CHAPTER 11

The Cognitive Science Approach to Learning and Memory

Matthew M. Walsh *and* Marsha C. Lovett

Abstract

The world is astonishingly complex, as is the manner in which we respond to and interact with it. Furthermore, change is an ever-present feature of the environment. Actions that once were favored may be superseded by newer, more appropriate responses. How do people acquire the knowledge to behave sensibly in the first place, and how do they continue to adapt to the changing world? These are key questions in the psychology of learning. In this chapter, the authors review research on learning and research on memory as it relates to learning using a cognitive science perspective. The defining feature of this perspective is its multidisciplinary nature. Work in various disciplines including experimental psychology, education, artificial intelligence, and neuroscience has led to discoveries about learning and memory. The insights gained advance each field of inquiry and support a more comprehensive understanding of learning and memory.

Key Words: learning, memory, skill acquisition, education, cognitive models

Introduction

The environment is startlingly complex. Humans are not born knowing how to respond to the complete range of circumstances they may face. Further complicating matters, change is an ever-present feature of the environment. Over time, actions that once were adaptive may lose their value, and new behaviors may become preferred instead. One way to confront these challenges is with general methods for learning about and adapting to the world (Walsh & Gluck, 2015*a*). These are the key issues in the psychology of learning: how do humans acquire the knowledge to act sensibly in the first place, and how do they continue to adjust to the changing world? Researchers have addressed these questions from a variety of perspectives, prompting the emergence of several major themes in the study of learning and memory.

Historical Trends in Learning Research

Research on learning and memory has been central to psychology since its inception as a scientific field in the late 19th century. Broadly, learning refers to change in behavior over time, and memory refers to the record of experience underlying change. Ebbinghaus conducted the first rigorous studies of memory (1885/1913). Using himself as a subject, he memorized lists of letter trigrams. He then measured the amount of time needed to relearn the lists after different delays. Ebbinghaus found that time saved—the difference between the initial time to learn and the relearning time—decreased with the length of the delay. Furthermore, when he studied lists repeatedly on multiple days, the amount of time to relearn decreased with each day. In addition to first demonstrating the forgetting and learning curves, Ebbinghaus's work established a methodological basis for future memory research.

At about the same time, Thorndike began studying cats as they learned to escape from an experimental apparatus called a puzzle box (1905). To escape, the cat needed to press a lever. Thorndike found that the cat initially behaved randomly and only pressed the lever by chance. Over the course of repeated trials, the cat began to move directly to the lever and escaped more quickly. Thorndike termed this the *law of effect*; animals tend to repeat actions that are followed by reinforcement (i.e., escaping from a puzzle box and receiving food) and to avoid actions that are followed by punishment. He generalized his findings to humans and used his laws of learning to guide instructional reform in what amounted to an early application of educational psychology.

Work on animal learning continued throughout the behaviorist era of the early 20th century. This movement stressed understanding associations between stimuli and responses and the study of overt behavior without recourse to nonobservable mental processes (for reviews, see Rachlin, 1970; Skinner, 1938). A series of grand theories dominated the behaviorist era; two of these were *stimulus-response* and *cognitive* theories. Stimulus–response theories (e.g., Hull, 1943) portrayed action as arising directly from associations between stimuli and responses. These theories emphasized the role of reinforcement in augmenting habit strength. Stimulus–response theories did not allow for an organism to mentally evaluate and weigh options before responding. Cognitive theories (e.g., Tolman, 1932), on the other hand, portrayed action as arising from what amounted to prospective inference over internal models or "maps" of the environment. Unlike stimulus–response theories, planning, anticipation, and outcome evaluation featured prominently in these theories.

The advent of computing and of artificial intelligence (AI) during the late 20th century led to the next major milestone in the development of theories of learning and memory. Computers became a metaphor for the mind. This reduced reluctance to think in terms of mental processes that could not be directly observed because the complex processes within a computer were not visible either. Using the microprocessor, researchers could instantiate complex models as running computational simulations (Newell & Simon, 1972). With this came renewed emphasis on how knowledge is acquired, how it is represented and stored in memory, and how it is used to guide behavior. Looking back, Ebbinghaus's research concerned memory, whereas Thorndike's concerned learning. These areas remained largely separate from one another. However, computational simulation allowed psychologists to recognize the importance of both to human cognition.

Neuroscience has increasingly driven advances in the study of learning and memory. Noninvasive techniques such as functional magnetic resonance imaging (fMRI) and electroencephalography (EEG) have allowed researchers to examine the neural change that accompanies behavioral adaptation. These methods have revealed that classic memory phenomena depend largely on the medial temporal lobe, whereas incremental learning occurs throughout structures in the basal ganglia. In addition to providing a measure of nonobservable mental processes in terms of neural activation, neuroscience has revealed insights about the brain's structure that have inspired a new generation of biologically plausible models of learning and memory.

What Is Learning?

Our survey of historical trends in learning research gives a sense of work in this area, but what precisely does "learning" mean? Researchers have proposed various definitions for this term. In their seminal text on theories of learning, Hilgard and Bower (1975) state: "Learning refers to the change in a subject's behavior to a given situation brought about by his repeated experiences in that situation" (p. 17).

Hilgard and Bower's definition emphasizes change in observable behavior. A widely held assumption implicit in their definition is that learning chiefly allows individuals to adapt their behavior to different circumstances in order to produce desirable outcomes. This definition also identifies repeated experience as being the lynchpin to learning. Experience certainly shapes behavior, but humans learn from nonexperiential sources of information as well. For example, by learning from instruction expressed using language, humans can avoid costly mistakes (Walsh & Anderson, 2011).

In the context of educational design, Mayer (2008) wrote:

> Learning is a change in the learner's knowledge that is attributable to experience. Learning depends on the learner's cognitive processing during learning and includes (a) selecting—attending to the relevant incoming material; (b) organizing—organizing the incoming material into a coherent mental representation; and (c) integrating—relating the incoming material with existing knowledge from long-term memory. (p. 761)

Mayer's definition emphasizes the object of learning—the information that is acquired and the manner in which it is integrated with existing mental representations. The key change, which may be inferred from overt behavior, involves knowledge in long-term memory. Mayer also stresses the individual's need to attend to and interpret experiences in the course of learning.

Turning to AI, Mitchell (2006) said:

> Machine Learning focuses on the question of how to get computers to program themselves (from experience plus some initial structure). . . . Machine Learning incorporates additional questions about what computational architectures and algorithms can be used to most effectively capture, store, index, retrieve and merge these data. (p. 1)

Mitchell refers to "algorithms." Analogizing to psychology, these are the mental processes that convert experience into a form suitable for storage and future use. He also refers to "initial structure" and "computational architecture." Learning algorithms and the lasting changes they produce must be realized in the biological hardware of the brain.

Collectively, these definitions identify a set of conditions for a learning system: (1) the system interacts with the environment; (2) the system extracts, processes, and records information from its experiences; and (3) the system selects future behavior based on these records. Common across all definitions is the idea that learning involves change sustained over time. The definitions further specify storage of knowledge. In the historical context, this maps onto the distinction between research on learning—the change in behavior over time, and memory—the record of experience underlying change.

What Distinguishes a Cognitive Science Approach?

Given the multidisciplinary nature of cognitive science, studying learning from a cognitive science approach involves extending the theory and methods of psychology to include perspectives from AI, education, or neuroscience. Whereas psychological research on learning typically includes testing hypotheses derived from verbally expressed theories, cognitive science research on learning often (but not always) involves instantiating theory in the form of a computational model—a set of computational processes that represent the theory and can be "run" by computer simulation. The theory's predictions can then be tested by comparing model output to observed data, with the goal of refining the theory (and hence, the model) so that it fits existing data and makes novel predictions that can be empirically tested. Thus, a cognitive science approach to learning differs from a pure AI approach in which the computational system's goal is to meet a given objective without necessarily matching human performance.

A cognitive science approach to learning may also incorporate other disciplines' perspectives via the nature of the research question and the source of the data. For example, a cognitive science study of learning may investigate not only whether a given learning mechanism explains participants' behavior and performance in the laboratory, but also whether an instructional intervention based on that mechanism improves student learning in a real educational context. This cognitive science approach takes a psychological research question and translates it to a related educational research question. Similarly, a cognitive science study of learning may involve combining behavioral and neuroimaging data to address how learning unfolds across time and to reveal the neural underpinnings of that change.

The goal of this chapter is to survey learning research from the cognitive science perspective. As we organized our thoughts, we realized that learning could not be adequately addressed without recourse to memory. As such, we also survey memory research as it relates to learning. Collectively, the sections that follow exemplify the multidisciplinary nature of learning research. We begin by describing key phenomena—classic empirical results discovered in psychology laboratory experiments. We move on to computational cognitive models—theories of learning and memory instantiated as running computational simulations. We close by discussing applications of learning and memory research to education. Throughout, we consider how advances in neuroscience relate to and inform work on learning and memory.

Setting the Stage: Key Phenomena

Psychologists use experiments to study behavior and the processes underlying it. Observation, in the form of empirical or measurable evidence, and reasoning are central to the experimental method. Furthermore, the experimental method involves manipulating one or more variables in a controlled manner in order to determine how they affect behavior. Returning to an example from the introduction, Ebbinghaus studied how serial lists are memorized and retained (1885/1913). Because prior knowledge of words could potentially affect learning, he created and memorized lists of meaningless syllables instead. He studied each list until he correctly recalled all syllables twice, after which he waited from 20 minutes to 31 days to restudy the list. As the length of the retention interval increased, Ebbinghaus needed more trials to reach a criterion of two correct trials.

Several aspects of Ebbinghaus's experiment are notable: he created a design to isolate the mental processes of interest, he manipulated a dependent variable (length of the retention interval), and he measured its effect on an independent variable (number of trials to criterion). The experimental method has since been applied to a myriad of questions related to learning, memory, and the acquisition of expert performance. These experiments have revealed robust phenomena that are grist for theories and computational cognitive models.

Learning

Human learning research has its roots in studies of animal learning. Many of these studies involve animals acquiring associations between stimuli or between a stimulus and response. Overt behavioral changes reveal the existence of these associations. Studies of animal learning underscore the diversity of associations that can be formed and identify factors that influence the acquisition and persistence of associations. An often-tacit assumption underlying this research is that the basic findings generalize across species and to humans. This would be problematic if not for ongoing efforts, beginning with the rise of cognitive psychology, to replicate and extend animal research to human subjects. Although there are clearly differences among species, there are similarities as well. Many of the results first obtained with animals hold for humans. Thus, animal research has motivated comparative work and has inspired computational models of learning and memory.

Animals can learn associations between neutral and biologically potent stimuli. In the standard classical conditioning paradigm, a neutral stimulus is repeatedly presented before an unconditioned stimulus (US). For example, in Pavlov's famous experiment (1927), a dog sees a light and receives food. Initially, food causes an unconditioned response (UR), salivation, but the light does not. Eventually, the dog associates the light with food and salivates when the light is turned on. At this point, the light is called a conditioned stimulus (CS), and salivation to the light is called a conditioned response (CR). Humans also acquire CRs. For example, in eyeblink conditioning, a stimulus precedes a puff of air directed toward the outside of the eye. The puff of air initially causes the person to blink. Eventually, the neutral stimulus causes the person to blink as well.

In classical conditioning, the individual reacts to an event but cannot alter the outcome. In other words, the response is merely preparatory. When they are allowed to enact behaviors that affect outcomes, animals (and humans) can learn which responses produce reinforcement. In the standard instrumental conditioning paradigm, an animal is presented with a stimulus, chooses a response, and is rewarded or punished. For instance, Thorndike's cats were rewarded with escape and food when they pressed the correct knob in the puzzle box. Many additional examples of instrumental conditioning involve humans. For example, in a study by Friedman et al. (1964), participants chose between two buttons, after which a light was illuminated. If the light matched the selected button, the participant received a small monetary bonus. Over the course of many trials, participants learned to favor the button that was rewarded more often.

Research on instrumental conditioning raises several nuanced questions. What types of associations are learned? According to one view, rewards "stamp in" stimulus–response associations. How to respond to a stimulus is remembered, but the type of reward is not. Some evidence suggests that animals do develop expectations about the specific reward, however. For example, Adams and Dickinson (1981) trained rats to press a lever for a food pellet. They then induced an aversion by pairing the food with illness. When the rats were later allowed to press the lever, they did so less frequently indicating a reduced preference for the devalued reward. Another question is whether the mere co-occurrence of the response and reward or the dependency of the reward on the response drives behavior. To address this question, Balleine and Dickinson (1998) weakened action–outcome associations by freely giving rats rewards whether they pressed a lever or not. Rats responded less frequently when rewards were delivered noncontiguously with lever presses, showing that outcomes must depend on actions for instrumental conditioning to occur.[1]

What about the stimulus is learned? Humans and animals typically encounter variations of a stimulus rather than exact matches. Learning requires generalizing knowledge to related stimuli. Guttman and Kalish (1956) trained pigeons to peck a key of a particular color. When they were exposed to new keys, pigeons pecked the matching key most often, and they pecked all other keys in proportion to their differences from the original color. In other words, pigeons showed a degree of generalization. Humans also generalize responses based on stimulus properties such as color, hue, orientation, size, and sound; the generalization gradient decreases exponentially with the psychological distance between physical properties of the original stimulus and the new stimulus (Shepard, 1987). Researchers have created two types of models to accommodate these results. In exemplar models, a new stimulus is compared against existing examples stored in memory. Alternatively, in rule-based models, the value of a stimulus along a specific dimension is encoded and used to categorize the stimulus (Anderson, 1991; Nosofsky, Palmeri, & McKinley, 1994).

A final question about instrumental conditioning involves the types of responses learned. Initially, researchers thought animals learned effector-specific motor responses. This definition was later expanded to include all instances of responses that had identical effects in the environment (i.e., operant classes, Skinner, 1938): for example, pressing the same lever using either paw. Animals learn still more general response rules. Packard and McGaugh (1996) placed rats in a cross-maze. Upon reaching the central intersection, they chose between two goal boxes located to the east and west. One box contained a food pellet. Depending on how the pellets were placed, rats learned a place strategy (i.e., always go to the eastern goal box) or a response strategy (i.e., always turn right). Other maze studies have established that rats can learn a *stay* strategy that involves returning to the previously rewarded arm or a *switch* strategy that involves alternating arms. Humans can also learn simple stimulus–response associations and abstract response rules. In a study by Walsh and Anderson (2013*a*), participants viewed number pairs. They learned to select specific numbers from certain pairs that were frequently rewarded. Additionally, they learned abstract rules like *choose the larger number in the pair*. These rules were abstract in that they could be applied to studied and novel number pairs alike. More generally, people learn to choose among complex multistep strategies based on the strategies' base rates of success (Lovett & Schunn, 1999; Walsh & Anderson, 2009).

Memory

Learning is a process of long-lasting change in behavior caused by experience. Memory is the record of experience underlying learning. A key distinction between these topics is that studies of learning focus on how an individual responds to different stimuli and in different contexts, whereas studies of memory focus on how information is encoded, stored, and retrieved. The two are not truly separable—memory processes are a subset of the psychological processes engaged by learning.

Distinction is made between transient memory systems, such as sensory and working memory, and long-term memory (Atkinson & Shiffrin, 1968). Transient systems have limited capacity and enable recall for a period of seconds to minutes. These systems support ongoing information processing and may serve as an intermediary to the formation of more-permanent memories, but they alone do not support the long-lasting change denoted by learning. Long-term memory, in contrast, has a seemingly unlimited capacity and enables recall over far greater periods of time. The acquisition of knowledge stored in long-term memory provides the records of experience underlying learning.

Much research in psychology concerns how the speed and success of retrieving information from memory changes with practice. This research shows that with increased practice, people can retrieve information faster and more consistently. In a study by Pirolli and Anderson (1985), participants memorized sentences. They were then presented with studied and novel sentences and asked to decide whether they had seen the sentence before. Over the course of 25 days, the speed of their responses to studied sentences increased. The speed-up followed a power-law of learning: response times dropped substantially over the first several days and continued to decrease at a diminishing rate over the course of the study. This type of learning function is practically universal: it characterizes changes in response times and error rates, and it applies to the acquisition of facts (i.e., declarative knowledge) and skills (i.e., procedural knowledge). The relationship between amount of practice and ease and durability of retrieval has significant educational implications. For example, longitudinal studies of arithmetic show that children shift from counting strategies to retrieval. The strategy shift coincides with the increased speed and accuracy of retrieving arithmetic facts from memory, which is mediated by amount of practice (Siegler, 1999).

Memories fade over time. In one study that examined knowledge retention, participants memorized Spanish vocabulary words (Bahrick, 1979). On the first day of the experiment, they were tested until they responded correctly to each word. One day later, memory performance dropped to 53% correct. Extending the retention interval to 1 week and 1 month reduced memory performance to 39% and 21% correct, respectively. In related work that examined knowledge retention on an educationally relevant timescale, Bahrick recruited a cross-section of individuals who had studied Spanish in high school or college from 1 to 50 years earlier (1984). Proficiency, measured in terms of grammar and vocabulary recall, dropped sharply during the years immediately following the class and more gradually over the ensuing decades. Similar long-term retention functions apply to doctors' recollection of basic science knowledge from medical school (Custers & ten Cate, 2011) and even for people's memories of historical and popular events (Roediger & DeSoto, 2014). The memory decline in all of these studies follows a power law: recollection drops most during the years immediately following education and continues to decrease at a diminishing rate.

The initial level of practice can partially offset the negative effects of time on retention. In Bahrick's study (1984), individuals who took more Spanish classes had greater retention up to 50 years later. The same is true in mathematics (Bahrick & Hall, 1991). However, regardless of their initial level of learning, people in both studies showed substantial forgetting in the long run. The implication is that simply practicing more up front will not eliminate forgetting. That is not to say that the benefits of initial study are wholly lost. When people review material after an extended delay, their starting level of performance is low but they quickly relearn. Interestingly, review seems to decrease the rate of subsequent forgetting, as evidenced by enhanced long-term retention following review after a delay (Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006).

Simply increasing the amount of practice may not improve recall. The manner in which one studies is also important. This idea is central to the depth of processing theory, which holds that rehearsal is only beneficial when material is studied in a deep and meaningful way. To test this theory, Craik and Tulving (1975) presented participants with words. They asked participants questions about the words' physical properties (i.e., *Is the word shown in capital letters?*), their phonemic properties (i.e., *Does the word rhyme with WEIGHT?*), or their semantic properties (i.e., *Is the word an animal name?*). The three types of questions were designed to evoke increasingly deeper levels of processing. In a subsequent memory test, recognition was about four times better (96% correct vs. 22%) for words that had been presented with semantic versus physical questions during the encoding phase. A related finding is that *elaborative processing*, which involves embellishing to-be-remembered information, enhances retention. This effect underlies *elaborative interrogation*, a study strategy in which the student reads a fact and generates an explanation for it (Dunlosky, Rawson, Marsh, Nathan, & Willingham, 2013).

Various theories in cognitive science account for the benefits of depth of encoding and elaborate processing. These theories propose that memory is organized as an interrelated network of concepts. A concept like the word *dog* is connected to other related concepts (e.g., *cat* and *bone*). When an external prompt such as a test question is given, multiple concepts become active, and activation spreads among related concepts. Deep encoding and elaborate processing facilitate integration of new knowledge with existing concepts. As such, when a student attempts to recall a newly learned fact, more activation spreads from related concepts. This increases the fact's activation and the corresponding probability of retrieval.

The distinction between learning and memory is more apparent than real; even in paradigms meant to isolate learning processes, some record of experience must be stored to support long-lasting behavioral change. Thus, all of the empirical phenomena we have described actually involve learning *and* memory. However, the phenomena differ in terms of the types of memory they involve—declarative or nondeclarative (Squire, Knowlton, & Musen, 1993). *Declarative memory* is available to consciousness and can be expressed using language. *Nondeclarative memory*, also called *procedural memory*, is not consciously available.

Neuroscientific findings substantiate this phenomenological distinction. The acquisition of declarative knowledge depends on structures in the medial temporal lobe. Damage to these causes anterograde amnesia—the inability to form new memories. Existing memories are not affected, however, indicating that other areas, most notably the associative cortex, are responsible for the long-term storage of declarative knowledge. The acquisition and storage of nondeclarative memory depends on a different set of structures in the basal ganglia. These are affected in patients with Parkinson's disease, who show impairments in probabilistic learning and skill acquisition. In a study that doubly dissociated these conditions, patients with amnesia or Parkinson's disease performed a weather prediction task (Knowlton, Mangels, & Squire, 1996).

They viewed sets of cues and predicted whether the correct response was sunshine or rain. Unknown to them, a probabilistic rule mapped cues to the correct responses. Patients with amnesia learned to correctly respond to cues despite remembering little about the training examples. Patients with Parkinson's disease, in contrast, recalled details about the training examples but did not learn to correctly respond to cues.

Expert Performance

Comparing the performance of experts and novices in a given domain provides a compelling demonstration of the transformative power of learning. Counter to the common belief that experts are innately talented, their performance as compared to novices constitutes an extreme point along a continuous learning trajectory. Experience, paired with learning, engenders change in experts' strategies, perception, and long-term knowledge structures. The development of expertise, then, depends on the same learning and memory processes described in the previous sections. In the case of experts, however, improvement is gradually amassed over thousands of hours of deliberate practice (Ericsson, Krampe, & Tesch-Römer, 1993).

Fitts (1964) and Anderson (1982) proposed three stages of skill acquisition. The first is called the *cognitive stage*. During this stage, people commit rules and domain knowledge to memory. The individual may learn from instruction or examples. Performance is slow and effortful. To solve problems, one must retrieve facts from memory and interpret them. The chief benefit of practice at this point is to increase the accessibility of information from memory. The second stage is called the *associative stage*. Initial errors of understanding are gradually detected and eliminated. Additionally, task knowledge is compiled into specialized procedures that minimize the amount of information that must be retrieved from long-term memory and held in working memory. Upon encountering a set of conditions, the individual no longer needs to reason about how to respond. Rather, he or she simply recognizes and implements applicable rules and responses. The final stage of skill acquisition is called the *autonomous stage*. Performance becomes increasingly automatic, requires less attention, and interferes less with other ongoing tasks. Performance continues to improve with practice, but more slowly. Skills that have become automatic, or proceduralized, decay more gradually than declarative knowledge. People remain proficient at highly trained skills, such as cycling or skiing, even after many years absence (Schmidt, 1988).

This model of skill acquisition is informative with respect to educational practice. It indicates that instructional designers must consider the learner's achievement level to strengthen currently active knowledge structures and support progression to the next stage. Instruction that does not take current ability into account can create unwanted aptitude-by-treatment interactions, including the expertise reversal effect (Kalyuga, Ayres, Chandler, & Sweller, 2003). Another implication of this model is that some amount of practice (i.e., *drilling*) is necessary to ensure that task knowledge and procedures are accessible from memory. This is a precondition to progressing to later stages of automaticity. A final implication is that part-task training may facilitate the acquisition of complex tasks by placing appropriate demands on the learner. This is especially true if the parts are performed sequentially rather than concurrently during full-task performance (Wickens, Hutchins, Carolan, & Cumming, 2013). Gradually increasing the speed or difficulty of the full-task may also facilitate acquisition by moderating performance demands.

What underlying differences account for experts' exceptional performance? Experts use specific procedures rather than general knowledge to solve problems. For example, in a study by Walsh and Gluck (2015*b*), participants were asked to think aloud as they performed a stock market selection task. In each trial, they considered multiple financial indicators about two stocks before selecting one. Over the course of the study, they made 58% fewer statements about the rules of the task and strategies for performing it. In contrast, they only made 2% fewer statements about the values of the indicators in a trial and 20% fewer statements about the selection. This indicates that participants continued to attend to indicators and responses but that they stopped using general knowledge to reason about and choose responses. The conversion of general knowledge to specific rules is called *proceduralization* and has been observed in academic domains like geometry and physics as well (Sweller, Mawer, & Ward, 1983).

Experts use more efficient strategies. In the course of math learning, children first solve addition problems by counting both addends (the *count* strategy), then by counting up from the larger addend (the *min* strategy), and then by retrieving the sum from memory (the *retrieval* strategy). The final transition coincides with the increased accessibility of arithmetic facts from memory (Siegler, 1999). Studies of adults performing alphabet arithmetic reveal a similar transition from strategies with many intermediate steps to a declarative memory strategy with one step (Logan & Klapp, 1991). Experts and novices also differ in the search strategies they apply to complex problems. In a study of mathematical problem-solving, Sweller et al. (1983) found that novices used means-ends analysis, which involved working backward from the goal to the problem's givens, whereas experts worked forward from the problem's givens to the goal. Most novices eventually transitioned to the forward-search strategy, after which they solved problems faster and with fewer steps. The actual search strategies used—forward, backward, or other—vary by domain and problem type. A hallmark of expert performance is mastery of multiple strategies and their flexible application to different problems.

Experts and novices perceive problems differently. Schoenfeld and Herrmann examined problem perception in mathematics (1982). They asked experts (mathematics professors) and novices (undergraduates) to sort math problems into related categories. Experts sorted problems according to the mathematical principles needed to solve them, whereas novices sorted problems according to their superficial details. In related work, Chi, Feltovich, and Glaser (1981) found that experts and novices in physics also sorted problems differently based on deep or superficial features, respectively. The ability to map surface features of a problem onto deeper principles is useful because deeper principles are more predictive of the solution method. A study strategy called *interleaving* provides an interesting demonstration of this principle. Practicing concepts related to different subtopics in an interleaved rather than a blocked manner facilitates test performance (Rohrer & Taylor, 2007). One explanation for the benefit of interleaved practice is that it allows students to practice identifying which solution method to use for a given problem.

Last, experts and novices differ in terms of the amount of information stored in long-term memory. Experts possess the basic collection of facts, definitions, and concepts needed to perform a task. In addition, they have amassed an extensive collection of memories based on meaningful patterns, conditions, and problems previously encountered. For example, Chase and Simon (1973) estimated that chess masters have acquired a repertoire of 50,000 to 100,000 different chess patterns and learned what to do in the presence of each. Expert performance in chess, then, is not a matter of thinking, but of recognizing patterns and acting. The intuitions of expert operators, managers, and commanders are thought to arise from the same process of mapping current conditions onto previously encountered patterns stored in memory (Klein, 2008).

Computational Models of Learning

One of the main challenges in psychology research is that the mental processes that give rise to behavior cannot be directly observed. For example, most researchers agree that people can direct attention to different parts of a diagram, temporarily hold information in short-term memory, learn about the usefulness of different strategies from experience, and retrieve past episodes from memory. Yet none of these processes is visible—each must be inferred from its impact on behavior.

One way to "see" the processes involved is with computational cognitive models. Cognitive models describe how a particular combination of mental processes gives rise to behavior. Models are evaluated by comparing their predictions to people's behavior. If the correspondence is high, the model may be useful for understanding a theoretical phenomenon or for predicting behavior. Computational cognitive models advance basic science. They show how variations in behavior arise from interactions between cognitive processes—approximated by the model—and properties of the task and context. In this way, they unify the results of past experiments and their predictions motivate future studies (McClelland, 2009). Cognitive models also enable application. For example, a multitude of factors in instructional design affect learning outcomes. By representing these factors and their interactions in a cognitive model, the model can be used to explore potential designs and curricula when it is not possible to do so with actual students (Koedinger, Booth, & Klahr, 2013).

Cognitive models employ different types of computational processes, and they represent knowledge in different ways. In this section, we describe three unified architectures—general models intended to account for all of cognition. The three are *State Operator and Result* (Soar), *Parallel Distributed Processing* (PDP), and *Adaptive Control of Thought—Rational* (ACT-R). Each emphasizes different types of learning processes, and each represents knowledge in different ways.

State Operator and Result

Soar is based on a production system (Newell, 1990). Knowledge is represented in the form of production rules, each of which is made up of a set of conditions (the *if* part of the rule) and an action (the *then* part of the rule). If the conditions are met, the rule applies and the action is performed. In Soar, intelligent behavior is viewed as a form of problem-solving. The problem space consists of states and operators. Conditions of the problem state at each point in time determine which productions are eligible for selection. Operators contained in the *then* part of productions are chosen to move through the problem space toward the goal state.

The selection of an Operator in Soar is divided into elaboration and decision. During elaboration, knowledge encoded as production rules is used to evaluate candidate operators. If knowledge is not sufficient to determine which operator to perform next, an *impasse* occurs. The new subgoal of resolving the impasse is formed. Once the subgoal is completed, focus returns to the main goal. A new production rule that consists of the set of conditions prior to the impasse and the operator that was eventually selected is added to memory. Stated more concretely, once the system has performed the deliberative processing necessary to solve a problem, it stores the result as a new rule. The accumulation of these rules gradually eliminates the need for internal deliberation. This form of learning, called *chunking*, is the core mechanism of change in Soar. The gradual elimination of intermediate steps by chunking produces a negatively accelerated speed-up in task completion time, characteristic of the power law of learning.

Soar's chunking mechanism is demonstrated in a model of the Tower of Hanoi puzzle (TOH; Ruiz & Newell, 1989). In the TOH, five disks of graded size form a tower on an initial peg. Participants can move disks to an intermediate peg, and they must ultimately move all disks to the destination peg. They can only move one disk at a time, and they cannot place a larger disk on top of a smaller disk. Across repeated trials, they solve the puzzle faster and using fewer moves. The Soar model, like people, begins by using a trial-and-error strategy. The model moves disks to the destination peg semi-randomly. The model improves in two ways. First, midway through learning, the model begins to notice subtowers rather than individual disks. This triggers the creation of chunks that recognize the existence of nested subtowers and the use of a recursive strategy for moving all of the disks that make up the subtower. Second, Soar gradually learns chunks that map each TOH state to the correct operator. With enough experience, these chunks completely eliminate the need for deliberative processing. This shift from general reasoning to specialized procedures mirrors the transition from the cognitive to the associative stage in models of skill acquisition (Anderson, 1982; Fitts, 1964).

Chunking was initially the only learning mechanism in Soar. Although chunking is applicable to a wide variety of tasks and strategies, other learning mechanisms have since been added (Laird, 2012). These include computational processes for forming numeric operator preferences based on history of reinforcement and processes for shaping the activation of knowledge in Soar's working memory.

Parallel Distributed Processing

PDP models provide a very different account of cognitive processing and representation (for a review, see also Flusberg & McClelland, this handbook). In PDP models, cognitive processes arise from interactions among simple, neuron-like units with activation values. Units are typically arranged in layers (Figure 11.1). One set of units receives input from the external world, and another is designated as the system's output. Activation passes from the input layer, through one or more hidden layers, and to the output layer. Weighted connections between units in the different layers control the spread of activation throughout the network and transform the input to the network's output. Knowledge is contained abstractly in the strengths of connections between units, which are learned.

Many PDP models employ a type of representation called a *distributed representation* (Figure 11.1). Concepts such as robin, oak, or salmon are represented by the pattern of activation across several units rather than by the activation of one unit. Individual units can be thought of as "microfeatures." A particular concept, then, is represented by the set of microfeatuers that it activates, and two concepts are related to the extent that they share microfeatures. Distributed representations naturally support the stimulus generalization seen in studies of classical and instrumental conditioning. When the network encounters a new item, it activates other items to the extent that they have microfeatures in common

The weighted connections that give rise to distributed representations are learned. *Backpropagation* is one technique for acquiring a set of weights. Briefly, backpropagation involves (1) providing inputs to the network, (2) transforming inputs to outputs based on connection weights between units in the network, (3) calculating differences between the network's actual outputs and the correct target

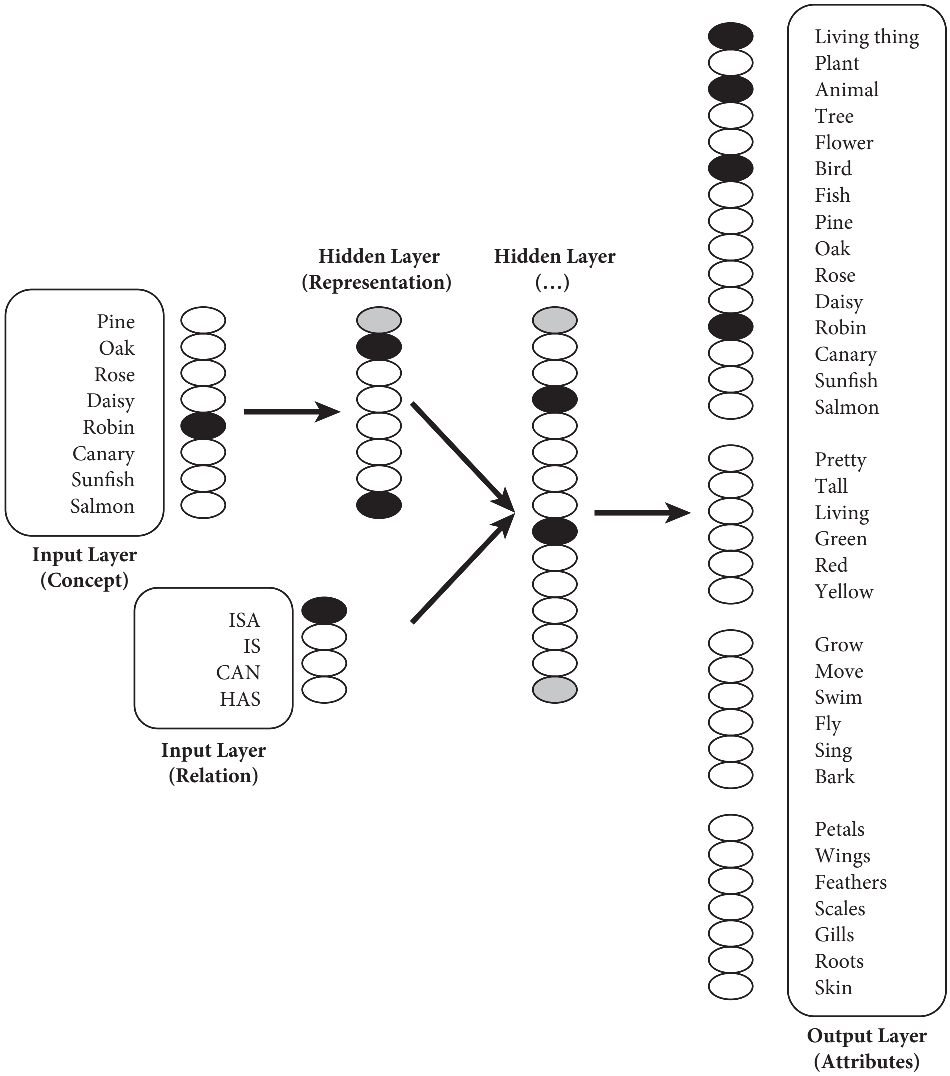

Fig. 11.1 Parallel distributed processing (PDP) network with units organized into input layers, hidden layers, and an output layer. Each unit is connected with every unit in the next layer. Activation (denoted by shading) is distributed across multiple units.

outputs, and (4) modifying connection weights based on the error signals to reduce differences.

Rumelhart and Todd's (1993) model of semantic knowledge illustrates the learning dynamics of a network with distributed representations (Figure 11.1). The network learns about concepts, relations, and attributes; for example, the fact that a robin is a bird, has wings, and can fly. The model is given inputs (e.g., *Robin ISA*. . .) and activates attributes in the output layer that complete the proposition (e.g., *living thing, animal, and bird*). During initial training, the network is also given target outputs, which it compares against its own outputs. The network uses backpropagation to adjust connection weights in order to reduce the difference between the actual and target outputs.

Concepts are represented in a distributed fashion across units in the network's hidden layer. The more similar two concepts, the more similar the set of hidden units they activate. Note that the network is never given target values for hidden units; their activation values are an emergent property of the set of weights learned to map inputs to target outputs. Over the course of learning, patterns of distributed activation in the hidden layer first become distinguishable between superordinate categories (plants/animals), next among intermediate categories (trees/plants and birds/fish), and last among individual items (oak/pine, rose/daisy, robin/canary, and salmon/sunfish). The sharpening of concept representations is paralleled by the network's success in completing general propositions first and item-specific propositions last. Developmentally, young children show the exact same progressive differentiation of concepts (McClelland & Rogers, 2003).

For a network to learn distributed representations, training must be interleaved. If the network is repeatedly exposed to one concept, connection weights change so much that information about other concepts is lost. This is called *catastrophic interference*. How, then, can people acquire concepts that are presented just once or a small number of times? To resolve this dilemma, McClelland et al. (McClelland, McNaughton, & O'Reilly, 1995) proposed the idea of complementary learning systems. The hippocampal system learns rapidly, but does not extract similarity across items. The hippocampus reinstates recent memories and plays them back to the neocortex in an interleaved fashion. This allows the neocortex to gradually extract similarity across the ensemble of experiences and to form distributed representations. In addition to resolving the problem of catastrophic interference, this model accounts for the pattern of memory deficits seen in amnesic patients. Damage to the hippocampus (a part of the medial temporal lobe) results in the inability to form new memories, whereas long-term memories stored in the neocortex (also called the associative cortex) are spared.

Adaptive Control of Thought–Rational

ACT-R is made up of a set of specialized information processing modules (Anderson, 2007; see also Salvucci, this handbook). These include a visual module for seeing, a manual module for responding, a declarative module for storing and retrieving information in memory, an imaginal module for holding intermediate problem representations, a goal module for maintaining information about context and intent, and a procedural module for coordinating other modules' behavior. During task performance, information is processed within modules and passed between modules. The order in which modules are engaged and the information transformations that occur within each depend on the task.

A fundamental distinction in ACT-R is between procedural and declarative knowledge. Procedural (i.e., *nondeclarative*) knowledge involves information about the history of successes and failures of different actions. This information is represented as *utility values*, which describe the reward associated with taking different actions at different times. An example of procedural learning comes from the building sticks task (BST; Lovett & Anderson, 1996). In the BST, participants are given an unlimited supply of sticks of three lengths. Their objective is to add and subtract sticks to create a new stick of a certain length. The two basic strategies are to start with a stick that is longer than the target length and subtract shorter sticks (i.e., *overshoot*), or to start with a stick that is shorter than the target length and add more sticks (i.e., *undershoot*). Lovett and Anderson created two sets of BST problems. One set could only be solved using the overshoot strategy and the other could only be solved using the undershoot strategy. The first set of problems was followed by a second where the opposite strategy worked. After participants completed the first set of problems, they continued to use the initially favored strategy even though it no longer worked.

Reinforcement learning (called *procedural learning* in ACT-R) produces this sort of gradual adaptation (Fu & Anderson, 2006). In reinforcement learning, the individual acquires expectations about the rewards associated with different actions—for example, the success of the overshoot and undershoot strategies in the BST (Walsh & Anderson, 2013*b*). After the individual enacts an action and receives feedback, a prediction error is calculated. This is the difference between the reward that the individual expected and the reward they received. Prediction errors are used to adjust expectations so that they gradually come to match the actual reward associated with each action. In addition to accounting for the acquisition of complex strategies in the BST, procedural learning explains how humans and animals acquire simpler contingencies between instrumental responses and rewards.

The second form of knowledge in ACT-R is declarative. Declarative knowledge involves facts such as "7 + 3 = 10," or "Lincoln was president of the United States." Items in declarative memory have *activation values*, which depend on their study histories. An item's activation and the corresponding probability that it can be retrieved from memory is greatest if the item was encountered recently and frequently. This maps onto the empirical finding that the amount of practice and the length of the retention interval are the two main determinants of memory performance. The mathematical function that characterizes how frequency and recency affect memory performance (called *base-level learning* in ACT-R) also happens to predict the real-world probability that an item encountered in the past will come up again (Anderson & Schooler, 1991). In other words, the information most likely to be needed in the world is also most accessible from memory.

Anderson et al. (Anderson, Fincham, & Douglass, 1999) tested base-level learning in an experiment where participants memorized statements such as "Skydiving was practiced on Saturday at 5 p.m. and Monday at 4 p.m." They then saw part of the statement (e.g., skydiving, Saturday, and Monday) and filled in the remaining details (e.g., 5 p.m. and 4 p.m.). Participants were tested multiple times within a session and again during another session that occurred from 1 day to 14 months later. The speed and accuracy of responses increased within sessions because items were encountered so recently and frequently. Speed and accuracy were lower after the long retention interval between sessions because items had not been encountered recently. These results held for declarative knowledge *and* for the application of a complex rule that could be used to complete the statements, perhaps because application of the rule also required retrieval of steps stored in declarative memory.

ACT-R contains a final learning mechanism, called *production compilation*, for combining successive pairs of model operations (Taatgen & Anderson, 2002). An interesting example of production compilation involves the retrieval of information from declarative memory based on an external prompt and the initiation of a response. With repeated practice, production compilation combines these two operations into one, allowing the external prompt to directly trigger the response. This is similar to chunking in SOAR and is as a computational instantiation of the transition from the cognitive to the associative stage in models of skill acquisition (Anderson, 1982; Fitts, 1964).

Applications of Learning Research to Education

The ramifications of research on learning and memory are perhaps most apparent in the classroom. Educational psychology focuses on diverse topics such as human development, individual differences, personality, motivation, and, most critically, their role in learning. Educational psychology shares much with cognitive science, including theories about the mental processes evoked during learning and commitment to formal observation and methodological rigor. Cognitive science is somewhat broader in that the educational setting is a particular context for observing cognitive phenomena related to learning and memory. In the case of educational psychology, however, learning research is squarely directed toward improving educational outcomes.

Already, we have alluded to applications of learning research to education. We now elaborate on these and introduce additional applications. We describe two applications of cognitive phenomena to education (the spacing effect and the testing effect), and we describe two applications of cognitive theories to education (cognitive load and cognitive tutors). The insights gained from translational research are of value to both educational psychology and cognitive science. The outcomes of educational interventions, much like the results of laboratory experiments, provide data that support or refute theories of learning and memory.

The Spacing Effect

The temporal distribution of practice affects retention (for a review, see Cepeda et al., 2006). When study devoted to a single item is uninterrupted, learning is *massed*. Alternatively, when measurable time or intervening items separate study opportunities, learning is *spaced*. Massed practice speeds learning, whereas spaced practice enhances retention. This is called the *spacing effect*. In a study by Kapler et al. (Kapler, Weston, & Wiseheart, 2015), undergraduates learned meteorological concepts during a lecture and completed an online review 1 or 8 days later. They were tested on concepts 35 days after the online review. Students' final recollection increased with the duration of time between the initial lecture and the online review. Furthermore, their ability to use studied material to answer higher level application questions increased with the spacing of study.

According to one theory, the spacing effect arises from a *study-phase retrieval* process (Bjork, 1994). Retrieval of past exposures, prompted by study, strengthens the original memory trace. The change in memory strength depends on the difficulty of retrieval: the more difficult the (successful) retrieval, the greater the gain. Because item retrieval is initially more difficult when practice is spaced, spaced practice is most beneficial. Pavlik and Anderson (2005) created a computational model of the spacing effect that relates to this theory. In their model, each repetition of an item contributes to the item's overall activation in memory. Each repetition's contribution decays at a different rate, which is determined by the item's activation at the time when the repetition occurred. If an item has low activation at the time of study, additional repetitions are stored with low decay. When practice is spaced, an item's activation rises slowly, and subsequent repetitions are stored with low decay. Alternatively, when practice is massed, an item's activation rises quickly, and repetitions are stored with higher decay. Over long retention intervals, the low decay causes spaced items to remain more accessible than items that were practiced in a massed fashion.

The spacing effect has clear education applications. To enhance retention, practice should be spaced. Practically, this means that students and teachers should periodically review earlier course material. Additionally, examples should be interleaved within a study session to create spaces between repetitions of the same concept. Computational cognitive models can be used to make even more precise recommendations about when course content should be reviewed and how practice repetitions should be interleaved to maximize retention (Lindsey, Shroyer, Pashler, & Mozer, 2014; Pavlik & Anderson, 2008). These models can be used to design study schedules for groups of students or to adapt review based on an individual's experience and performance.

The Testing Effect

Tests are generally conceived as instruments that measure learning, not as instruments that can change learning. Yet research has demonstrated that taking tests as a means of studying can significantly enhance memory. This phenomenon is called the *testing effect* or *test-enhanced learning*. The canonical study design involves the following phases: (1) participants are exposed to the new material, (2) they are randomly assigned to conditions involving additional study or practice tests on the material, and (3) they are given post-tests to measure their memory for the material. Across many studies that have explored variations on this experimental paradigm, participants who take tests consistently outperform those who engage in additional study (for reviews, see Pyc, Agarwal, & Roediger, 2014; Rawson & Dunlosky, 2011; Roediger & Butler, 2011).

A notable feature of the testing effect literature is the phenomena's robustness. The testing effect holds regardless of the format of the testing intervention (e.g., multiple choice, short answer, or essay; Kang, McDermott, & Roediger, 2007; McDaniel, Roediger, & McDermott, 2007). Similarly, the testing effect does not depend on whether the practice and final tests match in format (Carpenter & DeLosh, 2006). These and other experimental manipulations have led researchers to conclude that retrieval from memory is the key mechanism underlying the testing effect (Carpenter & Pashler, 2007; Putnam & Roediger, 2013).

Other manipulations mediate the benefits of practice tests. For example, although feedback is not necessary for the testing effect to occur (e.g., Kornell, Hays, & Bjork, 2009), presenting the correct answer after each practice test trial further enhances memory at the final test (Butler, Karpicke, & Roediger, 2008; Pashler, Cepeda, Wixted, & Rohrer, 2005). Additionally, the benefits of practice tests increase with the duration of the retention interval and may even be absent when the final test is immediate (Roediger & Karpicke, 2006).

Across all these results, it is natural to see educational applications of the testing effect. Providing students with low-stakes or practice tests is one way to implement test-enhanced learning in the classroom. Similarly, students who use flashcards or otherwise find ways to explicitly test themselves during study can benefit from the testing effect. Several studies have confirmed that these results apply outside the laboratory (McDaniel, Agarwal, Huelser, McDermott, & Roediger, 2011; Pennebaker, Gosling, & Ferrell, 2013) At the same time, incorporating test-enhanced learning in education can lead to other (indirect) benefits as well. For example, by taking multiple practice tests, students are naturally spacing their practice, rather than cramming (see previous subsection). Additionally, students who test themselves may become more self-aware of their strengths and weaknesses and hence improve their metacognitive abilities.

Cognitive Load Theory

Although the involvement of working memory in cognitive performance has long been known (e.g., Baddeley & Hitch, 1974), the theory of *cognitive load* was developed to capture related effects on *learning*, especially in problem-solving domains (Sweller, 1988).[2] Cognitive load is the "demand for working memory resources of a particular learner [imposed] by specific cognitive tasks or activities" (Lee & Kalyuga, 2014, p. 32). For example, when a learner solves a problem in an unfamiliar domain, demands on working memory may stem from a combination of (a) comprehending the problem, (b) applying domain-general strategies to solve the problem, and (c) learning problem-solving schemas for the domain. When the sum of these demands exceeds the limits of the learner's working memory, learning is impaired. *Cognitive load theory* (CLT) is concerned with the nature of working memory demands, their impact on learning, and instructional strategies that leverage cognitive load to optimize learning.

Early CLT research highlighted the need to reduce cognitive load in order to improve learning. For instance, the *worked example effect* is the finding that students who solve a series of problems (for instructional purposes) are slower and less accurate during a post-test than were those who receive the same problems but alternate between studying worked examples and solving (Sweller & Cooper, 1985). CLT explains this effect in terms of the demands of problem-solving taking resources away from learning. The solution is to reduce cognitive load by presenting worked examples. Similarly, the *split-attention effect* (Chandler & Sweller, 1991, 1992; Moreno & Mayer, 2000; Ward & Sweller, 1990) occurs when an instructional display requires learners to integrate disparate information, thus imposing greater working memory demands and producing poorer learning outcomes. As with the worked example effect, the design heuristic is the same: reduce the cognitive load of instruction to improve learning outcomes.

Further research into cognitive load, however, has led to refinement of the theory and nuances to its application. Most notably, CLT now distinguishes different types of load and their effects on learning (Sweller, Van Merriënboer, & Paas, 1998). *Intrinsic load* describes the inherent working memory demands of a particular task and is considered difficult to change. *Extraneous load* describes extra demands imposed by instruction and should be reduced through careful design. Last, *germane load* describes demands from added cognitive processing that is productive for learning, even if it goes above and beyond the basic requirements of the task. These distinctions help to explain why *increasing* load can sometimes benefit learning. For example, asking students to self-explain as they study worked examples (Bielaczyc, Pirolli, & Brown, 1995; Chi, DeLeeuw, Chiu, & LaVancher, 1994; Lovett, 1992) or having them identify errors in incorrect worked examples (Booth, Lange, Koedinger, & Newton, 2013) introduces additional germane load and produces greater learning gains compared to studying worked examples alone.

The refined CLT also helps to explain the *expertise reversal effect* (Kalyuga et al., 2003; Lee & Kalyuga, 2014) in which an instructional design technique that reduces cognitive load and enhances learning for novices may do the opposite—increase load and harm learning—for "experts" (high-knowledge learners). For example, even though providing worked examples benefits novice learners by reducing the working memory demands associated with solving problems, this support can introduce extraneous load (and impede learning) for high-knowledge learners by causing them to process the extra information and try to align it with what they already know. Thus, the expertise reversal effect emphasizes the importance of considering learners' knowledge level when creating instructional interventions.

ACT and Cognitive Tutors

A notable application of cognitive science research to education is *intelligent tutoring systems*. When students work with an intelligent tutoring system, they receive hints and feedback as they solve practice problems at a computer-based interface. Intelligent tutoring systems generally employ AI techniques to solve target problems in parallel with students (enabling evaluation of the students' responses) and to interact with students in sophisticated ways (e.g., via natural language). In addition, intelligent tutoring systems incorporate psychological theory to enhance their design for student learning. For example, *cognitive tutors* (Anderson, Boyle, & Reiser, 1985; Anderson, Conrad, & Corbett, 1989) represent a form of intelligent tutoring system based on the ACT* theory (Anderson, 1983), which posits that complex task performance can be deconstructed into independent components of procedural knowledge—that is, production rules.

Even though cognitive tutors were originally developed with a key goal being to test the ACT theory outside the lab, they have significantly benefited students in a variety of contexts (Anderson, Corbett, Koedinger, & Pelletier, 1995). Studies have shown that cognitive tutors lead to greater learning gains (Koedinger, Anderson, Hadley, & Mark, 1997; Lovett, 2001; Shute, 1995), greater learning efficiency (Anderson et al., 1989), or both (VanLehn, 2011). Recently, in a randomized, controlled study among 147 school sites, schools that adopted cognitive tutors for algebra I showed significant improvement in students' post-test scores compared to matched schools that continued with the current algebra curriculum (Pane, Griffin, McCaffrey, & Karam, 2014).

Cognitive tutors—along with intelligent tutoring systems in general—have also been leveraged as a cognitive science research platform. They naturally support cognitive science research on learning because they enable (a) collection of rich learning data from students in real classes; (b) instructional interventions to be deployed experimentally at the school, class, student, or (within-student) topic level; and (c) integration with other state-of-the-art computational tools for studying learning and behavior. For example, cognitive tutors have been designed to foster metacognition and motivation through instructional interventions that support help-seeking, self-explanation, and positive affect (Aleven & Koedinger, 2002; Aleven, McLaren, Roll, & Koedinger, 2006; Rodrigo et al., 2012). Intelligent tutoring systems have been used to collect clickstream and other data for automated sensors that detect affective states related to learning (Arroyo et al., 2014; Baker, Corbett, & Koedinger, 2004). Finally, tutoring systems have been further instrumented to track (and respond to) other sources of information such as eye movements (D'Mello, Olney, Williams, & Hays, 2012; Gluck & Anderson, 2001) and changes in neural activation (Anderson, Betts, Ferris, & Fincham, 2010).

Conclusion and Future Directions

Research on learning and memory dates back to the late 19th century, yet it remains a vibrant and dynamic field of study. A survey of the literature 10 years hence might look very different from our current review. We now shift focus from the classic perspectives and findings that have defined the field to the emerging trends that are giving rise to new ideas about learning and memory. Some of these trends are driven by theoretical advances (i.e., biologically inspired cognitive models), some by the development of new methods and sources of data (i.e., learning analytics and educational data mining), and some by a combination of both (i.e., educational neuroscience).

Educational Neuroscience

Neuroscience has shed light on the mental processes involved in learning and memory. These insights may advance our understanding of instruction and learning, and, in doing so, improve education. Most basically, neuroscience research has demonstrated the plasticity of the brain throughout adolescence and into adulthood (Lövdén, Bäckman, Lindenberger, Schaefer, & Schmiedek, 2010). This underscores the potential for lifelong learning. Neuroscience has also contributed to our understanding of how factors such as nutrition, exercise, and sleep impact the brain and learning (for a review, see Sigman, Peña, Goldin, & Ribeiro, 2014). These findings point to simple but powerful interventions.

Neuroimaging techniques such as EEG and fMRI reveal information about the neural basis of language and mathematics. For example, during language processing, speech sounds (along with other linguistic features) cause a characteristic sequence of voltage deflections in scalp-recorded event-related potentials (ERPs). The amplitudes and latencies of certain ERPs, recorded at birth, are predictive of whether an individual will later develop dyslexia (Molfese, 2000). In this way, neuroimaging techniques provide a tool for early screening and intervention. Neuroimaging techniques have also been used to track student performance. For instance, Anderson and Fincham (2014) used fMRI to discover the sequence of mental states that occurred as individuals solved math problems. Based on variations in state durations, they could predict errors before they occurred. Although fMRI is not suitable for classroom application, portable EEG systems have also been used to detect students' mental states as they complete education tasks (Xu, Chang, Yuan, & Mostow, 2014). This sort of monitoring capability could be harnessed to improve the responsiveness of intelligent tutoring systems.

Learning Analytics and Educational Data Mining

With advances in educational technology, instruction is increasingly occurring online. As with cognitive tutors and intelligent tutoring systems, this creates a platform for instrumenting and studying learning outside the laboratory. Whereas the research in intelligent tutoring systems was largely motivated by the goal of testing cognitive theories, the focus of early work in learning analytics and educational data mining has been on analyzing the new (and often large) datasets coming from a variety of online learning environments that are now available. Learning analytics is "the measurement, collection, analysis, and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs" (LAK '11, 2011). Similarly educational data mining is "concerned with developing methods for exploring the unique types of data that come from educational settings, and using those methods to better understand students, and the settings in which they learn" (JEDM, 2015). Both fields are interdisciplinary and involve integrating research across many of the same fields that participate in cognitive science research.

Work in this area has already led to new methods and discoveries about learning and has yielded practical tools for making educational technologies more intelligent and adaptive (Baker & Yacef, 2009). Looking ahead, the future of learning analytics is in effectively integrating data-driven and theory-driven approaches to understand and improve learning, especially in the context of massively open online courses (MOOCs), educational games, and motivational and affective issues.

Biologically Inspired Computational Models

Computational cognitive models grew out of the AI tradition (Newell & Simon, 1972). As cognitive science matured, modelers increasingly adopted algorithmic constraints—assumptions about the limits of cognitive processing. Yet they remained agnostic about the biological basis of cognition. The current movement is toward increased biological realism; that is, consideration of how the brain actually implements cognitive processes.

An example of this is the attempt to map ACT-R modules to neural regions (Anderson, 2007). By comparing time-varying activity in modules with activation observed in the brain, architectural modules can be localized to specific regions. Activation in those regions, or the onset of other predefined brain signatures, can then be used to infer the sequence of covert mental operations taking place in a task (Anderson, Zhang, Borst, & Walsh, 2016). This approach allows us to ask new questions, such as which stage(s) of a task are most affected by practice (Tenison & Anderson, 2016)?

Another example is Leabra, a cognitive architecture built in the PDP framework (O'Reilly, 1998; see also O'Reilly, Hazy, & Herd, this handbook). Backpropagation, although a powerful and widely used learning algorithm, is inconsistent with known properties of neurobiology. Leabra uses biologically plausible learning algorithms instead. Low-level processing units in Leabra are organized into an architecture consisting of a posterior cortex, a prefrontal cortex, and a hippocampus. Complex human behavior emerges from interactions among the low-level units and the high-level regions that make up the architecture.

Parting Thoughts: The Cognitive Science Approach

Cognitive science entails a multidisciplinary perspective. The benefit of such a perspective is exemplified in research on learning and memory. Work in various disciplines, including experimental psychology, education, AI, and neuroscience, has led to discoveries about learning and memory. The insights gained are valuable to each field of inquiry. Consider four examples from this chapter:

- *Laboratory experiments demonstrated that spaced practice produces superior retention*. Subsequent applications of spacing strategies in the classroom improved educational outcomes and, in the process, furnished additional evidence for computational models of learning and memory. The same is true of laboratory studies and classroom applications of testing effect research.
- *Cognitive tutors are based on computational student models*. In addition to enabling intelligent tutoring systems to tailor guidance to individual learners, cognitive tutors provide tests of psychology theories outside the laboratory.
- *Patient studies revealed a dissociation between structures involved in the acquisition and long-term storage of declarative knowledge (the medial temporal lobe and the associative cortex, respectively)*. A subsequent model showed that such a dissociation would allow the human memory system to reinstate memories in an interleaved fashion in order to extract similarities across ensembles of experience.
- *Patient studies revealed yet another set of structures within the basal ganglia involved in the acquisition of procedural knowledge*. The distinction between procedural and declarative knowledge is part of computational models of learning and memory and is supported by evidence from behavioral experiments.

As we set out to write this review, we were struck by the diversity of research on learning and memory, both in terms of topics and methodologies. However, far from existing as a fractionated field of unrelated facts, the various research themes, when integrated, support a consistent view of learning and memory.

Acknowledgments

We thank Katharine Walsh for providing simulating conversation and comments on this topic, and John Anderson, Terence Andre, and Stu Rodgers for giving constructive feedback on an earlier version of this chapter.

Notes

1. Contingencies of reinforcement need not be of an all-or-nothing nature. The strength of reinforcement in maintaining behavior depends critically on the schedule of reinforcement; that is, whether reinforcement is delivered on a fixed or variable basis (fixed/variable), and whether it depends on number of responses or elapsed time (ration/interval) (for a review, see Fester & Skinner, 1957).

2. Consistent with a cognitive science approach, Sweller's introduction of cognitive load theory included a computational model to demonstrate the theory's implications for learning.

References

Adams, C. D., & Dickinson, A. (1981). Instrumental responding following reinforcer devaluation. *Quarterly Journal of Experimental Psychology B: Comparative and Physiological Psychology*, *33B*, 109–121.

Aleven, V. A., & Koedinger, K. R. (2002). An effective metacognitive strategy: Learning by doing and explaining with a computer-based Cognitive Tutor. *Cognitive Science*, *26*(2), 147–179.

Aleven, V., McLaren, B., Roll, I., & Koedinger, K. (2006). Toward meta-cognitivetutoring: A Model of help-seeking with a Cognitive Tutor. *International Journal of Artificial Intelligence in Education*, *16*(2), 101–128.

Anderson, J. R. (1982). Acquisition of cognitive skill. *Psychological Review*, *89*(4), 369–406.

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

Anderson, J. R. (1991). The adaptive nature of human categorization. *Psychological Review*, *98*(3), 409–429.

Anderson, J. R. (2007). *How can the human mind occur in the physical universe?* New York: Oxford University Press.

Anderson, J. R., Betts, S., Ferris, J. L., & Fincham, J. M. (2010). Neural imaging to track mental states while using an intelligent tutoring system. *Proceedings of the National Academy of Sciences*, *107*(15), 7018–7023.

Anderson, J. R., Boyle, C. F., & Reiser, B. J. (1985). Intelligent tutoring systems. *Science*, *228*, 456–462.

Anderson, J. R., Conrad, F. G., & Corbett, A. T. (1989). Skill acquisition and the LISP Tutor. *Cognitive Science*, *13*(4), 467–506.

Anderson, J. R., Corbett, A. T., Koedinger, K. R., & Pelletier, R. (1995). Cognitive tutor: Lesson learned. *Journal of the Learning Sciences*, *4*(2), 167–207.

Anderson, J. R., & Fincham, J. M. (2014). Discovering the sequential structure of thought. *Cognitive Science*, *38*, 322–352.

Anderson, J. R., Fincham, J. M., & Douglass, S. (1999). Practice and retention: A unifying analysis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(5), 1120.

Anderson, J. R., & Schooler, L. J. (1991). Reflections of the environment in memory. *Psychological Science*, *2*(6), 396–408.

Anderson, J. R., Zhang, Q., Borst, J. P., & Walsh, M. M. (2016). The measurement of processing stages: Extension of Sternberg's method. http://psycnet.apa.org/psycarticles/2016-21447-001

Arroyo, I., Woolf, B. P., Burleson, W., Muldner, K., Rai, D., & Tai, M. (2014). A multimedia adaptive tutoring system for mathematics that addresses cognition, metacognition and affect. *International Journal of Artificial Intelligence in Education*, *24*(4), 387–426.

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence and J. T. Spence (Eds.), *The psychology of learning and motivation*, *2* (pp. 89–195). New York, NY: Academic.

Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 8, pp. 47–89). New York, NY: Academic Press.

Bahrick, H. P. (1979). Maintenance of knowledge: Questions about memory we forgot to ask. *Journal of Experimental Psychology: General*, *108*(3), 296–308.

Bahrick, H. P. (1984). Semantic memory content in permastore: Fifty years of memory for Spanish learned in school. *Journal of Experimental Psychology: General*, *113*(1), 1–29.

Bahrick, H. P., & Hall, L. K. (1991). Lifetime maintenance of high school mathematics content. *Journal of Experimental Psychology: General*, *120*, 20–33.

Baker, R. S., Corbett, A. T., & Koedinger, K. R. (2004). Detecting student misuse of intelligent tutoring systems. *Proceedings of the 7th International Conference on Intelligent Tutoring Systems*, 531–540.

Baker, R. S. J. D., & Yacef, K. (2009). The state of educational data mining in 2009: A review and future visions. *Journal of Educational Data Mining*, *1*, 3–16.

Balleine, B. W., & Dickinson, A. (1998). Goal-directed instrumental action: Contingency and incentive learning and their cortical substrates. *Neuropharmacology*, *37*(4), 407–419.

Bielaczyc, K., Pirolli, P., & Brown, A. L. (1995). Strategy training in self-explanation and self-regulation strategies for learning computer programming. *Cognition and Instruction*. *13*(2), 221–252.

Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe & A. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 185–205). Cambridge, MA: MIT Press.

Booth, J. L., Lange, K. E., Koedinger, K. R., & Newton, K. J. (2013). Using example problems to improve student learning in algebra: Differentiating between correct and incorrect examples. *Learning and Instruction*, *25*, 24–34.

Butler, A. C., Karpicke, J. D., & Roediger, H. L. (2008). Correcting a metacognitive error: Feedback increases retention of low-confidence correct responses. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(4), 918–928.

Carpenter, S. K., & DeLosh, E. L. (2006). Impoverished cue support enhances subsequent retention: Support for the elaborative retrieval explanation of the testing effect. *Memory and Cognition*, *34*(2), 268–276.

Carpenter, S. K., & Pashler, H. (2007). Testing beyond words: Using tests to enhance visuospatial map learning. *Psychonomic Bulletin & Review*, *14*(3), 474–478.

Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin*, *132*(3), 354–380.

Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction*, *8*(4), 293–332.

Chandler, P., & Sweller, J. (1992). The split-attention effect as a factor in the design of instruction. *British Journal of Educational Psychology*, *62*(2), 233–246.

Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology*, *4*(1), 55–81.

Chi, M., DeLeeuw, N., Chiu, M., & LaVancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science*, *18*(3), 439–477.

Chi, M. T., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science*, *5*(2), 121–152.

Craik, F. I., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General*, *104*(3), 268–294.

Custers, E. J., & ten Cate, O. T. (2011). Very long-term retention of basic science knowledge in doctors after graduation. *Medical Education*, *45*(4), 422–430.

D'Mello, S., Olney, A., Williams, C., & Hays, P. (2012). Gaze tutor: A gaze-reactive intelligent tutoring system. *International Journal of Human-Computer Studies*, *70*(5). 377–398.

Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques. Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest*, *14*(1), 4–58.

Ebbinghaus, H. (1885/1913). *Memory: A contribution to experimental psychology* (H. A. Ruger & C. E. Bussenues, Trans.). New York, Teachers College, Columbia University.

Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review*, *100*(3), 363–406.

Fester, C. B., & Skinner, B. F. (1957). *Schedules of reinforcement*. New York: Appleton-Century-Crofts.

Fitts, P. M. (1964). Perceptual-motor skill learning. In A. W. Melton (Ed.), *Categories of human learning* (pp. 381–391). New York, NY: Academic Press.

Friedman, M. P., Burke, C. J., Cole, M., Keller, L., Millward, R. B., & Estes, W. K. (1964). Two-choice behavior under extended training with shifting probabilities of reinforcement. In R. C. Atkinson (Ed.), *Studies in mathematical psychology* (pp. 250–316). Stanford, CA: Stanford University Press.

Fu, W. T., & Anderson, J. R. (2006). From recurrent choice to skill learning: A reinforcement-learning model. *Journal of Experimental Psychology: General*, *135*(2), 184–206.

Gluck, K. A., & Anderson, J. R. (2001). What role do cognitive architectures play in intelligent tutoring systems? In D. Klahr & S. M. Carver (Eds.), *Cognition & instruction: Twenty-five years of progress* (pp. 227–262). Mahwah, NJ: Erlbaum.

Guttman, N., & Kalish, H. I. (1956). Discriminability and stimulus generalization. *Journal of Experimental Psychology*, *51*(1), 79–88.

Hilgard, E. R., & Bower, G. H. (1975). *Theories of learning*. New York, NY: Appleton-Century-Crofts

Hull, C. L. (1943). *Principles of behavior: An introduction to behavioral theory*. New York, NY: Appleton-Century-Crofts.

JEDM. (2015). Call for papers. *Journal of Educational Data Mining*. Retrieved October 30, 2015, from http://www.educationaldatamining.org/JEDM/index.php/JEDM.

Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The expertise reversal effect. *Educational Psychologist*, *38*(1), 23–31.

Kang, S. H. K., McDermott, K. B., & Roediger, H. L. (2007). Test format and corrective feedback modify the effect of testing on long-term retention. *European Journal of Cognitive Psychology*, *19*(4-5), 528–558.

Kapler, I. V., Weston, T., & Wiseheart, M. (2015). Spacing in a simulated undergraduate classroom: Long-term benefits for factual and higher-level learning. *Learning and Instruction*, *36*, 38–45.

Klein, G. (2008). Naturalistic decision making. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *50*(3), 456–460.

Knowlton, B. J., Mangels, J. A., & Squire, L. R. (1996). A neostriatal habit learning system in humans. *Science*, *273*(5280), 1399–1402.

Koedinger, K. R., Anderson, J. R., Hadley, W. H., & Mark, M. A. (1997). Intelligent tutoring goes to school in the big city. *International Journal of Artificial Intelligence in Education*, *8*, 30–43.

Koedinger, K. R., Booth, J. L., & Klahr, D. (2013). Instructional complexity and the science to constrain it. *Science*, *342*(6161), 935–937.

Kornell, N., Hays, M. J., & Bjork, R. A. (2009). Unsuccessful retrieval attempts enhance subsequent learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(4), 989–998.

Laird, J. (2012). *The Soar cognitive architecture*. Cambridge, MA: MIT Press.

LAK'11 (2011). Call for papers. First International Conference on Learning Analytics and Knowledge 2011. Retrieved October 30, 2015, from https://tekri.athabascau.ca/analytics/.

Lee, C. H., & Kalyuga, S. (2014). Expertise reversal effect and its instructional implications. In V. A. Benassi, C. E. Overson, & C. M. Hakala (Eds.), *Applying science of learning in education: Infusing psychological science into the curriculum* (pp. 31–44). Retrieved from http://teachpsych.org/ebooks/asle2014/index.php

Lindsey, R. V., Shroyer, J. D., Pashler, H., & Mozer, M. C. (2014). Improving students' long-term knowledge retention through personalized review. *Psychological Science*, *25*(3), 639–647.

Logan, G. D., & Klapp, S. T. (1991). Automatizing alphabet arithmetic: I. Is extended practice necessary to produce automaticity? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*, 179–195.

Lövdén, M., Bäckman, L., Lindenberger, U., Schaefer, S., & Schmiedek, F. (2010). A theoretical framework for the study of adult cognitive plasticity. *Psychological Bulletin*, *136*(4), 659–676.

Lovett, M. C. (1992). Learning by problem solving versus by examples: The benefits of generating and receiving information. *Proceedings of the Fourteenth Annual Conference of the Cognitive Science Society*. (pp. 956–961). Hillsdale, NJ: Erlbaum.

Lovett, M. C. (2001). A Collaborative convergence on studying reasoning processes: A case study in statistics. In S. Carver, & D. Klahr (Eds.), *Cognition and instruction: Twenty-five years of progress* (pp. 347–384). Mahwah, NJ: Erlbaum.

Lovett, M. C., & Anderson, J. R. (1996). History of success and current context in problem solving. Combined influences on operator selection. *Cognitive Psychology*, *31*(2), 168–217.

Lovett, M. C., & Schunn, C. D. (1999). Task representations, strategy variability, and base-rate neglect. *Journal of Experimental Psychology: General*, *128*(2), 107–130.

Mayer, R. E. (2008). Applying the science of learning: Evidence-based principles for the design of multimedia instruction. *American Psychologist*, *63*(8), 760–769.

McClelland, J. L. (2009). The place of modeling in cognitive science. *Topics in Cognitive Science*, *1*(1), 11–38.

McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*, *102*(3), 419–457.

McClelland, J. L., & Rogers, T. T. (2003). The parallel distributed processing approach to semantic cognition. *Nature Reviews Neuroscience*, *4*(4), 310–322.

McDaniel, M. A., Agarwal, P. K., Huelser, B. J., McDermott, K. B., & Roediger, H. L. (2011). Test-enhanced learning in a middle school science classroom: The effects of quiz frequency and placement. *Journal of Educational Psychology*, *103*(2), 399–414.

McDaniel, M. A., Roediger, H. L., & McDermott, K. B. (2007). Generalizing test-enhanced learning from the laboratory to the classroom. *Psychonomic Bulletin & Review*, *14*(2), 200–206.

Mitchell, T. (2006). *The discipline of machine learning* (Technical Report CMU–ML-06-108). Pittsburgh, PA: Carnegie Mellon University.

Molfese, D. L. (2000). Predicting dyslexia at 8 years of age using neonatal brain responses. *Brain and Language*, *72*(3), 238–245.

Moreno, R., & Mayer, R. E. (2000). A coherence effect in multimedia learning: The case for minimizing irrelevant sounds in the design of multimedia instructional messages. *Journal of Educational Psychology, 92*(1), 117–125.

Newell, A. (1990). *Unified theories of cognition*. Cambridge, MA: Cambridge University Press.

Newell, A., & Simon, H. A. (1972). *Human problem solving* (Vol. 104, No. 9). Englewood Cliffs, NJ: Prentice-Hall.

Nosofsky, R. M., Palmeri, T. J., & McKinley, S. C. (1994). Rule-plus-exception model of classification learning. *Psychological Review*, *101*(1), 53–79.

O'Reilly, R. (1998). Six principles for biologically-based computational models of cortical cognition. *Trends in Cognitive Sciences*, *2*(11), 455–462.

Packard, M. G., & McGaugh, J. L. (1996). Inactivation of hippocampus or caudate nucleus with lidocaine differentially affects expression of place and response learning. *Neurobiology of Learning and Memory*, *65*(1), 65–72.

Pane, J. F., Griffin, B. A., McCaffrey, D. F., & Karam, R. (2014). Effectiveness of cognitive tutor algebra I at scale, *Educational Evaluation and Policy Analysis*, *36*(2), 127–144.

Pashler, H., Cepeda, N. J., Wixted, J. T., & Rohrer, D. (2005). When does feedback facilitate learning of words? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(1), 3–8.

Pavlik, P. I., & Anderson, J. R. (2005). Practice and forgetting effects on vocabulary memory: An activation-based model of the spacing effect. *Cognitive Science*, *29*(4), 559–586.

Pavlik, P. I., & Anderson, J. R. (2008). Using a model to compute the optimal schedule of practice. *Journal of Experimental Psychology: Applied*, *14*(2), 101–117.

Pavlov, I. P. (1927). *Conditioned reflexes: An investigation of the physiological activity of the cerebral cortex*. (Translated and edited by G. V. Anrep). London: Oxford University Press.

Pennebaker, J. W., Gosling, S. D., & Ferrell, J. D. (2013). Daily online testing in large classes: Boosting college performance while reducing achievement gaps. *PLoS ONE 8*(11): e79774.

Pirolli, P. L., & Anderson, J. R. (1985). The role of practice in fact retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *11*, 136–153.

Putnam, A. L., & Roediger, H. L. (2013). Does response mode affect amount recalled or the magnitude of the testing effect? *Memory and Cognition*, *41*(1), 36–48.

Pyc, M. A., Agarwal, P. K., & Roediger, H. L. (2014). Test-enhanced learning. In V. A. Benassi, C. E. Overson, & C. M. Hakala (Eds.), *Applying science of learning in education: Infusing psychological science into the curriculum*. Retrieved from http://teachpsych.org/ebooks/asle2014/index.php

Rachlin, H. (1970). *Introduction to modern behaviorism*. San Francisco: W. H. Freeman.

Rawson, K. A., & Dunlosky, J. (2011). Optimizing schedules of retrieval practice for durable and efficient learning: How much is enough? *Journal of Experimental Psychology: General*, *140*(3), 283–302.

Rodrigo, M. M. T., Baker, R. S. J. d., Agapito, J., Nabos, J., Repalam, M. C., Reyes, S. S., & San Pedro, M. C. Z. (2012). The effects of an interactive software agent on student affective dynamics while using an intelligent tutoring system. *IEEE Transactions on Affective Computing*, *3*(2), 224–236.

Roediger, H. L., & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences*, *15*(1), 20–27.

Roediger, H. L., & DeSoto, K. A. (2014). Confidence and memory: Assessing positive and negative correlations. *Memory*, *22*(1), 76–91.

Roediger, H. L., & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, *17*(3), 249–255.

Rohrer, D., & Taylor, K. (2007). The shuffling of mathematics problems improves learning. *Instructional Science*, *35*(6), 481–498.

Ruiz, D., & Newell, A. (1989). Tower-noticing triggers strategy-change in the Tower of Hanoi: A Soar model. In *Proceedings of the Eleventh Annual Conference of the Cognitive Science Society* (pp. 522–529). Hillsdale, NH: Lawrence Erlbaum Associates.

Rumelhart, D. E., & Todd, P. M. (1993). Learning and connectionist representations. In D. E. Meyer & S. Kornblum (Eds.), *Attention and performance XIV: Synergies in experimental psychology, artificial Intelligence, and cognitive neuroscience* (pp. 3–30). Cambridge, MA: MIT Press.

Schmidt, R. A. (1988). *Motor control and learning: A behavioral emphasis*. Champaign, IL: Human Kinetics.

Schoenfeld, A. H., & Herrmann, D. J. (1982). Problem perception and knowledge structure in expert and novice mathematical problem solvers. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *8*(5), 484–494.

Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science*, *237*(4820), 1317–1323.

Shute, V. J. (1995). SMART: Student modeling approach for responsive tutoring. *User Modeling and User-Adapted Interaction*, *5*(1), 1–44.

Siegler, R. S. (1999). Strategic development. *Trends in Cognitive Sciences*, *3*(11), 430–435.

Sigman, M., Peña, M., Goldin, A. P., & Ribeiro, S. (2014). Neuroscience and education: Prime time to build the bridge. *Nature Neuroscience*, *17*(4), 497–502.

Skinner, B. F. (1938). *The behavior of organisms: An experimental analysis*. Cambridge, MA: B. F. Skinner Foundation.

Squire, L. R., Knowlton, B., & Musen, G. (1993). The structure and organization of memory. *Annual Review of Psychology*, *44*(1), 453–495.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science 12*(2), 257–285.

Sweller, J., & Cooper, G. A. (1985). The use of worked examples as a substitute for problem solving in learning algebra. *Cognition and Instruction 2*(1), 59–89.

Sweller, J., Mawer, R. F., & Ward, M. R. (1983). Development of expertise in mathematical problem solving. *Journal of Experimental Psychology: General*, *112*(4), 639–661.

Sweller, J., Van Merriënboer, J., & Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, *10*(3), 251–296.

Taatgen, N. A., & Anderson, J. R. (2002). Why do children learn to say "broke"? A model of learning the past tense without feedback. *Cognition*, *86*(2), 123–155.

Tenison, C., & Anderson, J. R. (2016). Modeling the distinct phases of skill acquisition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 42, 749-767.

Thorndike, E. L. (1905). *Elements of psychology*. New York, NY: A. G. Seiler.

Tolman, E. C. (1932). Purposive behavior in animals and men. New York, NY: Century.

VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems and other tutoring systems. *Educational Psychologist*, *46*(4), 197–221.

Walsh, M. M., & Anderson, J. R. (2009). The strategic nature of changing your mind. *Cognitive Psychology*, *58*, 416–440.

Walsh, M. M., & Anderson, J. R. (2011). Modulation of the feedback-related negativity by instruction and experience. *Proceedings of the National Academy of Sciences*, *108*(47), 19048–19053.

Walsh, M. M., & Anderson, J. R. (2013*a*). Electrophysiological responses to feedback during the application of abstract rules. *Journal of Cognitive Neuroscience*, *25*(11), 1986–2002.

Walsh, M. M., & Anderson, J. R. (2013*b*). Navigating complex decision spaces: Problems and paradigms in sequential choice. *Psychological Bulletin*, *140*, 466–486.

Walsh, M. M., & Gluck, K. A. (2015*a*). Mechanisms for robust cognition. *Cognitive Science*, *39*(6), 1131–1171.

Walsh, M. M., & Gluck, K. A. (2015*b*). Verbalization of decision strategies in multiple-cue probabilistic inference. *Journal of Behavioral Decision Making*, *29*, 78-91.

Ward, M., & Sweller, J. (1990). Structuring effective worked examples. *Cognition and Instruction*, *7*(1), 1–39.

Wickens, C. D., Hutchins, S., Carolan, T., & Cumming, J. (2013). Effectiveness of part-task training and increasing-difficulty training strategies: A meta-analysis approach. *Human Factors*, *55*(2), 461–470.

Xu, Y., Chang, K. M., Yuan, Y., & Mostow, J. (2014). EEG helps knowledge tracing! In *Proceedings of the 12th International Conference on Intelligent Tutoring Systems Workshop on Utilizing EEG Input in Intelligent Tutoring Systems*, Honolulu, Hawaii.