

Learning and Memory

Henry L. Roediger III

Washington University in St. Louis, St. Louis, Missouri, USA

Jeffrey D. Karpicke

Washington University, St. Louis, Missouri, USA

Glossary

conceptual tests Measures of priming using retrieval cues that are conceptually related to studied items; these tests are considered to be conceptually driven because performance on them is affected by manipulating conceptual (meaning-based) processing.

cued recall A technique in which participants are provided with retrieval cues and are asked to recall studied items related to the cues.

explicit memory tests Measures of conscious, intentional recollection of specific learning events.

free recall A test in which participants recall as many studied items as they can, in any order they choose.

implicit memory tests Indirect measures of retention.

perceptual tests Measures in which participants are asked to identify or complete degraded stimuli; these tests are considered to be perceptually driven because performance on them is affected by manipulating perceptual processing.

priming Improvement in performance on a test that assesses retention indirectly; the measure of primary interest in implicit memory tests.

recognition tests Measures in which test participants are asked to decide whether they recognize test items from the study list, amid distracters or lures (a yes/no recognition test), or are asked to pick out the correct test item from a set containing distracter items (a forced-choice recognition test).

word fragment completion Test in which participants are presented with word fragments and are asked to complete them; priming is measured as the improved ability to complete fragments with words that had been previously presented.

Learning and memory can be measured in a variety of ways. Some of the prominent methods of measuring

retention are based on several popular paradigms used in cognitive psychology. These are surveyed in this article; coverage is selective and other techniques exist for more specialized purposes. Different measures of memory, even for the same experiences, may not always agree in their assessment, and such dissociations among memory tests provide important insights into the proper understanding of human memory.

Origins of Memory Measurement

Achievements of Hermann Ebbinghaus

The measurement of learning and memory dates back to the seminal work of Hermann Ebbinghaus, who developed the first reliable measures of memory and published his original results in 1885 in his great monograph, *Memory: A Contribution to Experimental Psychology*. Ebbinghaus created lists of nonsense syllables (e.g., JUK-WAF-KOB, etc.) that varied in terms of their length. He measured the number of trials (or the amount of time) it took to learn each series of syllables, to accomplish one perfect recitation of the entire list, a measure he called original learning. After a retention interval that could last several minutes, hours, days, or even weeks, Ebbinghaus measured the number of trials (or the time) it took to relearn the list of syllables. The difference between measures of original learning (OL) and later relearning (RL) is known as the savings in relearning. For example, if it took 20 trials to learn a series of syllables perfectly during original learning and, later, it took 10 trials to relearn the series, the savings in relearning would be 10 trials. Ebbinghaus expressed

retention performance as a savings score, defined as $[(OL - RL) \div OL] \times 100$ (or 50% savings, in this example).

One beauty of the savings method, and one reason Ebbinghaus preferred it, was that retention could be measured independently from whether the person being tested could consciously recall anything from the original learning experience. Even if a person does not have any recollection of the original list, retention of the experience can be detected by the method of calculating savings in relearning. Thus, a decade before Freud proposed his ideas about unconscious memories, which led several generations of psychologists to wonder how unconscious memories could be studied empirically, Ebbinghaus had already supplied one promising solution to the problem: the method of savings in relearning.

Strength Theory and Measurement

One beguiling idea about human memory is that memories vary in terms of a unidimensional quantity, often referred to as “memory strength.” Memories that can be easily recalled are thought to have greater strength, whereas those that can be recalled only with difficulty are thought to be weaker. Although (without further elaboration) strength theory simply redescribes the phenomenon to be explained, nonetheless, the notion of memory strength has been reified in several conceptions of memory and still holds currency in various forms even today. Strength theory also makes a strong prediction regarding the measurement of memory: if memory strength is the critical quality determining performance on any memory test, then different measures of memory should be correlated (even though they may vary in their ability to detect memories of different strengths). That is, although some measures may be more sensitive than others, all memory tests ultimately assess the same underlying quality of memory strength.

Consider the following simple example. A typical laboratory method of assessing retention involves the initial presentation of a list of words and the measurement of performance on one or more memory tests given at a later time. Consider five words (items A, B, C, D, and E) presented within a longer list of words. After studying the list, one group of individuals is given a sheet of paper and simply asked to recall as many of the words as they can remember, in any order they choose. Assume that B and D are recalled on this free recall test. A second group of individuals is given a memory test in which words that are associated with the studied words are presented as cues. For example, if “chair” had been a word presented on the study list, “table” might be given as a cue. The study participants might now recall more items on this cued recall test than they could on the free recall test. Finally, a third group of individuals is given a memory test in which

items that had actually been presented on the study list (such as chair) are mixed in with other plausible candidate items that had not been presented (such as stool). In this recognition memory test, study participants are asked to identify the items that had been presented on the original study list.

The ideas behind strength theory are illustrated in Fig. 1. The different items presented on the study list (A through E) vary in terms of their strength, with items B and D being the strongest. According to strength theory, free recall is an insensitive measure of memory, because only memories with the greatest level of strength are recalled. Free recall tests require a great amount of strength to cross the conversion threshold for recall, indicated by the top dashed line in Fig. 1. Cued recall tests are more sensitive measures and are able to detect weaker memories, such as item C in Fig. 1. Finally, recognition memory tests are the most sensitive measures of memory strength; they are able to detect all items recalled in free and cued recall tests and one additional item (E) that was too weak to be detected by the other methods. In short, the principal assumptions of strength theory are that particular memories vary in terms of their strength and that methods of assessing retention differ only in terms of their sensitivity to memory strength.

This story is tidy but wrong. In fact, different measures of memory assessed across items, people, or other independent variables are not always positively correlated, as predicted by strength theory. In fact, measures of memory may be completely uncorrelated or even negatively correlated with each other. For example, consider the effects of word frequency (i.e., the frequency with which a word occurs in a language) on memory performance measured on two different memory tests. In free recall tests, high-frequency words tend to be better recalled

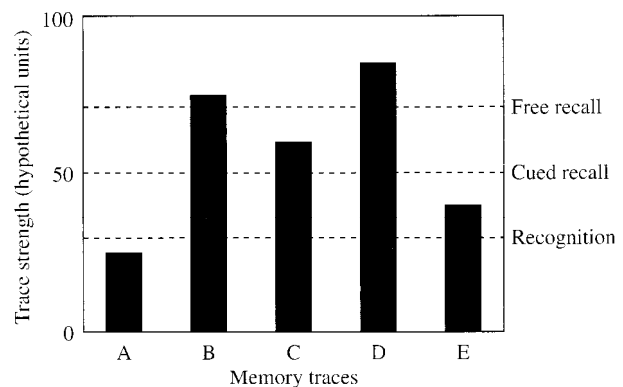


Figure 1 Strength theory. Different memory traces (A through E) vary in terms of their strength, and different measures of memory differ in their sensitivity to memory strength.

than low-frequency words are. Thus, it is reasonable to conclude that high-frequency words are stronger compared to low-frequency words. However, in recognition memory tests, low-frequency words better recognized than high-frequency words, leading to the odd conclusion that low-frequency words must have greater memory strength than high frequency words. Neither conclusion is correct. Instead, interactions between independent variables and performance on different memory tests indicate that strength theory is simply the wrong way to conceptualize human memory. Different measures of memory do not assess a single quality that varies along a single dimension.

Encoding/Retrieval Interactions and Their Implications

Processes of learning and memory are typically conceptualized as involving three stages: encoding, storage, and retrieval. Encoding is the initial registration and acquisition of information, storage is the maintenance of information over time in the nervous system (represented as a memory trace), and retrieval is the process whereby stored information is brought back into conscious awareness or otherwise affects ongoing behavior. Strength theory essentially proposes that encoding conditions will produce main effects on performance measured on different memory tests and will never interact with retrieval conditions, because different memory tests simply vary in terms of their sensitivity to memory strength. However, the literature on human memory is replete with examples in which encoding and retrieval conditions interact. The example mentioned previously, that high-frequency words are better recalled than low-frequency words are, whereas low-frequency words are better recognized than high-frequency words are, is one example of an encoding/retrieval interaction.

Larry Jacoby has designed compelling experiments demonstrating that two different measures of memory with great surface similarity can be uncorrelated or even negatively correlated. Participants in Jacoby's experiments were presented with lists of words under various study conditions and were given one of two different memory tests. One group of individuals was given a standard yes/no recognition memory test, in which they were presented with a long list of test words and were asked to determine which words had been previously studied. The other group of individuals was given a test that involved identifying words presented at very fast rates (around 30 msec per word). The proportion of words correctly identified was the dependent measure. Some of the words flashed during the test had been presented on the study list but other test words had not been previously studied. In this speeded word identification test,

the improved ability to name the briefly flashed words that had been presented during the study phase is known as priming.

In one of Jacoby's experiments, the independent variable was the level of processing of words during the study phase. Students were presented with one of three questions that oriented them toward either the surface features of the target word (e.g., "Is the word in all capitals?"), the sound of the word (e.g., "Does the word rhyme with chair?"), or the meaning of the word (e.g., "Is the word a type of animal?") before the presentation of each target word (e.g., BEAR). These three different orienting questions manipulated the level of processing that individuals performed on each word. The effects of levels of processing on performance in the two different memory tests are depicted in Fig. 2. In the recognition memory test, the typical levels of processing effect was observed: individuals were best at recognizing words they had processed at a meaningful level and worst at recognizing words they had processed at only a surface level, whereas processing the sounds of the words produced intermediate recognition performance. In contrast, consider performance on the speeded word identification test, shown in the right panel of Fig. 2. In this test, priming was measured as the difference in performance between naming studied and nonstudied words. Although all of the priming scores were positive, indicating retention of the studied words, all three encoding conditions produced equivalent levels of priming! Levels of processing, which had such a profound effect on recognition memory, had no effect on priming. Although both tests were measuring retention of the same list of items, the two measures were completely uncorrelated in this experiment.

In another experiment, Jacoby demonstrated that measures of recognition memory and speeded word identification could even be negatively correlated. In this experiment, individuals studied a list of words under one of three different encoding conditions: they were either asked to read the target words in a neutral context (XXXX-cold), to read each word paired with its opposite (hot-cold), or to generate each target word given its opposite (hot-????). Thus, in all three conditions, participants said out loud the same list of target words, but the means of having participants produce the words differed dramatically. The effects of these encoding manipulations on performance in the two different memory tests are shown in Fig. 3. Although generating the target words produced the best performance on the recognition test and reading the words in a neutral context produced the worst performance (a finding known as the "generation effect"), the opposite pattern of results was observed in the speeded word identification test: reading the words produced better identification performance than did generating the words! The results of Jacoby's experiment demonstrate that two

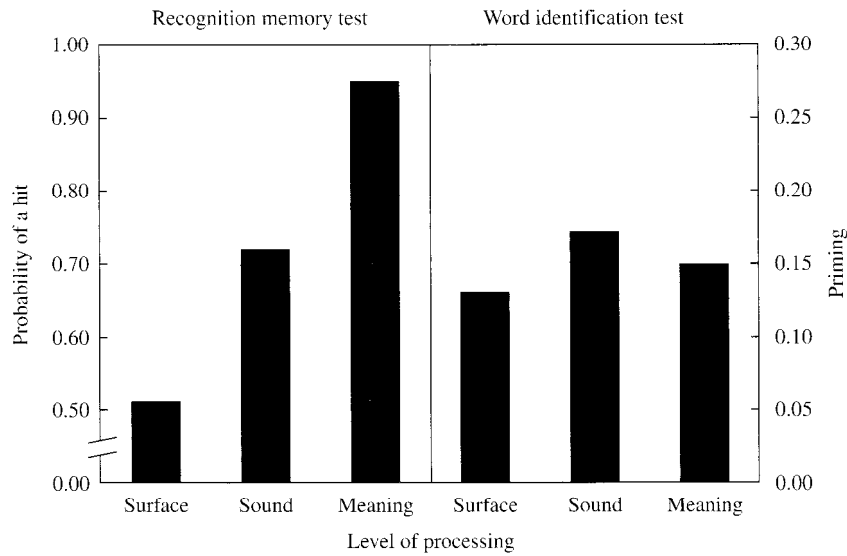


Figure 2 Levels of processing during encoding have a profound effect on recognition memory performance, but no effect on speeded word identification. The two memory tests are not correlated. Based on data from Jacoby and Dallas (1981).

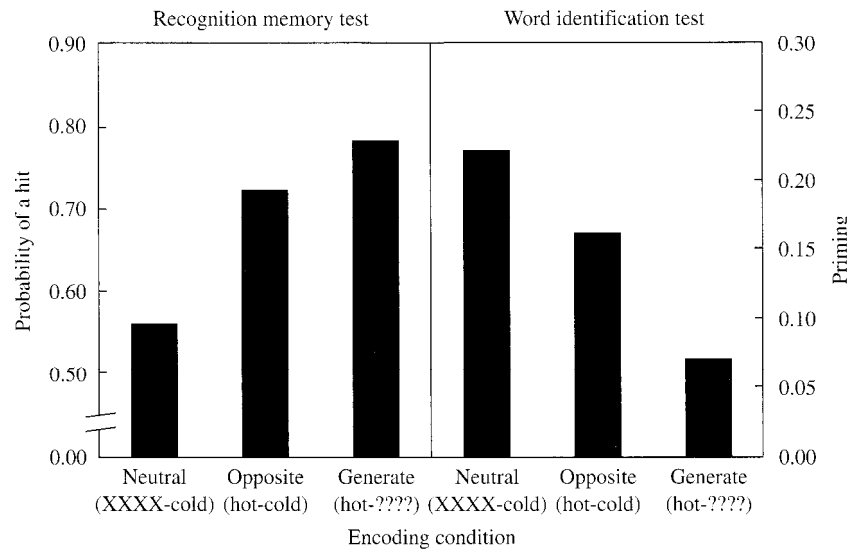


Figure 3 Generating words during encoding produced better recognition, compared to simply reading words (the generation effect). In contrast, reading words produced better performance on the word identification test, compared to generating words. The two memory tests are negatively correlated. Based on data from Jacoby (1983).

measures of memory that appear to be very similar on the surface may be negatively correlated with each other under some circumstances.

Interactions between encoding and retrieval conditions demonstrate that measures of retention can reveal positive, zero, or negative correlations with one another. These encoding/retrieval interactions have an important

implication for understanding human memory: although the concept of memory is labeled with a single word, it is hardly a single entity. Many different types of memory exist, and there are several different valid measures of memory. Some of the most prominent ways to measure memory are considered in the following discussions.

Explicit Measures of Retention

One important distinction in the field of human memory is that between explicit and implicit measures of memory. Briefly, explicit memory tests involve conscious, intentional recollection. If you are asked to describe what you had for dinner two nights ago, you must mentally travel back in time and retrieve information about that specific event. Explicit retrieval involves conscious awareness of this mental time travel. Endel Tulving has referred to explicit retrieval as involving the use of episodic memory, a memory system that enables people to remember their personally experienced past. There are several differences between explicit and implicit measures of retention; in the next section, several different explicit tests commonly used in research on human memory are described.

Free Recall

In most explicit tests of memory, participants are exposed to a set of material on which they will be tested at a later point in time. The material can consist of words, pictures, sentences, stories, videotapes, and other types of data. In a free recall test, participants are asked to recall as many of the studied items as they can, in any order they choose. Because no overt retrieval cues are provided, they must rely on their own strategies in order to retrieve the items. Free recall is perhaps the most effortful explicit memory test.

Cued Recall

In a cued recall test, participants are provided with retrieval cues and are asked to recall the studied items related to the cues. Retrieval cues can have a high or low associative strength (a measure of their degree of association with the target words). For example, “table” might be a strong cue to retrieve the word “chair,” whereas “glue” would be a weak cue in this case. A retrieval cue can also be a graphemic cue, in which a portion of the word is provided at test (e.g., “ta__ or __le” or “t_b_”). Yet another type of retrieval cue is a phonemic or rhyming cue: participants may be told to recall studied words that rhyme with the cue words. For example, “bear” might be provided as a cue to recall “chair.” Although these are the most common types of retrieval cues, others can also be developed.

Recognition Tests

Recognition memory tests are perhaps the most popular measures of explicit memory used in cognitive psychology. There are two basic types of recognition memory tests. In a free choice or yes/no recognition memory test, participants are presented with a long list of test words and are asked to say “yes” if they recognize a word from

the study list (thus, they are free to say “yes” or “no” to each item). For example, if participants studied 100 words, they might be presented with 200 words during the test (the 100 studied words mixed in with 100 distracter words) and asked to make a yes/no response for each item. Two basic measures of recognition performance are obtained in this test: the hit rate (the probability of saying “yes” to words that were actually studied) and the false alarm rate (the probability of saying “yes” to words that were not studied). The false alarm rate can be interpreted as a measure of false recognition and can also be used to correct performance for guessing (by subtracting the false alarm rate from the hit rate). Although a yes/no recognition memory test can be thought of as a 2-point scale for judging recognition, the scale can also be expanded by asking participants to provide confidence ratings. For example, participants might be asked to judge each item according to a 6-point scale, ranging from 1 (sure the item was not studied) to 6 (sure the item was studied). Such fine-grained scales provide the researcher with more information regarding participants’ decisions during the recognition memory test.

The other primary type of recognition memory test is the forced choice recognition memory test. In a forced choice test, participants are presented with several alternatives on the test (one previously studied item embedded among several distracter or lure items) and are required to choose the item that they think was presented in the study list. For example, participants might be presented with “chair table dresser lamp” and asked to choose the item that they studied. If all of the response alternatives are equally likely to have been studied, the guessing level can be determined depending on the number of alternatives. For example, in a two-alternative test, 50% reflects chance-level performance, whereas in a three-alternative test, 33% reflects chance-level performance.

Although recognition and recall tests are both explicit measures of memory, they do not always produce the same results. As mentioned previously, although people are better at recalling high-frequency words than low-frequency words, they are better at recognizing low-frequency words than high-frequency words. As another example, when adults of different ages are compared in their performance on a free recall test, older adults typically recall less than younger adults do. However, on recognition memory tests, older adults often perform just as well as younger adults do. These dissociations illustrate that recall and recognition memory tests measure different aspects of retention and cannot be considered equivalent.

Paired Associate Learning

Another popular method of studying memory is paired associate learning. In a paired associate learning test,

participants are presented with pairs of items, such as knight–heaven, elephant–house, flower–camera, and so on. Later, they are given the left-hand member of the pair (e.g., knight) and are asked to recall the right-hand member of the pair (heaven). This type of paired associate learning test measures the forward association between the pairs of words that is formed during learning. Researchers can also investigate backward associations by providing the right-hand member (heaven) and asking participants to recall the left-hand member (knight). In general, recall of forward associations is better than recall of backward associations. Because forming associations among experiences is thought to be a basic mechanism of learning, paired associate learning has been investigated since the beginning of memory research, notably by Mary Calkins, in 1894. Note that the paired associate learning procedure can be considered a type of cued recall.

Serial Recall

Although paired associate learning has been used for over 100 years, serial recall is actually the oldest memory paradigm, first used by Francis Nipher in 1878 and by Ebbinghaus in 1885. In a serial recall test, participants are presented with a series of items (such as digits, letters, words, or pictures) and are asked to recall the items in the order in which they were presented, working from the beginning to the end of the series. Recalling a telephone number is an everyday example of a serial recall task: the digits in a telephone number must be remembered in correct sequential order. Thus, memory both for the items and for the order in which they occurred is critical. A more difficult variation on this task is backward serial recall, in which a person is presented with a sequence, such as 7923948, and is asked to recall the sequence in backward order (8493297). Backward recall is often used in memory assessment batteries, because people with brain damage, for example, show greatly impaired performance on this type of task.

Implicit Measures of Retention

In addition to the explicit measures of memory, there is an entire other class of tests, implicit memory tests; these are superficially similar to explicit tests, but measure retention indirectly. For example, consider an experiment in which some participants are presented with a list of pictures (with one being of an elephant) and other participants are presented with a list of words (the names of the pictures, so the word “elephant” is presented). Following this initial presentation phase, participants are told that they will now perform a different, unrelated task, and nothing is said about testing their memory for the list

of items. In one such test, participants are presented with word fragments (“e_e_h_n_”) and are asked to complete them (as in the game Wheel of Fortune). Some of the correct answers, such as “elephant,” were presented on the study list, whereas others are not. The difference in the ability to complete studied versus nonstudied fragments is a measure of priming, the measure of primary interest in studies of implicit memory.

In the rest of this section, two main types of implicit memory tests, perceptual tests and conceptual tests, are described.

Perceptual Implicit Memory Tests

Perceptual tests present a puzzle or challenge to the perceptual system. Participants are typically asked to identify or complete a degraded stimulus, such as identifying a briefly flashed item or completing an object presented in a fragmented format. These implicit memory tests are considered to be perceptually driven because performance on them is affected greatly by manipulations of perceptual processing and less (if at all) by manipulations of meaning-based processing. Several examples are considered in the following discussions.

Word Fragment Completion

In a word fragment completion test, individuals are presented with word fragments and are asked to complete them, which is difficult if the correct answer has not been recently primed and much easier if it has been primed. In a typical word fragment completion experiment, the probability of completing a fragment that was not primed is about 0.30, and the probability of completing fragments after studying a list of words is about 0.60 (a 30% priming effect). However, after studying a series of pictures (such as a picture of an elephant), negligible priming effects are observed (0–5%), which indicates that even though the concept (elephant) may be primed, performance on a word fragment completion test depends on priming specific perceptual representations. In contrast, on explicit memory tests such as recall or recognition, pictures are remembered better than words are, a finding known as the “picture superiority effect.” Changes in presentation modality (i.e., auditory or visual) also affect priming in word fragment completion. Visual presentation of words produces the greatest amount of priming, whereas auditory presentation produces reduced levels of priming. The effects of presentation modality on word fragment completion provide additional evidence that it is a perceptually driven test.

Word Stem Completion

In a word stem completion test, individuals are presented with the first three letters of a word (“ele_____”) and are asked to complete the stem with the first word that comes

to mind. This test differs from word fragment completion because participants could potentially produce several different words from the same stem (“element, elegant, elevate, eleven,” etc.). Priming in a word stem completion test is measured as the difference between the ability to complete primed word stems versus those that were not primed. Like word fragment completion tests, performance on word stem completion tests is also affected by perceptual manipulations.

Word Identification

Priming can also be measured in a word identification test. In a typical word identification test, participants are asked to identify briefly presented words, and priming is measured as the difference between identifying previously studied versus nonstudied words. Priming in word identification tests can also be measured under auditory presentation conditions, in which participants might be asked to identify words presented against a background of noise.

Other Perceptual Tests

There are several other implicit memory tests that are perceptual in nature. For example, participants might be presented with picture fragments and, following the same logic presented previously, asked to guess the identity of the picture. Likewise, there are several auditory implicit memory tests, in which test words are presented in noise or with portions of the words deleted. Because these tests are also affected greatly by manipulations of perceptual processing during encoding, they are classified as perceptual implicit memory tests.

Conceptual Implicit Tests

In conceptual implicit memory tests, participants are provided with cues that are conceptually related to primed items. Whereas perceptual tests are affected by manipulations of perceptual processing, conceptual tests are not affected by perceptual manipulations but, instead, are affected by manipulating conceptual (meaning-based) processing. Three examples are category association, word association, and general knowledge tests.

Category Association

In category association tests, participants are presented with lists of items (such as “elephant”) during an initial phase and are later asked to generate as many instances from a category (such as “animals”) as they can during a brief testing period (lasting 30–60 seconds). Priming is measured as the greater probability of producing words that had been presented earlier than words that were not presented. In category association and other conceptually driven tests, priming is increased when the words are initially encoded in a way that emphasizes their meaning

rather than other aspects of the words (such as the surface features or sound of the words).

Word Association

Word association tests are very similar to category association tests. However, the retrieval cue in a word association test is an associate of the target word (such as “tusk”), rather than the name of a category. Priming in word association tests is also affected by manipulations of conceptual processing during encoding.

General Knowledge Tests

As the name implies, general knowledge tests require participants to answer general knowledge questions. For example, they might be asked, “What animal aided the general Hannibal in his attack on Rome?” Participants who had previously been presented with the word “elephant” will be more likely to produce the correct answer to this question, compared to participants who had not been primed with the correct answer. Once again, priming on this test is not affected by manipulating perceptual processing of the studied words (e.g., modality of presentation) but is affected by manipulating conceptual processing.

Other Measures of Memory

In addition to the measures of learning and memory already discussed, there are many more ways to test memory. Several tests have been designed to measure working memory capacity, which is the ability to hold information in mind and resist interfering information. One popular measure of working memory capacity is the operation span task, in which test participants are presented with a series of math problems followed by target words (for example, “ $12 \times 8 = 96$, WINE”). Participants must read the problem out loud, say whether it is true or false, and then read the target word. After a series of 2 to 10 operations such as these, participants are asked to recall all of the target words in the order in which they were presented. The average length of the lists that are recalled in correct serial order is the participant’s operation span. Measures of working memory capacity such as the operation span task are correlated with measures of general fluid intelligence and also predict performance on a wide variety of other cognitive tasks.

Other memory tests have been designed to measure autobiographical memory, which is a person’s memory for their own life history. One popular technique for measuring autobiographical memory is the Galton–Crovitz cue word method, in which participants are presented with a word or phrase naming a common object or activity and are asked to recollect an experience from their lives related to this cue. For example, a person who is presented with the cue “nurse” may recall an event

from when they were 5 years old in which they went to the hospital and were frightened by a scary nurse. One finding from research using this technique is that people in later life (50 years and older) most frequently retrieve memories from the period of their life lasting from late adolescence to young adulthood (roughly ages 16 to 25). Experiences from these formative years seem especially likely to come to mind.

There are also several standardized measures of memory used in educational testing and neuropsychological testing (for example, to assess whether a person with brain damage has memory impairments). The Wechsler Memory Scale and the California Verbal Learning Test are two widely used standardized memory tests. Tests such as the Scholastic Assessment Test (SAT) or even different measures of intelligence can also be considered as tests that, in part, measure recall of information learned during a lifetime (similar to the general knowledge tests).

Concluding Remarks

Although measuring human learning and memory may seem to be a straightforward enterprise at first blush, there is no single correct method of measuring retention. Instead, several diverse measures are used for different purposes to assess various aspects of a person's memory. Most importantly, different methods of testing memory for the same experiences can be completely uncorrelated or even negatively correlated with each other. These findings have important implications for our conceptualization of memory. Memory does not refer to a single, unitary property of the mind, and dissociations among measures of retention indicate that several different memory processes and systems work together to produce the complexity of human memory.

See Also the Following Articles

Behavioral Psychology • Cognitive Maps • Cognitive Psychology • Music Cognition • Spatial Cognition and Perception

Further Reading

- Balota, D. A., Dolan, P. O., and Duchek, J. M. (2000). Memory changes in healthy older adults. In *The Oxford Handbook of Memory* (E. Tulving and F. I. M. Craik, eds.), pp. 395–410. Oxford University Press, Oxford.
- Blaxton, T. A. (1989). Investigating dissociations among memory measures: Support for a transfer-appropriate processing framework. *J. Exp. Psychol. Learn. Mem. Cognit.* **15**, 657–668.
- Calkins, M. W. (1894). Association: I. *Psychol. Rev.* **1**, 476–483.
- Conway, M. A. (1996). Autobiographical memory. In *Memory* (E. L. Bjork and R. A. Bjork, eds.), pp. 165–194. Academic Press, San Diego.
- Craik, F. I. M., and Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *J. Exp. Psychol. Gen.* **104**, 268–294.
- Ebbinghaus, H. (1964). *Memory: A Contribution to Experimental Psychology* (H. A. Ruger and C. E. Bussenius, trans.). Dover, New York. (Original work published in 1885.)
- Engle, R. W. (2002). Working memory capacity as executive attention. *Curr. Dir. Psychol. Sci.* **11**, 19–23.
- Freud, S., and Breuer, J. (1960). *Studies on Hysteria* (J. Strachey, trans.). Avon Books, New York. (Original work published in 1895.)
- Gregg, V. (1976). Word frequency, recognition and recall. In *Recall and Recognition* (J. Brown, ed.), pp. 183–216. Wiley, New York.
- Jacoby, L. L., and Dallas, M. (1981). On the relationship between autobiographical memory and perceptual learning. *J. Exp. Psychol. Gen.* **110**, 306–340.
- Jacoby, L. L. (1983). Remembering the data: Analyzing interactive processes in reading. *J. Verbal Learn. Verbal Behav.* **22**, 485–508.
- Nipher, F. E. (1878). On the distribution of errors in numbers written from memory. *Trans. Acad. Sci. St. Louis* **3**, ccx–ccxi.
- Roediger, H. L., Marsh, E. J., and Lee, S. C. (2002). Varieties of memory. In *Stevens' Handbook of Experimental Psychology, Third Edition, Volume 2: Memory and Cognitive Processes* (D. L. Medin and H. Pashler, eds.), pp. 1–41. Wiley, New York.
- Tulving, E. (1983). *Elements of Episodic Memory*. Clarendon, Oxford.
- Weldon, M. S., and Roediger, H. L. (1987). Altering retrieval demands reverses the picture superiority effect. *Mem. Cognit.* **15**, 269–280.