new ones will not be found.

This review raises the issue of whether there are laws of memory, but because psychology has seen a distinct decline in discussion of laws in general, we can ask if there are general laws of psychology. Teigen (2002) noted that most laws still described in psychology texts are those first proposed long ago (e.g., Weber's and Fechner's laws in psychophysics). Again, I leave this discussion to others, but finding general principles of behavior that hold widely across the variables in Jenkins' (1979) tetrahedron may be difficult. Also, Jenkins' (1979) approach does not explicitly take into account such factors as culture, species, or even type of experimental design.

In short, my claims are about a particular arena of study of learning and memory. Perhaps others can defend general laws of learning and memory (or of psychology) in other domains.

A COURSE FOR THE FUTURE

The message of this review should not, in my opinion, be taken as depressing. Although our search for general laws of memory has not succeeded, we have learned a huge amount about human memory and its complexities in the past 125 years. Some have argued that even seeking laws and general principles is misguided. Baddeley (1978, p. 150) argued 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," a point made too by Battig (1978) at about the same time.

In the past 30 years, dozens of measures of retention deemed valid indicators of memory and knowledge have been studied. Although cognitive psychologists rely heavily on a certain restricted class of procedures to study recall and recognition, these are only two of many possible assessments of past learning. Despite the fact that researchers may spend their careers concentrating on one or another measure, consideration of many measures simultaneously can upset any general characterization of "how memory works." Claims from our research must therefore be specific rather than general, but this does not prevent our seeking more general principles at higher levels of analysis.

Research programs that examine the effects of several independent variables across many dependent variables almost always show complex interactions. Such studies reveal dissociations across variables and point to a multiplicity of forms of retention. One focus for the future is how various conditions of learning transfer across various measures of retention. Transfer as a topic of study per se has fallen a bit out of favor, but it remains a fundamental concept for the science of memory (McDaniel 2007). Ideas such as transfer appropriate processing (Morris et al. 1977) and the encoding specificity principle (Tulving & Thomson 1973), among others, help to focus on the complex interactions inherent in memory research (see too Jenkins 1979; Kolers & Roediger 1984; Roediger et al. 1989a,b). However, these broad classes of ideas are just the beginning step to understanding these puzzling facts, and clearly much theoretical work is needed to develop them.

The fact that our science does not provide straightforward laws is perhaps well known. McKeachie (1974) made similar points years ago. After all, hardly any researcher writes about general laws any longer (Teigen 2002). At the most, one sees a passing reference to, say, Jost's second law, in a particular context. The aim of this review has been to remind us of the quest for laws and the difficulty in achieving them. In a sense, this state of affairs is no surprise. E.O. Wilson, in *Consilience* (1998, p. 183), commented that the social sciences "are inherently far more difficult than physics and chemistry, and as a result they, not physics and chemistry, should be called the hard sciences." The reason is that in the social sciences, which deal with the huge number of variables that affect behavior of individuals and groups, pinpointing cause and effect relations for complex behaviors is a daunting challenge. The study of human memory may even seem tractable relative to many other issues. Therefore, confronting the complexity in the field of learning and memory is a strong challenge, but not an insurmountable one.

CONCLUSION

When one reads or hears the claim that "repetition improves retention" or that "generating information enhances recall" or that "forgetting follows a power function" or all other statements of the form that "variable X has a consistent effect on variable Y," one can be certain that the claim is both true and false. It is true in that conditions can be found under which the rule holds (otherwise the claim would not be made), but false in that a skeptic can always say: "Very nice work, but your finding depends on many other conditions. Change those and your effect will go away."

The proponent of a general rule that "X affects Y" is, on analysis, usually claiming, "Our study has shown that variable X has a positive effect on a particular measure of memory in a particular situation holding many other variables constant." If some other perfectly valid measure of retention were used in a slightly different situation, with other variables held constant at different levels, no effect, a different effect, or even the opposite effect can often be shown to obtain. The most fundamental principle of learning and memory, perhaps its only sort of general law, is that in making any generalization about memory one must add that "it depends." Of course, only future research can typically tell whether some finding is widely generalizable or holds under a narrow set of conditions.

Should we be discouraged that laws of memory do not seem to exist? I would argue that the answer is no, and that we should be all the more impressed by our science because of its inherent complexity. The cognitive psychology (and cognitive neuroscience) of memory have led to many wonderful discoveries with new ones occurring yearly (if not monthly). The great majority of the findings are highly replicable, at least under the same or similar conditions as in the original studies. The field has a strong scientific base. We have learned a tremendous amount in the years since Melton (1950) wrote the first *Annual Review of Psychology* article on this general topic. There may be no general laws of memory yet discovered, but there is an exciting, robust, and increasingly detailed scientific base in the study of memory. Further, many of the discoveries can make important applications outside the laboratory, in fields such as law and education. Future research can only profit from acknowledging the inherent complexity of the phenomena of memory.

ACKNOWLEDGMENTS

This review benefited from comments by Yadin Dudai, Gilles Einstein, John Jonides, Kathleen McDermott, James Nairne, Dan Schacter and, especially, Endel Tulving. I thank them all and apologize for some requests that remain unfulfilled. The writing of the review was supported by a Collaborative Activity Grant of the James S. McDonnell Foundation and through grant R305H060080 from the Cognition and Student Learning program of the Institute of Education Sciences, U.S. Department of Education. The opinions expressed are those of the author and do not represent views of the Institute or the U.S. Department of Education.

LITERATURE CITED

Alin LH. 1997. The memory laws of Jost. *Göteborg Psychol. Rep.* 27(1):1–21

Anderson JR. 1982. Acquisition of cognitive skill. *Psychol. Rev.* 89:369–406

Baddeley AD. The trouble with levels: a re-examination of Craik and Lockhart's framework for memory research. *Psychol. Rev.* 85:139–52

Ballard PB. 1913. Oblivescence and reminiscence. *Br. J. Psychol. Monogr. Suppl.* 1:1–82

Balota DA, Duchek JM, Paullin R. 1989. Age-related differences in the impact of spacing, lag, and retention interval. *Psychol. Aging* 4:3–9

Battig WF. 1962. Parsimony in psychology. *Psychol. Rep.* 11:555–72

Battig WF. 1978. *Parsimony or psychology.* Presidential Address, Rocky Mountain Psychol. Assoc., Denver, CO

Blaxton TA. 1989. Investigating dissociations among memory measures: support for a transfer-appropriate processing framework. *J. Exp. Psychol.: Learn. Mem. Cogn.* 15:657–68

Bousfield WA. 1953. The occurrence of clustering in the recall of randomly arranged associates. *J. Gen. Psychol.* 49:229–40

Bower GH. 1972. Mental imagery and associative learning. In *Cognition in Learning and Memory*, ed. LW Gregg, pp. 51–88. New York: Wiley

Bugelski BR. 1962. Presentation time, total time, and mediation in paired associate learning. *J. Exp. Psychol.* 63:409–12

Calkins MW. 1894. Association. Studies from the Harvard Psychology Laboratory. *Psychol. Rev.* 1:476–83

Carrier M, Pashler H. 1992. The influence of retrieval on retention. *Mem. Cogn.* 20:633–42

Cepeda NJ, Pashler H, Vul E, Wixted JT, Rohrer D. 2006. Distributed practice in verbal recall tasks: a review and quantitative synthesis. *Psychol. Bull.* 132:354–80

Cermak LS, Craik FIM, eds. 1979. *Levels of Processing in Human Memory*. Hillsdale, NJ: Erlbaum

Cermak LS, Reale L. 1978. Depth of processing and retention of words by alcoholic Korsakoff patients. *J. Exp. Psychol.: Hum. Learn. Mem.* 4:165–74

Challis BH, Sidhu R. 1993. Dissociative effect of massed repetition on implicit and explicit measures of memory. *J. Exp. Psychol.: Learn. Mem. Cogn.* 19:115–27

Challis BH, Velichkovsky BM, Craik FIM. 1996. Levels-of-processing effects on a variety of memory tasks: new findings and theoretical implications. *Conscious. Cogn.* 5:142–64
Collyer SC, Jonides J, Bevan W. 1972. Images as memory aids: Is bizarreness helpful? *Am. J. Psychol.* 85:31–38
Cooper EH, Pantle AJ. 1967. The total-time hypothesis in verbal learning. *Psychol. Bull.* 68:221–34
Cowan N. 2005. *Working Memory Capacity*. Hove, UK: Psychol. Press
Craik FIM. 1977. Age differences in human memory. In *Handbook of the Psychology of Aging*, ed. JE Birren, W Schaie, pp. 384–420. New York: Van Nostrand
Craik FIM, Lockhart RS. 1972. Levels of processing: a framework for memory research. *J. Verb. Learn. Verb. Behav.* 11:671–84
Craik FIM, Tulving E. 1975. Depth of processing and the retention of words in episodic memory. *J. Exp. Psychol. Gen.* 104:268–94
Craik FIM, Watkins MJ. 1973. The role of rehearsal in short-term memory. *J. Verb. Learn. Verb. Behav.* 12:599–607
Crowder RG. 1976. *Principles of Learning and Memory*. Hillsdale, NJ: Erlbaum
Deese J. 1957. Serial organization in the recall of disconnected items. *Psychol. Rep.* 3:577–82
Dudai Y. 2004. The neurobiology of consolidation, or, how stable is the engram? *Annu. Rev. Psychol.* 55:51–86
Ebbinghaus HE. 1885/1964. *Memory: A Contribution to Experimental Psychology.* Transl. HA Ruger, CE Bussenius, 1913. New York: Dover
Erdelyi MH, Becker J. 1974. Hypermnesia for pictures: incremental memory for pictures but not words in multiple recall trials. *Cogn. Psychol.* 6:159–71
Fisher RP, Craik FIM. 1977. Interaction between encoding and retrieval operations in cued recall. *J. Exp. Psychol.: Hum. Learn. Mem.* 3:701–11
Gates AI. 1917. Recitation as a factor in memorizing. *Arch. Psychol.* 40:104
Glanzer M, Adams JK. 1985. The mirror effect in recognition memory. *Mem. Cogn.* 13:8–20
Glenberg A, Smith SM, Green C. 1977. Type I rehearsal: maintenance and more. *J. Verb. Learn. Verb. Behav.* 16:339–52
Graf P, Mandler G. 1984. Activation makes words more accessible, but not necessarily more retrievable. *J. Verb. Learn. Verb. Behav.* 23:553–68
Graf P, Squire LR, Mandler G. 1984. The information that amnesic patients do not forget. *J. Exp. Psychol.: Learn. Mem. Cogn.* 10:164–78
Greene RL. 2003. Recognition memory for pseudowords. *J. Mem. Lang.* 50:259–67
Greene RL. 2007. Foxes, hedgehogs, and mirror effects: the role of general principles in memory research. In *The Foundations of Remembering: Essays in Honor of Henry L. Roediger, III*, ed. JS Nairne, pp. 53–66. New York: Psychol. Press
Hall JF. 1971. *Verbal Learning and Retention*. Philadelphia, PA: Lippincott
Hauck P, Walsh C, Kroll N. 1976. Visual imagery mnemonics: common vs bizarre mental images. *Bull. Psychon. Soc.* 7:160–62
Heathcote A, Brown S, Mewhort DJK. 2000. The power law repealed: the case for an exponential law of practice. *Psychon. Bull. Rev.* 7:185–207
Herrnstein RJ. 1961. Relative and absolute strength of response as a function of frequency of reinforcement. *J. Exp. Anal. Behav.* 4:267–72
Hintzman DL, Curran T. 1997. Comparing retrieval dynamics in recognition memory and lexical decision. *J. Exp. Psychol.: Gen.* 126:228–47
Holton GJ, Brush SG. 1985. *Introduction to Concepts and Theories in Physical Science*. Princeton, NJ: Princeton Univ. Press

Hyde TS, Jenkins JJ. 1969. Differential effects of incidental tasks on the organization of recall of a list of highly associated words. *J. Exp. Psychol.* 82:472–81

Hyde TS, Jenkins JJ. 1973. Recall for words as a function of semantic, graphic, and syntactic orienting tasks. *J. Verb. Learn. Verb. Behav.* 12:471–80

Intraub H, Nicklos S. 1985. Levels of processing and picture memory: the physical superiority effect. *J. Exp. Psychol.: Learn. Mem. Cogn.* 11:284–98

Jacoby LL. 1978. On interpreting the effects of repetition: solving a problem vs remembering a solution. *J. Verb. Learn. Verb. Behav.* 17:649–67

Jacoby LL. 1983. Remembering the data: analyzing interactive processes in reading. *J. Verb. Learn. Verb. Behav.* 22:485–508

Jacoby LL. 1984. Incidental vs intentional retrieval: remembering and awareness as separate issues. In *Neuropsychology of Memory*, ed. LR Squires, N Butters, pp. 145–56. New York: Guilford

Jacoby LL, Dallas M. 1981. On the relationship between autobiographical memory and perceptual learning. *J. Exp. Psychol.: Gen.* 110:306–40

Jenkins JJ. 1974. Remember that old theory of memory? Well forget it. *Am. Psychol.* 29:785–95

Jenkins JJ. 1979. Four points to remember: a tetrahedral model of memory experiments. In *Levels of Processing in Human Memory*, ed. LS Cermak, FIM Craik, pp. 429–46. Hillsdale, NJ: Erlbaum

Jonides J, Lewis RL, Nee DE, Lustig CA, Berman MG, Moore KS. 2008. The mind and brain of short-term memory. *Annu. Rev. Psychol.* 59:193–224

Jost A. 1897. Die Assoziationsfestigkeit in ihrer Abhängigkeit von der Verteilung der Wiederholungen. *Z. Psychol. Physiol. der Sinnesorg.* 14:436–72

Karpicke JD, Roediger HL. 2007. Expanding retrieval practice promotes short-term retention, but equally spaced retrieval enhances long-term retention. *J. Exp. Psychol.: Learn. Mem. Cogn.* 34:707–19

Kausler DH. 1974. *Psychology of Verbal Learning and Memory*. New York: Academic

Keppel G. 1968. Verbal learning and memory. *Annu. Rev. Psychol.* 19:169–202

Keppel G, Underwood BJ. 1962. Proactive inhibition in short-term retention of single items. *J. Verb. Learn. Verb. Behav.* 1:153–61

Kirkpatrick EA. 1894. An experimental study of memory. *Psychol. Rev.* 1:602–9

Kirsner K, Dunn JC. 1985. The perceptual record: a common factor in repetition priming and attribute retention. In *Mechanisms of Attention: Attention and Performance, XI*, ed. MI Posner, OWM Marrin, pp. 547–56. Hillsdale, NJ: Erlbaum

Kolers PA, Roediger HL III. 1984. Procedures of mind. *J. Verb. Learn. Verb. Behav.* 23:425–49

Lane DM, Robertson L. 1979. The generality of the levels of processing hypothesis: an application to memory for chess positions. *Mem. Cogn.* 7:253–56

Lockhart RS, Craik FIM. 1990. Levels of processing: a retrospective commentary on a framework for memory research. *Can. J. Psychol.* 44:87–112

Mandler G. 1967. Organization and memory. In *The Psychology of Learning and Motivation*, ed. KW Spence, JT Spence, 1:327–72. New York: Academic

Mandler G, Pearlstone Z. 1966. Free and constrained concept learning and subsequent recall. *J. Verb. Learn. Verb. Behav.* 5:126–31

McDaniel MA. 2007. Transfer: rediscovering a central concept. In *Science of Memory: Concepts*, ed. HL Roediger, Y Dudai, SM Fitzpatrick, pp. 267–70. Oxford: Oxford Univ. Press

McDaniel MA, Bugg JM. 2007. Instability in memory phenomena: a common puzzle and a unifying explanation. *Psychon. Bull. Rev.* In press

McDaniel MA, Einstein GO. 1986. Bizarre imagery as an effective memory aid: the importance of distinctiveness. *J. Exp. Psychol.: Learn. Mem. Cogn.* 12:54–65

McDaniel MA, Friedman A, Bourne LE. 1978. Remembering the levels of information in words. *Mem. Cogn.* 6:156–64

McGaugh JL. 2000. Memory: a century of consolidation. *Science* 287:248–51

McGeoch JA. 1932. Forgetting and the law of disuse. *Psychol. Rev.* 39:352–70

McGeoch JA. 1942. *The Psychology of Human Learning*. New York: Longmans, Green

McKeachie WJ. 1974. The decline and fall of the laws of learning. *Educ. Res.* 3:7–11

Melton AW. 1950. Learning. *Annu. Rev. Psychol.* 1:9–30

Melton AW. 1970. The situation with respect to the spacing of repetitions and memory. *J. Verb. Learn. Verb. Behav.* 9:596–606

Morris CD, Bransford JD, Franks JJ. 1977. Levels of processing vs transfer appropriate processing. *J. Verb. Learn. Verb. Behav.* 16:519–33

Moscovitch M, Vriezen E, Goshen-Gottstein Y. 1993. Implicit tests of memory in patients with focal lesions or degenerative brain disorders. In *Handbook of Neuropsychology*, ed. F Boller, J Grafman, pp. 133–73. Amsterdam: Elsevier

Murdock BB. 2003. The mirror effect and the spacing effect. *Psychon. Bull. Rev.* 20:570–88

Murphy MD, Brown AL. 1975. Incidental learning in preschool children as a function of level of cognitive analysis. *J. Exp. Child Psychol.* 19:509–23

Nairne JS, Riegler GL, Serra M. 1991. Dissociative effects of generation on item and order retention. *J. Exp. Psychol.: Learn. Mem. Cogn.* 17:702–9

Newell A, Rosenbloom PS. 1981. Cognitive skills and their acquisition. In *Mechanisms of Skill Acquisition and the Law of Practice*, ed. JR Anderson, pp. 1–51. Hillsdale, NJ: Erlbaum

Nipher FE. 1876. Probability of error in writing a series of numbers. *Am. J. Sci. Arts* 12:79–80

Nipher FE. 1878. On the distribution of errors in numbers written from memory. *Trans. Acad. Sci. St. Louis* 3:ccx–ccxi

Paivio A. 1969. Mental imagery in associative learning and memory. *Psychol. Rev.* 76:241–63

Payne DG. 1987. Hypermnesia and reminiscence in recall: a historical and empirical review. *Psychol. Bull.* 101:5–27

Payne JD, Ellenbogen JM, Walker MP, Stickgold R. 2007. The role of sleep in memory consolidation. In *Cognitive Psychology of Memory. Vol. 2 of Learning and Memory: A Comprehensive Reference*, ed. HL Roediger III, 4 vols. (J. Byrne, ed.). Oxford: Elsevier. In press

Peterson LR, Wampler R, Kirkpatrick M, Saltzman D. 1963. Effect of spacing presentations on retention of a paired-associate over short intervals. *J. Exp. Psychol.* 66:206–9

Poulton EC. 1982. Influential companions: effects of one strategy on another in the within-subjects designs of cognitive psychology. *Psychol. Bull.* 91:673–90

Premack D. 1962. Reversibility of the reinforcement relation. *Science* 136:255–57

Rajaram S, Roediger HL. 1993. Direct comparison of four implicit memory tests. *J. Exp. Psychol.: Learn. Mem. Cogn.* 19:765–76

Reitman JS. 1974. Without surreptitious rehearsal, information in short-term memory decays. *J. Verb. Learn. Verb. Behav.* 13:365–77

Rescorla RA. 1988. Pavlovian conditioning: It's not what you think it is. *Am. Psychol.* 43:151–60

Rescorla RA, Wagner AR. 1972. A theory of Pavlovian conditioning: variations in the effectiveness of reinforcement and nonreinforcement. In *Classical Conditioning II: Current Research and Theory*, ed. AH Black, WF Prokasy, pp. 64–99. New York: Appleton-Century-Crofts

Richardson-Klavehn A, Bjork RA. 1988. Measures of memory. *Annu. Rev. Psychol.* 39:475–543

Richardson-Klavehn A, Gardiner JM, Java RI. 1999. Involuntary conscious memory and the method of opposition. *Memory* 2:1–29

Rickard TC. 1997. Bending the power law: a CMPL theory of strategy shifts and the automatization of cognitive skills. *J. Exp. Psychol.: Gen.* 126:288–311

Roediger HL, Blaxton TA. 1987. Retrieval modes produce dissociations in memory for surface information. In *Memory and Learning: The Ebbinghaus Centennial Conference*, ed. DS Gorfein, RR Hoffman, pp. 349–79. Hillsdale, NJ: Erlbaum

Roediger HL, Dudai Y, Fitzpatrick SM, eds. 2007. *Science of Memory Concepts*. Oxford: Oxford Univ. Press

Roediger HL, Gallo DA. 2002. Levels of processing: some unanswered questions. In *Perspectives on Human Memory and Cognitive Aging: Essays in Honour of Fergus I.M. Craik*, ed. M Naveh-Benjamin, M Moscovitch, HL Roediger, pp. 28–47. Philadelphia, PA: Psychol. Press

Roediger HL, Karpicke JD. 2006a. The power of testing memory: basic research and implications for educational practice. *Perspect. Psychol. Sci.* 1:181–210

Roediger HL, Karpicke JD. 2006b. Test-enhanced learning: taking memory tests improves long-term retention. *Psychol. Sci.* 17:249–55

Roediger HL, Payne DG. 1982. Hypermnesia: the effects of repeated testing. *J. Exp. Psychol.: Learn. Mem. Cogn.* 8:66–72

Roediger HL, Srinivas K, Weldon MS. 1989a. Dissociations between implicit measures of retention. In *Implicit Memory: Theoretical Issues*, ed. S Lewandowsky, JC Dunn, K Kirsner, pp. 67–84. Hillsdale, NJ: Erlbaum

Roediger HL, Weldon MS, Challis BH. 1989b. Explaining dissociations between implicit and explicit measures of retention: a processing account. In *Varieties of Memory and Consciousness: Essays in Honour of Endel Tulving*, ed. HL Roediger, FIM Craik, pp. 3–39. Hillsdale, NJ: Erlbaum

Roediger HL, Weldon MS, Stadler ML, Riegler GL. 1992. Direct comparison of two implicit memory tests: word fragment and word stem completion. *J. Exp. Psychol.: Learn. Mem. Cogn.* 18:1251–69

Rubin DC, Wenzel AE. 1996. One hundred years of forgetting: a quantitative description of retention. *Psychol. Rev.* 103:734–60

Schacter DL. 1987. Implicit memory: history and current status. *J. Exp. Psychol.: Learn. Mem. Cogn.* 13:501–18

Schacter DL, Bowers J, Booker J. 1989. Intention, awareness, and implicit memory: the retrieval intentionality criterion. In *Implicit Memory: Theoretical Issues*, ed. S Lewandowsky, JC Dunn, K Kirsner, pp. 47–65. Hillsdale, NJ: Erlbaum

Schmidt SR. 2007. A theoretical and empirical review of the concept of distinctiveness in memory research. In *Cognitive Psychology of Memory. Vol. 2 of Learning and Memory: A Comprehensive Reference*, ed. HL Roediger III, 4 vols. (J. Byrne, ed.). Oxford: Elsevier. In press

Schmidt SR, Cherry K. 1989. The negative generation effect: delineation of a phenomenon. *Mem. Cogn.* 17:359–69

Seligman ME. 1970. On the generality of the laws of learning. *Psychol. Rev.* 77:406–18

Shimamura AP. 1986. Priming effects in amnesia: evidence for a dissociable memory function. *Q. J. Exp. Psychol.* 38A:619–44

Slamecka NJ, Graf P. 1978. The generation effect: delineation of a phenomenon. *J. Exp. Psychol.: Hum. Learn. Mem.* 4:592–604

Smith AD, Winograd E. 1978. Adult age differences in remembering faces. *Dev. Psychol.* 14:443–44

Spitzer HF. 1939. Studies in retention. *J. Educ. Psychol.* 30:641–56

Squire LR. 1987. *Memory and Brain*. New York: Oxford Univ. Press
Squire LR, Wetzel CD, Slater PC. 1978. Anterograde amnesia following ECT: an analysis of the beneficial effect of partial information. *Neuropsychologia* 16:339–47
Srinivas K, Roediger HL. 1990. Classifying implicit memory tests: category association and anagram solution. *J. Mem. Lang.* 29:389–412
Stein BS. 1978. Depth of processing reexamined: the effects of precision of encoding and test appropriateness. *J. Verb. Learn. Verb. Behav.* 17:165–74
Stickgold R. 2005. Sleep-dependent memory consolidation. *Nature* 437:1272–78
Stickgold R, Hobson JA, Fosse R, Fosse M. 2001. Sleep, learning, and dreams: off-line memory reprocessing. *Science* 294:1052–57
Stigler SM. 1978. Some forgotten work on memory. *J. Exp. Psychol.: Hum. Learn. Mem.* 4:1–4
Strong EK Jr. 1912. The effect of length of series upon recognition memory. *Psychol. Rev.* 19:447–62
Teigen KH. 2002. One hundred years of laws in psychology. *Am. J. Psychol.* 115:103–18
Thorndike EL. 1898. Animal intelligence: an experimental study of the association processes in animals. *Psychol. Rev. Monogr.* 2:Whole No. 8
Thorndike EL. 1911. *Animal Intelligence: Experimental Studies*. Lewiston, NY: MacMillan
Toth JP. 2000. Nonconscious forms of human memory. In *The Oxford Handbook of Memory*, ed. E Tulving, FIM Craik, pp. 245–61. New York: Oxford Univ. Press
Tresselt ME, Mayzner MS. 1960. A study of incidental learning. *J. Psychol.* 50:339–47
Tulving E. 1962. Subjective organization in free recall of "unrelated" words. *Psychol. Rev.* 69:344–54
Tulving E. 1966. Subjective organization and the effects of repetition in multi-trial free recall verbal learning. *J. Verb. Learn. Verb. Behav.* 5:193–97
Tulving E. 1972. Episodic and semantic memory. In *Organization and Memory*, ed. E Tulving, W Donaldson, pp. 381–403. New York: Academic
Tulving E. 1983. *Elements of Episodic Memory.* New York: Oxford Univ. Press
Tulving E. 1985a. How many memory systems are there? *Am. Psychol.* 40:385–98
Tulving E. 1985b. Memory and consciousness. *Can. Psychol.* 26:1–12
Tulving E. 2007. Are there 256 different kinds of memory? In *The Foundations of Remembering: Essays in Honor of Henry L. Roediger, III*, ed. JS Nairne, pp. 39–52. New York: Psychol. Press
Tulving E, Madigan SA. 1970. Memory and verbal learning. *Annu. Rev. Psychol.* 21:437–84
Tulving E, Schacter DL. 1990. Priming and human memory systems. *Science* 247:301–5
Tulving E, Thomson DM. 1973. Encoding specificity and retrieval processes in episodic memory. *Psychol. Rev.* 80:352–73
Underwood BJ. 1957. Interference and forgetting. *Psychol. Rev.* 53:49–60
Underwood BJ. 1961. Ten years of massed practice on distributed practice. *Psychol. Rev.* 68:229–47
Warrington EK, Weiskrantz L. 1968. New method of testing long-term retention with special reference to amnesic patients. *Nature* 217:972–74
Warrington EK, Weiskrantz L. 1970. Amnesic syndrome: consolidation or retrieval? *Nature* 228:629–30
Wertsch J. 2002. *Voices of Collective Memory.* New York: Cambridge Univ. Press
Webber SM, Marshall PH. 1978. Bizarreness effects in imagery as a function of processing level and delay. *J. Mental Imagery* 2:291–300
Weldon MS, Roediger HL. 1987. Altering retrieval demands reverses the picture superiority effect. *Mem. Cogn.* 15:269–80

Wilson EO. 1998. *Consilience*. New York: Knopf
Wixted JT. 2004. The psychology and neuroscience of forgetting. *Annu. Rev. Psychol.* 55:235–69
Wixted JT, Carpenter SK. 2007. The Wickelgren power law and the Ebbinghaus savings function. *Psychol. Sci.* 18:133–34
Woodworth RS. 1929. *Psychology*. New York: Holt & Co.
Yates FA. 1966. *The Art of Memory*. Middlesex, UK: Penguin

Contents

Annual Review of Psychology

Volume 59, 2008

Errata

An online log of corrections to *Annual Review of Psychology* articles may be found at http://psych.annualreviews.org/errata.shtml