2013

Metamemory

John Dunlosky
Kent State University

Keith W. Thiede
Boise State University

"Metamemory" by John Dunlosky and Keith W. Thiede (pp. 283-298) is from *The Oxford Handbook of Cognitive Psychology*, edited by Daniel Reisberg (2013), is reproduced by permission of Oxford University Press. doi: 10.1093/oxfordhb/9780195376746.013.0019

CHAPTER

19 Metamemory

John Dunlosky *and* Keith W. Thiede

Abstract

Metamemory refers to people's beliefs about their memory and to how people monitor and control their learning and retrieval. In this chapter, we describe monitoring and control processes involved in learning and retrieval, how these processes have been measured, and key outcomes relevant to human metamemory. Based on these outcomes, general conclusions include the following: (a) people's judgments of their memory are based on a variety of cues; hence (b) judgment accuracy arises from the diagnosticity of the cues, so that above-chance accuracy of any metamemory judgment only arises when the available cues are predictive (or diagnostic) of criterion performance; and finally, (c) people use their memory judgments to guide their study and retrieval. Thus, people's memory monitoring plays a pivotal role in the effectiveness of their self-regulated learning and retrieval, so a major aim of metamemory research is to discover techniques that yield high levels of judgment accuracy and optimal regulation.

Key Words: metamemory, self-regulated learning, memory monitoring, control, judgments of learning, feeling of knowing, confidence judgments, metacognition

Metamemory refers to people's thoughts about their memory and how memory operates. Although the term *metamemory* may seem esoteric to some, people rely on their metamemory as they perform many activities. When heading to a grocery store, one person may believe that he can remember all 10 items that need to be purchased, whereas another person with the same items to purchase may believe that her memory will fail and hence decides to take a list along. An eyewitness may point the finger at the accused, and do so with extreme confidence that the accused had committed the crime; however, such high confidence can be illusory, such as when confidence is inadvertently based on a memory of seeing the accused in a lineup instead of actually witnessing the accused commit the crime (for real-life examples, see Loftus & Ketcham, 1991). Unfortunately, jurors believe the testimony of a highly confident eyewitness, regardless of whether his or her confidence is well placed. On a lighter note, when playing games (such as Trivial Pursuit), players may withhold answers when they believe that those answers are wrong, yet they will try to persuade their team to respond with an answer when their confidence in it is high. And when failing to generate answers to some questions, an emotionally charged tip-of-the-tongue state may arise when they believe the sought-after answer is available in memory. Being a student can put an even higher premium on the effective use of metamemory, because to learn efficiently, students must be able to accurately judge how well key concepts have been learned and make appropriate decisions about which concepts require further study.

This short list illustrates several metamemory processes that can occur while people are learning new materials or are attempting to retrieve old ones. Moreover, they illustrate the multifaceted nature of metamemory, which includes knowledge (and beliefs) about memory, monitoring of memory, and control of memory. *Metamemory knowledge* refers to declarative knowledge or beliefs that an individual holds about how memory operates and whether those beliefs are accurate. In the earlier example, the individual who decided to write down the grocery list may know from experience that memory tends to fail after four or five items from a grocery list need to be remembered. The individual who went without a list may believe that he has a very good memory—that is, he has high memory self-efficacy—and hence will be able to recall even lengthier lists without having difficulties.

Memory monitoring refers to assessing progress during learning or the current state of a previously studied item. For example, when studying for an upcoming test, students may attempt to monitor and evaluate their ongoing progress while studying. Their monitoring of memory yields a confidence judgment about whether they will remember the key concepts on the upcoming exam. Finally, *memory control* involves regulating any aspect of learning or retrieval. For the control of learning, one example of a control process includes deciding to spend more time studying materials that one believes have not been well learned. For retrieval, accusing a defendant of committing a crime is a prime example, because the eyewitness allegedly must be highly confident in his or her memory of the crime to potentially condemn the defendant.

Based on these examples, it may be evident why so many people have become interested in metamemory: Faulty metamemory can lead to poor memory and low achievement, whereas accurate metamemory can enhance memory and achievement. For instance, if students inaccurately assess that they are ready for an upcoming test, then they may prematurely stop studying. In this case, poor monitoring leads to a nonoptimal control decision, which in turn would lead to less-than-expected performance on the examination. To expand on this example, consider two students who are studying for an upcoming test in a class of introductory psychology. Both students are trying to learn the core concepts relevant to memory, such as what is short-term memory, long-term memory, and encoding specificity. Both students judge that they will be able to remember about 90% of the concepts, which they believe is fine for the test, so they stop studying to join each other for a late-night movie. Whereas Julie actually will retain 90% of the concepts (her judgments were accurate), Mike was highly overconfident and will remember only 50%. In this case, Mike was overconfident and hence he stopped studying before he met his desired learning goal; unfortunately, Mike will not perform well on the test, and he may even tell the teacher, "But I thought I knew all these concepts. Why did I earn such a poor grade on the exam?"

Although the importance of accurate metamemory may be intuitive, research in this area did not begin in earnest until about 1970 when John Flavell coined the term "metamemory." In the 1970s, the term "meta" began to appear in articles and in conference papers, and groundbreaking research was also conducted by John Flavell, Ellen Markman, and Ann Brown, among others. Perhaps most important for solidifying a specialized field of metamemory was Flavell's (1979) *American Psychologist* article, called "Metacognition and Cognitive Monitoring: A New Area of Cognitive-Developmental Inquiry." This article has been highly influential because it is here where Flavell defines core concepts for the field, such as metacognitive knowledge and metacognitive experiences (which are most closely aligned with metamemory knowledge and memory monitoring, as described earlier). As important, he developed numerous testable hypotheses about how the development of metamemory in childhood would in turn influence the developmental progress of core cognitive processes.

John Flavell and his colleagues—most notably, Henry Wellman—first captured our attention with persuasive arguments about the importance of metamemory, but it was Joseph Hart who provided the first method to assess metamemory in an objective manner. One can best understand Hart's breakthrough within the context of the introspection method that was used in the infancy of psychological research. In the early 1900s, a researcher interested in associations may ask a (trained) participant to introspect about the psychological processes that produce a free association. For instance, what is the first word that comes to mind when you read "guitar"? Perhaps you think "Stratocaster," but if you were an introspectionist, you would also need to describe the cognitive processes that preceded the thought "Stratocaster." Doing so may seem difficult, but participants in these studies (at times the experimenters themselves) were highly practiced. A downfall of this method was that the introspective reports were viewed as a window into the mind; that is, the reports were assumed by some to be accurate and complete. This view is most evident in R. S. Woodworth's (1921) definition of introspection as the "observation by an individual of his own conscious action... Notice that it is a form of observation, and not speculation or reasoning from past experience. It is a direct observation of fact" (p. 10). This definition emphasizes introspection as a metacognitive act, because introspection involves directly observing a mental action. Unfortunately, even in 1901, it was evident that introspective methods would fall well short of direct observation of mental actions involved in a response (for a historical review, see Humphrey, 1951).

Six decades later, Joseph Hart introduced a method to systematically explore distortions in people's introspections about their memory. In particular, participants were asked general-information questions. When they did not answer a question, they were simply asked to predict whether they could choose the correct answer on a multiple-choice recognition test. These predictions—which are called *feeling-of-knowing* (FOK) judgments—are introspective reports, but in contrast to earlier introspection research, Hart (1965) did not assume that they were valid. Instead, he evaluated their accuracy against objective performance by administering recognition tests. By doing so, people's judgments could be validated against their performance: When they were in a tip-of-the-tongue state and were sure they knew the correct answer even when they could not retrieve it, would they then recognize the correct answer? Hart (1965, 1966) found that people's FOK judgments showed above-chance accuracy. That is, their judgments were higher for correctly recognized answers than for ones that were not recognized.

Hart's methods focused on the monitoring of retrieval, as measured by FOK judgments. The method is invaluable because it has allowed researchers to systematically explore the biases in people's FOK judgments, which we consider in some detail later. As important, his methods can be readily applied to evaluate how people monitor all aspects of learning—from study through retrieval. And extensions of his methods have been used to explore people's control of study and retrieval. In the remainder of this chapter, we first provide a bird's eye view of the kinds of monitoring and control processes investigated in the field. We then discuss current theory pertaining to two widely investigated metamemory judgments relevant to study and retrieval: judgments of learning (JOLs) and FOK judgments, respectively.

In 1990, Nelson and Narens unified the field of metamemory by organizing the various monitoring and control processes into a single framework. An expanded version of this framework is presented in Figure 19.1, which illustrates measures of memory monitoring and memory control that correspond to each phase of learning. For instance, during study ("acquisition" in Fig. 19.1), one may monitor memory for to-be-learned items, which is measured by having people judge how well an item has been learned. This JOL may also influence the control of study, such as by informing people's decisions about when to terminate study of a given item. The measures depicting memory monitoring are presented in the top portion of Figure 19.1, and the measures depicting memory control are presented in the bottom portion. Definitions of each measure are in Table 19.1.

Modern research on metamemory largely focuses on answering just a few questions about monitoring and control. Concerning memory monitoring, one primary question is, How do people monitor various phases of learning? Answers to this question are obtained by investigating how people make the various monitoring judgments. For instance, JOLs are investigated to understand how people monitor their learning, whereas confidence judgments and FOK judgments are investigated to understand how people monitor their retrieval. The accuracy of these judgments is often of central interest, and taking the lead from Hart (1965), accuracy is measured by comparing a given judgment to its corresponding criterion measure. JOLs (which are predictions of future performance) are compared to future performance, FOK judgments for predicting future recognition of currently unrecallable information are compared to future recognition performance, and confidence judgments are compared to performance on the criterion test being judged. These and other judgments often demonstrate a positive correlation with criterion performance. For instance, Souchay, Moulin, Clarys, Taconnat, and Isingrini (2007) had older (M age = 72 years) and younger (M age = 27) adults attempt to answer general-information questions and then make an FOK judgment for any question they did not answer. After this judgment phase, the

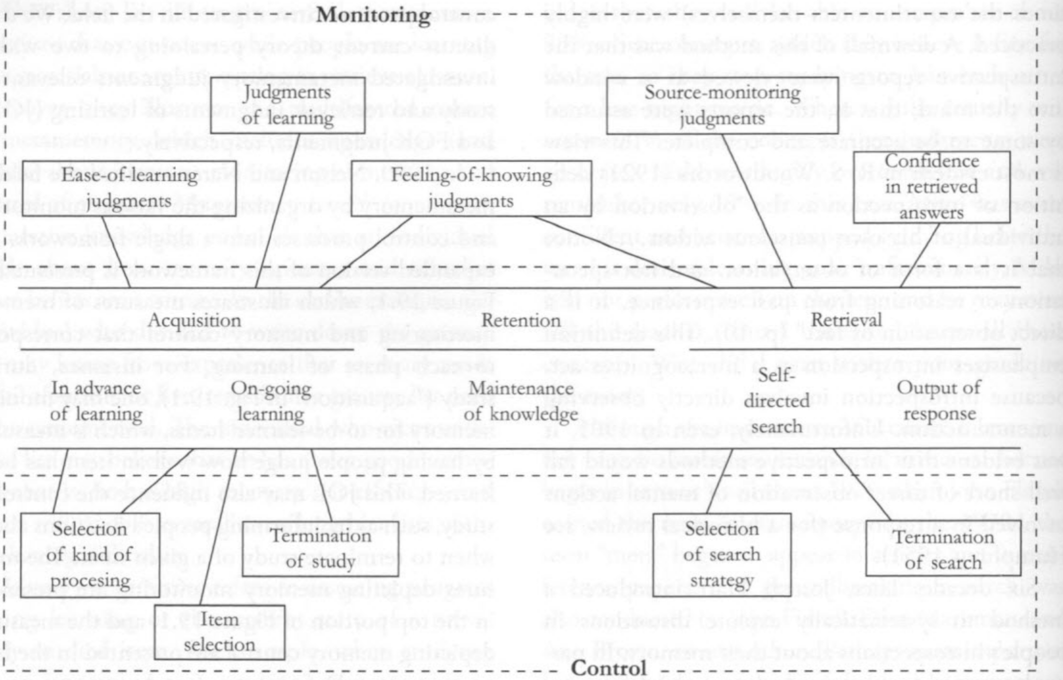

Figure 19.1 Overview of metamemory monitoring and control components that occur throughout study (acquisition), retention, and retrieval. (Adapted from Nelson & Narens, 1990, to include judgments that were absent from their original framework.)

Table 19.1 Names and Common Definitions of Metacognitive Judgments and Control Processes

Name	Definition
Metacognitive judgments	
Ease-of-learning (EOL) judgments	Judgments of how easy to-be-studied items will be to learn.
Judgments of learning (JOLs)	Judgments of the likelihood of remembering recently studied items on an upcoming test
Feeling-of-knowing (FOK) judgments	Judgments of the likelihood of recognizing currently unrecallable answers on an upcoming test
Source-monitoring judgments	Judgments made during a criterion test pertaining to the source of a particular memory
Confidence in retrieved answers	Judgments of the likelihood that a response on a test is correct. Often referred to as retrospective confidence (RC) judgments
Control processes	
Selection of kind of processing	Selection of strategies to employ when attempting to commit an item to memory
Item selection	Decision about whether to study an item on an upcoming trial
Termination of study	Decision to stop studying an item currently being studied
Selection of search strategy	Selecting a particular strategy in order to produce a correct response during a test
Termination of search	Decisions to terminate searching for a response

Source: From Dunlosky, J., Serra, M., & Baker, J. M. C. (2007). Metamemory applied. In F. Durso et al. (2nd ed.) *Handbook of applied cognition*. New York: Wiley.

participants completed a recognition test for each of the unrecalled questions. Means across intraindividual correlations between FOK judgments and recognition performance were .35 for younger adults and .46 for older adults, indicating that people of all ages can accurately predict future recognition performance for currently unrecallable information (for extended primers on metamemory measurement and analyses, see Dunlosky & Bjork, 2008; Dunlosky & Metcalfe, 2009).

Concerning memory control, a primary question is, How do people use monitoring to control their memory? To address this question, researchers often examine the relationship between a measure of monitoring and a measure of control. For instance, to examine control of study, JOLs made during an initial study trial may be compared to subsequent self-paced study time (Mazzoni, Cornoldi, & Marchitelli, 1990). JOLs are often negatively related to later self-paced study time—that is, learners tend to spend more time studying information that has been less well learned (lower JOLs) than better learned (higher JOLs). As we discuss later, explaining this negative relationship between JOLs and self-paced study times—and exceptions to it-has become a principal focus of current theories of self-paced study.

Theory and Data of Basic Monitoring and Control Processes

The literature on metamemory includes basic research on metamemory knowledge, monitoring, and control; analyses of changes in metamemory that occur throughout the life span; and the neuroscience and neuropsychology of monitoring and control processes (Dunlosky & Metcalfe, 2009). Given the size of this literature, our overview here must be limited in scope, so we have decided to focus largely on two questions: How do people make metamemory judgments during study and retrieval? And how do people control their study and retrieval? To answer them, we could rely on the literature of almost any of the measures presented in Figure 19.1. Presently, however, we consider only JOLs and FOK judgments, because the literature in both cases is extensive and offers some basic principles that apply to all metamemory judgments. A more thorough introduction to the literature on each of the monitoring and control functions in Figure 19.1 is provided by Dunlosky and Metcalfe (2009) and Dunlosky, Serra, and Baker (2007).

Monitoring of Study and Retrieval

HOW DO PEOPLE MAKE JUDGMENTS OF LEARNING AND FEELING-OF-KNOWING JUDGMENTS?

Some of the earliest theoretical work on the bases of FOK judgments explored the degree to which they directly measure memory. According to this *direct-access view*, people's judgments reflect the underlying strength of an item in memory (cf. trace-access mechanisms in Nelson, Gerler, & Narens, 1984). So, if you made a JOL after you had just studied "dog—spoon," your judgment would directly tap how well the association between "dog" and "spoon" was stored in memory. Such direct access echoes the early introspectionists' belief that self-evaluations arise from a "direct observation of fact," and perhaps not surprisingly, predictions from the direct-access view have been disconfirmed with regard to JOLs and FOK judgments.

Let's just consider a disconfirmation of this view from the JOL literature, which will also serve to introduce cue-based accounts for metamemory judgments. Benjamin, Bjork, and Schwartz (1998) had participants answer general-knowledge questions that were moderate to easy in difficulty level, so participants could generate an answer to most questions. Immediately after they generated an answer, participants made a JOL concerning the likelihood that they would recall the answer on a test of free recall; for this test, they were given a blank sheet of paper and had to recall only the previously retrieved answers (without the question cues). JOLs were related to the fluency of initial retrieval—the faster they could generate an answer to a general-knowledge question, the more likely they judged that they would be able to freely recall it on the criterion test. By contrast, the opposite was true for the criterion test; that is, the faster the answer was retrieved during the initial phase, the less likely they were to recall it during the free-recall criterion test. Thus, people's JOLs did not directly access how well each target answer was stored in memory. Instead, JOLs tracked retrieval fluency: As they retrieved answers more quickly during the initial test, they judged the answer would be easier to remember, even though this assessment did not reflect actual storage strength (as assessed by free recall on the criterion test).

This outcome, among many others, suggests that people use processing fluency as a cue to make JOLs. In fact, people tend to use fluency for almost all metamemory judgments when processing fluency differs across to-be-judged items (Alter

& Oppenheimer, 2009). This conclusion not only pertains to retrieval fluency (as in Benjamin et al., 1998) but other forms of processing fluency as well. Another example involves fluency of generating encoding strategies. In one study (Hertzog, Dunlosky, Robinson, & Kidder, 2003), college students studied word pairs and were asked to generate an interactive image to associate both words. During study, participants pressed the space bar once they had generated an image. The time from an item's presentation and this key press was a measure of the fluency of image generation. Across three experiments, encoding fluency was negatively related to JOLs; that is, the more fluently participants generated an image (faster latency), the higher people's judgment that they would correctly recall the item. Even subtle aspects of the stimulus environment can influence encoding fluency and JOLs. For instance, Rhodes and Castel (2009) manipulated the fluency of perceptually processing items. Participants listened to items that were presented at different volumes and made a JOL after each item was presented. JOLs were higher for items presented loudly than those presented quietly, presumably because the former were more fluently processed. Thus, both encoding fluency and retrieval fluency influence people's JOLs.

People also use fluency when making FOK judgments. The potential influence of this cue is most evident in Koriat's (1997) accessibility hypothesis. According to this hypothesis, when people make an FOK judgment for an unrecalled item, it is based on the information that they had accessed while searching for an answer. So, if you are asked, "What city is the capital of Washington State?" you may fail to come up with an answer, but in trying to retrieve one from memory, you may remember "emerald," that it begins with an "s," and that it is in the far northwest portion of the state. For another question, "Who wrote *Moll Flanders*?" you do not respond but think the author's name begins with a "D." A prediction from this hypothesis is that the more information you access—and the more fluently you access it—prior to making an FOK judgment, the higher your FOK is expected to be. So you may judge that you will absolutely recognize the capital of Washington State but will unlikely recognize the name of the author who wrote *Moll Flanders*.

Importantly, Koriat (1993) proposed that the quality of what is accessed is irrelevant—even if a lot of incorrect information is fluently accessed, then a high FOK will result. In the present example, note that all the accessed information about Washington's capital pointed toward Seattle (the "Emerald City"), when in fact, the correct answer is Olympia. By contrast, "D" is the first letter of the author's first and last names (Daniel Defoe). Nevertheless, according to the accessibility hypothesis, the sheer amount retrieved and the fluency of retrieval are most influential, so FOK judgments are expected to be greater for the question about Washington's capital, even though the partial information retrieved about it was incorrect. In a series of creative studies, Koriat (1993) demonstrated the powerful influence of accessibility on FOK judgments. In one experiment, participants studied tetragrams—four-letter strings that were nonwords, such as RFSC and FKRD. A tetragram was briefly presented for study, which was followed by a short retention interval and a retrieval attempt. After the retrieval attempt, an FOK judgment was made for that tetragram, and then after all tetragrams were studied and judged, a recognition test was administered so that FOK accuracy could be measured. Most relevant for now, people's FOK judgments were related to how many letters were recalled prior to making the judgment, even when those letters were not in the studied tetragram. Put differently, FOK judgments were based on how much information was accessed, regardless of whether the information recalled was correct.

We have focused almost exclusively on people's use of fluency as a basis for their judgments. More generally, however, people appear to base their judgments on many cues that appear to distinguish between items in a manner that may be relevant to criterion performance. For instance, another cue that is available when people make FOK judgments is the familiarity of the cue used to prompt the retrieval attempt. The cue for "What city is the capital of Washington State?" is the question itself. For a variety of reasons (e.g., perhaps you recently read an article about Washington State), you may find the cue itself familiar and hence judge that you are likely to recognize the correct answer. Research has established that FOK judgments are influenced by cue familiarity in this manner (Metcalfe, Schwartz, & Joaquim, 1993; Miner & Reder, 1994).

Given that metamemory judgments can be based on multiple cues, researchers are beginning to explore how various cues are integrated into a judgment. For FOK judgments, Koriat and Levy-Sadot (2001) proposed that cue familiarity has the first influence. If your familiarity with a cue is low, then you would spend little, if any, time attempting to retrieve the correct answer and your FOK would be low. However, if you are familiar with the cue, then you would search for an answer, and any information you accessed during this search would in turn influence your FOK judgment. That is, accessibility is expected to have a greater influence on FOK judgments when cue familiarity is high than when it is low. For instance, if you were asked, "Who is the mayor of Ravenna, Ohio?" you would likely have no familiarity with this particular cue and quickly respond with the lowest FOK judgment without even attempting to search for the answer (cf. Kolers & Palef, 1976). However, if you were asked, "Who was the second president of the United States?" if you had familiarity with the presidents of the United States, then you would likely search for the correct answer. While searching, any information that you accessed along the way would then influence your subsequent FOK judgments. Across three experiments, Koriat and Levy-Sadot (2001) provided evidence consistent with this interactive hypothesis. Moreover, Benjamin (2005) demonstrated how cue familiary and target accessibility both influence people's JOLs, but the time course for their influence differs. When JOLs are made under time pressure, cue familiarity has an influence. Without time pressure, people have more time to attempt retrieval, and hence this cue has a larger influence on people's JOLs.

To conclude this section, we will return to our initial question, How do people make JOLs and FOK judgments? Processing fluency and cue familiarity have a joint influence on these judgments, but they also can be influenced by any number of cues that are available when the judgments are made. For instance, in some experiments on JOLs, participants study paired associates consisting of either unrelated (dog—spoon) or related (king—crown) words. JOLs are much higher for related than unrelated items (e.g., Carroll, Nelson, & Kirwan, 1997). In some cases, people's JOLs are influenced by the serial position of items on a list, with higher judgments being made for items in the first positions (the primacy items) than for those in the middle of the list (Castel, 2008; Dunlosky & Matvey, 2001).

Given the number of cues that do influence JOLs, it is perhaps surprising that some relatively obvious ones have a rather minimal (or inconsistent) influence. For instance, when studying paired associates, people can use numerous strategies, such as interactive imagery (for "dog—spoon," imaging a dog paddling in a large spoon filled with milk) or repetition (repeating "dog—spoon" together during study). The strategies differ in their effectiveness, with final recall performance typically being much higher after interactive imagery than rote repetition. When people judge items studied under these two different strategies, however, their JOLs typically do not differ for items that were studied using imagery rather than repetition (Rabinowitz, Ackerman, Craik, & Hinchley, 1982; Shaughnessy, 1981). Likewise, other studies have demonstrated that people's JOLs can be insensitive to retention interval and to the number of learning trials (e.g., Carroll et al., 1997; Koriat, Bjork, Sheffer, & Bar, 2004; Kornell & Bjork, 2008, respectively), which often have a substantial influence on criterion recall performance.

Why are some cues influential (e.g., fluency), whereas other cues (e.g., encoding strategy or number of learning trials) are not? One answer is based on the assumption that all judgments are inferential. That is, people use various cues to infer the likelihood of correctly performing on the criterion test. What makes one cue more potent than another can partly be ascertained from Koriat and colleagues' (Koriat, Nussinson, Bless, & Shaked, 2008) inference-based account that emphasizes a distinction between whether the source of an inference is experience based (EB) or information based (IB). IB (or theory-based) judgments rely on people's declarative knowledge and beliefs about how a particular cue influences memory. For instance, as noted earlier, JOLs do not distinguish between items that are studied by imagery versus rote repetition, and this null effect presumably arises because most people do not know that imagery is a more effective strategy. Once people are given experience using both strategies, however, they obtain declarative knowledge that "imagery is better," and in subsequent trials, JOLs (and other metmamemory judgments) are higher after imagery than repetition (Hertzog, Price, & Dunlosky, 2008).

EB judgments involve two processing stages: "first a process that gives rise to a sheer subjective feeling and second a process that uses that feeling as a basis for memory predictions" (Koriat et al., 2008, p. 118). The idea here is that various cues trigger feelings, which in turn boost metamemory judgments. According to Koriat et al. (2008), EB judgments for JOLs rely on the ease with which items are encoded or retrieved during learning, and EB judgments for FOKs rely on the fluency of accessing partial information when searching for a target answer. Put differently, differences in the fluency of processing presumably give rise to different feelings that then influence people's metamemory judgments.

This taxonomy does provide some insight into why various cues influence—or do not influence—people's metamemory judgments. For instance, if a participant does not believe a particular cue will influence memory (IB) and if the cue does not give rise to a metacognitive experience (EB), then it should not influence the judgments. This situation presumably holds for different encoding strategies that most participants would not have experience using in a novel associative task, such as imagery and repetition. Another possibility is that a cue which produces a strong metacognitive experience (e.g., processing fluency) could override the influence of other IB cues. Thus, when people study "archer—tree" and then make a JOL for it, they may be captured by the ease of developing an image for the word pair (and hence make a high JOL), and altogether ignore the fact that the word pair is slated to be tested weeks after study (and hence should probably receive a low JOL). In this way, even obvious IB cues (e.g., retention interval) may not influence JOLs when other strong EB cues are available in the environment. This taxonomy is useful for explaining the influence of various cues on metamemory judgments, and it has intuitive appeal. Even so, the taxonomy is not entirely orthogonal, because people apparently have theories about fluency cues; that is, learners believe that fluently processed information is easier to remember. So this EB cue (fluency) may actually have its influence on metamemory judgments vis-à-vis IB processing (as in Matvey, Dunlosky, & Guttentag, 2001).

WHEN WILL METAMEMORY JUDGMENTS BE ACCURATE?

Hart's groundbreaking research demonstrated that the accuracy of people's FOK judgments was above chance, and the seminal work on JOLs by Arbuckle and Cuddy (1969) also revealed that they had above-chance accuracy. Why? One straightforward answer is offered by cue-based accounts of metamemory judgments, which are themselves founded on Egon Brunswik's lens model of perceptual judgments (Brunswik, 1956). These accounts highlight the importance of the three relationships among a cue, judgment, and criterion performance—which are illustrated in Figure 19.2. The relationship between cue and judgment is referred to as *cue utilization*, which represents the degree to which a cue influences (or is related to) the judgment. As described earlier, cue utilization for processing fluency is usually high for metamemory judgment. A second relationship occurs between

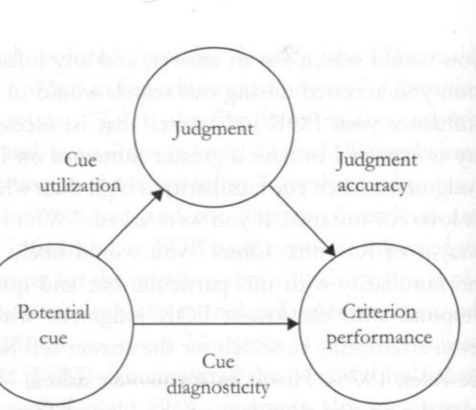

Figure 19.2 Relationships among cues, judgments, and recall that emphasize the central role of cue utilization and diagnosticity in determining judgment accuracy.

judgment and criterion performance, which represents *judgment accuracy*. Finally, the relationship between cue and performance is referred to as *cue diagnosticity* and represents the degree to which a particular cue predicts criterion test performance (cf. ecological validity, Brunswik, 1956).

According to the cue-based accounts, judgment accuracy is a function of cue utilization and cue diagnosticity, and hence, above-chance accuracy occurs whenever (a) a cue influences (or is related to) a judgment and (b) the cue itself accurately predicts—or rather, is diagnostic of—criterion performance. Above-chance accuracy is not ensured, because even when a cue is used to make a judgment, the cue itself may have nil diagnosticity or even negative diagnosticity. In the latter case, the judgments would show below-chance accuracy, with higher confidence in performance (e.g., higher JOLs) being indicative of *lower* criterion performance. In fact, the most impressive support for such inference-based accounts of accuracy comes from studies in which the accuracy of people's judgments was zero or negative.

One excellent example comes from Benjamin et al. (1998), which was described earlier. Recall that participants used initial fluency of retrieving answers to general-information questions as a cue to make JOLs for predicting subsequent free recall of the answers. In this case, the cue of retrieval fluency was *negatively* related to cued recall: The more quickly they retrieved answers to general-information questions (the cue), the less likely they would later recall the answers on the criterion test. Such negative cue diagnosticity in turn led to inaccurate JOLs. The general point here is that when people use a cue, its diagnosticity will in part drive judgment accuracy.

Thus, for Benjamin et al. (1998), if retrieval fluency on the first test had been predictive of performance on the second (criterion) test, then JOL accuracy should be above chance. This prediction has received empirical support. In particular, when the first test is a cued recall test (as in Benjamin et al., 1998) but the second criterion test is *also* cued recall, then (a) JOLs are still based on the fluency of retrieval on the first test, but (b) now this cue of retrieval fluency is predictive of final recall performance. As the cue is diagnostic of criterion performance, JOL accuracy is relatively high (Serra & Dunlosky, 2005).

This example also suggests one way that the accuracy of JOLs and FOK judgments can be improved or fine-tuned to match levels of performance. Namely, to assess memory well, one must evaluate it in contexts that yield diagnostic cues. For instance, when studying paired associates (e.g., dog—spoon), two kinds of JOLs can be made. An *immediate* JOL occurs immediately after an item is studied. So one would study "dog—spoon," and then ask, "Will I later recall spoon when shown "dog" on the test?" Immediate JOLs are not highly accurate because the cues available for the JOL (e.g., processing fluency) usually are not highly predictive of criterion performance. By contrast, a *delayed* JOL occurs after other items have been studied, so when the prompt for JOL is presented (i.e., "dog—?"), one must try to recall the response from memory (i.e., "spoon"). For paired associates, the outcome of this delayed retrieval attempt is highly diagnostic of future memory performance, so delayed JOLs are highly accurate (Nelson & Dunlosky, 1991).

Another avenue to improve judgment accuracy is to provide people with experience about the diagnosticity of various cues. For instance, as noted earlier, people's JOLs do not differ for paired associates studied by imagery more than for those studied by rote repetition, even though this cue (strategy used) is highly diagnostic of performance. That is, criterion recall performance is much higher after imagery than repetition. When people are given experience using these strategies during an initial study-test trial and then make JOLs on a second study-test trial in which imagery and repetition are used, people's JOLs begin to favor imagery items (Hertzog et al., 2008). Thus, even if people's JOLs and FOK judgments are not initially accurate, they can be tuned to better reflect memory performance (for further discussion of a variety of tactics to improve judgment accuracy, see Dunlosky & Metcalfe, 2009).

THE BASES AND ACCURACY OF OTHER METAMEMORY JUDGMENTS

Although we have focused exclusively on JOLs and FOK judgments, the ideas described earlier pertain to any of the judgments listed in the top half of Figure 19.1. Each of the judgments is presumably inferential in nature in that people infer the likelihood of criterion performance based on available cues. Even so, they are not always empirically correlated because they also differ in other ways. Concerning their intercorrelations, Nelson and Leonesio (1988) had participants make multiple judgments during a multiphase experiment—including ease-of-learning (EOL) judgments, JOLs, and FOK judgments. Within individual participants, the judgments were not highly correlated: +.19 between EOL judgments and JOLs, +.12 between EOL judgments and FOKs, and +.17 between JOLs and FOK judgments. One explanation for these small intercorrelations follows from cue-based accounts of metamemory judgments, because the cues available can differ as people judge different aspects of memory. As the cues differ, the relationship between the judgments will diverge; however, when the same cues are available for different judgments, the judgments will show a stronger relationship.

These points can be illustrated with a simple experiment. Participants study 40 paired associates; 20 pairs consist of related words (king—crown) and 20 consist of unrelated words (turtle—bean). Before studying any pairs, participants are shown each pair and asked to make this ease-of-learning (EOL) judgment: How difficult will this item be to learn (0 = difficult to 100 = easy)? Then, each pair is presented individually for 6 seconds, and a JOL is made (0 = 0% likelihood of recall to 100 = 100% likelihood). Finally, a cued recall test is administered across all pairs, and for those in which a correct response could not be recalled, participants made a FOK judgment for a subsequent recognition test. Notice that some cues are similar but others differ across the judgments. For EOL judgments and JOLs, the relatedness of pairs is available as a cue and it has a major influence: Both judgments are higher for related than unrelated pairs. In this case, the judgments are highly correlated (e.g., +.71 and +.65 in two different conditions; Dunlosky & Matvey, 2001). By contrast, note that pair relatedness is not available when people make FOK judgments, because these judgments are made when people cannot retrieve the response to a pair. Thus, the cue of pair relatedness would not be available for FOK judgments, and hence other cues would be used to infer the likelihood that the correct answers will be recognized. This context more closely matches the method used by Nelson and Leonesio (1988), which yielded low correlations.

Given the usefulness of a cue-based account for understanding metamemory judgments, a general agenda for research has been to identify (a) the cues that are available when any particular judgment is made, (b) the degree to which people use a given cue when making a judgment, (c) the diagnosticity of the available cues, and, of course, (d) the degree to which the available cues support above-chance accuracy. Any particular research study typically only explores a subset of these issues, but the entire enterprise has consistently led to converging conclusions about all the metamemory judgments listed in Table 19.1. In particular, when monitoring our memories, we do not possess the skill to *directly* assess how well a memory has been stored or how accurately it can be retrieved, and hence the accuracy of people's metamemory judgments is not ensured. Instead, judgment accuracy depends on the available cues, the degree to which those cues are used in constructing a metamemory judgment, and the diagnosticity of the cues for predicting criterion performance.

Control of Memory

A major focus of metamemory research concerns how people control their study and retrieval processes. In the next two sections, we review the extensive literature on these control processes. The main issues that we address here are whether and how monitoring is used to make control decisions.

HOW DO LEARNERS CONTROL THEIR STUDY?

To begin answering this question, we describe a standard method used to investigate the control of study. Participants are asked to learn a list of word pairs, and during an initial trial, each pair (e.g., cat—boat) is presented at a fixed rate (e.g., 2 seconds an item). Participants then make a JOL for each item. Afterward, they have another chance to study each pair. During this self-paced study trial, participants may be shown all the pairs in an array (such as in a textbook page of foreign language translation equivalents), and then they select pairs that they want to restudy. Alternatively, each pair may be presented individually, and participants study each pair as long as they want. A key question is, Do people use their monitoring (as measured by JOLs) to make decisions about which items to select for restudy or to decide how long to study each item?

The modal (and nearly universal) finding from the earliest studies using this method was a negative relationship between JOLs and subsequent study time (either item selection or self-paced study). That is, as people rated items as more difficult to remember (lower JOLs), they spent more time studying (for reviews, see Son & Kornell, 2008; Son & Metcalfe, 2000). This relationship suggests that people do use monitoring to control their study time. Nevertheless, it is correlational, and hence some other variable may be responsible for why JOLs are related to the allocation of study time. To evaluate whether JOLs are causally related to allocation, Metcalfe and Finn (2008) used a variant of the aforementioned method. On an initial study trial, participants either studied pairs (cat—boat) once or three times; a cued recall test then occurred (e.g., cat—?), and of course, recall was greater for pairs presented thrice than once. Next, a second study trial occurred in which items initially presented once were now presented three times, whereas items initially presented three times were presented only once. After each pair was studied (either once or thrice), participants made a JOL for the pair and then made a decision about whether to restudy the pair.

On this judgment-selection trial, memory performance for the two sets of items was statistically equivalent, because all items were studied four times, either being studied once during the first trial and three times during the second, or vice versa. Thus, when people made decisions about which items to restudy, on average, memory across all items was the same. However, people's JOLs were different for the two sets of items: On the second trial, people made higher JOLs for items that were *initially* studied three times on the first trial than for those studied only once. Note that JOLs made on the second study trial are inaccurate here: People believe they will remember items studied more time on the first trial, when in fact, all items were studied the same number of times when JOLs were made, so recall did not differ for them. Thus, if people's JOLs drive study-time allocation, then they should more often select to restudy those items studied once on the first trial, because people believe those items have been less well learned. Across three experiments, Metcalfe and Finn (2008) demonstrated that people use their JOLs to control their allocation of study time (see also, Hines, Touron, & Hertzog, 2009).

Now that we have established that people do use their monitoring of learning to guide study, an intriguing issue becomes how it is used in allocating study time. According to the discrepancy-reduction model, people set a goal for learning items, and then continue studying each one until the goal is met. That is, learners continue studying an item until the discrepancy between current learning and the goal is completely reduced. As it takes longer to reduce this discrepancy for more difficult-to-learn items than easier ones, people use more time studying the more difficult items. In this manner, the discrepancy-reduction model predicts the modal outcome in which people spend more time studying items given lower than higher JOLs (Son & Metcalfe, 2000, 2005).

Although discrepancy reduction may underlie some study behaviors (Benjamin & Bird, 2006), people's allocation of study time is more flexible. We do more than (perhaps mindlessly) continue studying items until they all meet the same preset goal. In some situations, learners use monitoring to prioritize the *easiest* items for study and do not even study the more difficult items (Son & Metcalfe, 2000; Thiede & Dunlosky, 1999), and in other situations, they abandon memory monitoring in deciding how to allocate study time across items (Ariel, Dunlosky, & Bailey, 2009). Let's consider the former case first. The modal outcome (i.e., more study to items judged as most difficult) has typically been found when participants are encouraged to learn all the items on a list. To learn them all (i.e., meet a mastery goal), more time would likely need to be spent studying the more difficult items. In some situations, however, learners do not have mastery goals: Perhaps they have little time to study and cannot master all the items, or perhaps they do not desire to master the list (e.g., students who only need to pass a final exam to obtain their desired grade). If learners are trying to efficiently meet their goals, in both cases, they should focus on the easiest items, because using extra effort on the more difficult ones would be a waste of time. Consistent with such possibilities, when study time is limited (e.g., 15 seconds to study 30 items) and when people set low performance goals (e.g., learn only 6 of 30 items), they spend the most time studying a subset of the easiest-to-learn items (Dunlosky & Thiede, 2004; Son & Metcalfe, 2000).

Note that in these cases, people are using monitoring of learning to decide which items should be selected and prioritized for study. Beyond monitoring, people also tend to focus their restudy on items that they expect will provide the most reward. For instance, college students more often select to study (a) items that are more likely (than less likely) to appear on the criterion test and (b) items that will be worth more points (or reward) if correctly recalled on the criterion test (Ariel et al., 2009; Thiede & Dunlosky, 1999). In both cases, college students prioritize the more highly valued items for study, regardless of whether they are difficult or easy to learn.

Thus, students often appear to attempt to maximize their performance in the most efficient manner, which is a central premise of the *agenda-based regulation* (ABR) framework of study-time allocation (Ariel et al., 2009; Dunlosky, Ariel, & Thiede, 2011). According to the ABR framework, people construct and execute agenda—or simple plans—for allocating study time to meet these dual criteria: meeting current task goals and doing so efficiently. Although this framework highlights the importance of agendas, factors other than agenda construction and execution are expected to influence how (and how well) learners allocate their study time. For instance, individual differences in working-memory abilities can limit the quality of agenda use that leads to the dysregulation of study time (Dunlosky & Thiede, 2004). Bottom-up processes that are triggered by prepotent responses to the stimulus environment also can influence study time (Rhodes & Castel, 2009).

The ABR framework is an example of just one of several theories of people's allocation of study time (see also, Koriat, Ma'ayan, & Nussinson, 2006; Metcalfe, 2009; Winne & Hadwin, 1998). These theories differ with respect to their explanatory scope and are not mutually exclusive. For instance, Metcalfe's (2009) *region-of-proximal-learning* framework emphasizes how students in some cases prioritize the easiest of the unlearned items for study, because focusing on those items presumably would increase the likelihood that students would efficiently obtain their learning goals (Kornell & Metcalfe, 2006). In the context of the ABR framework, prioritizing items within this region of proximal learning constitutes just one of many agendas that people can use to allocate their time across items. Given that research on self-regulated learning is largely in its infancy, new evidence will likely help to shape these recent theories as we move toward a general theory of study-time allocation that can both describe how people control their study as well as prescribe how they should control study to efficiently meet their goals.

CONTROL OF OTHER FACETS OF LEARNING

Researchers have just begun exploring how students control other facets of learning, such as making decisions to mass or space their study (Son, 2004; Toppino, Cohen, Davis, & Moors, 2009) or whether they elect to test themselves during study (Karpicke, 2009; Kornell & Bjork, 2007). Deciding how (or whether) to use these study strategies is important, because they can boost performance if used properly. For instance, memory performance is often better when study of each item is spaced throughout a study session than when it is massed together; fortunately, when given the choice, students often decide to space their study (Pyc & Dunlosky, 2010). Memory performance also benefits from testing; in fact, testing oneself on previously studied items typically is better for memory than is restudying the same items again (Roediger & Karpicke, 2006). Unfortunately, college students tend to underuse testing as a study strategy (Kornell & Bjork, 2007). Understanding why students appropriately use some effective strategies and underuse others is becoming an important goal for metamemory research, partly because it has obvious implications for improving student scholarship.

HOW DO PEOPLE CONTROL THEIR RETRIEVAL?

A great way to begin answering this question involves asking another one: Which country won the World Cup in 2006? If you are a soccer fan, you would be familiar with the World Cup. Thus, if you cannot recall the answer, you may spend quite a bit of time trying to retrieve it. If your first thought was, "What is the World Cup?" you may not even search for an answer. Understanding how people control their retrieval has been central to metamemory research on retrieval processes. Just like research on the control of study, much of the research on the control of retrieval has sought to understand how people use their monitoring to control the time. As implied by the earlier example, FOK judgments as well as confidence in retrieved answers may play a role in people's decisions about how long to search for a response ("termination of search") and whether to output a response that had been retrieved.

Concerning termination of search, Singer and Tiede (2008) investigated whether FOK judgments are related to search times for general-knowledge questions that lead to "don't know" responses. That is, when people ultimately cannot retrieve answers to questions, do their feelings of knowing drive them to use more or less time while searching? To answer this question, some participants were used to develop FOK norms for a set of general-knowledge questions. These participants attempted to answer each question and then rated their confidence in recognizing the correct answer (1 = sure I would not recognize it to 9 = sure I would recognize it). An average FOK for each item was computed, and such norms across participants were positively correlated with individual FOK judgments from single individuals. Thus, using them allowed Singer and Tiede (2007) to estimate the FOK judgments of a second group of participants who were instructed to merely retrieve the answers. The prediction was that the normative FOKs would be positively correlated with retrieval latencies (i.e., the time taken to retrieve an answer before giving up) when participants could not retrieve an answer. Consistent with this prediction, the correlation between FOKs and retrieval latencies for unrecalled answers was +.50.

This outcome is intuitive: When we really feel we know an answer, we typically spend more time trying to pull it from memory. Even when we do retrieve an answer to a question, however, it does not mean we will share it with others. In fact, withholding incorrect responses is often just as important as responding correctly, such as when a witness is admonished to "tell the truth, and nothing but the truth." Koriat and Goldsmith (1996) proposed an influential model of memory retrieval that explained why (and when) people would either volunteer a response or withhold it (see also Higham & Arnold, 2007). Importantly, this model illustrates how accurate monitoring is critical for the quality of people's decision to output a response. Their model is presented in Figure 19.3. After being asked a question ("Input Question" in Fig. 19.3), people attempt to retrieve candidate answers from memory. For the question "Which country won the World Cup in 2006?" you may have retrieved United States, Brazil, and Italy. As answers are retrieved, you would evaluate their quality, such as by judging your confidence in the retrieved answer from "0" (meaning there is no way it is correct) to "1.0" (meaning you are sure it is correct). Perhaps you feel that there is only a .10 probability that United States is correct, a .70 chance that Brazil is correct, and a .40 chance that Italy is correct. In the model, these are the assessed probabilities (*P*a). If you did not retrieve any other candidates, you would choose the best one (in this case, Brazil), but even now, you would not necessarily respond with "Brazil." Instead, you would compare your assessed probability (*P*a = .70) with a response criterion (*P*rc). This criterion can be influenced by a variety of factors,

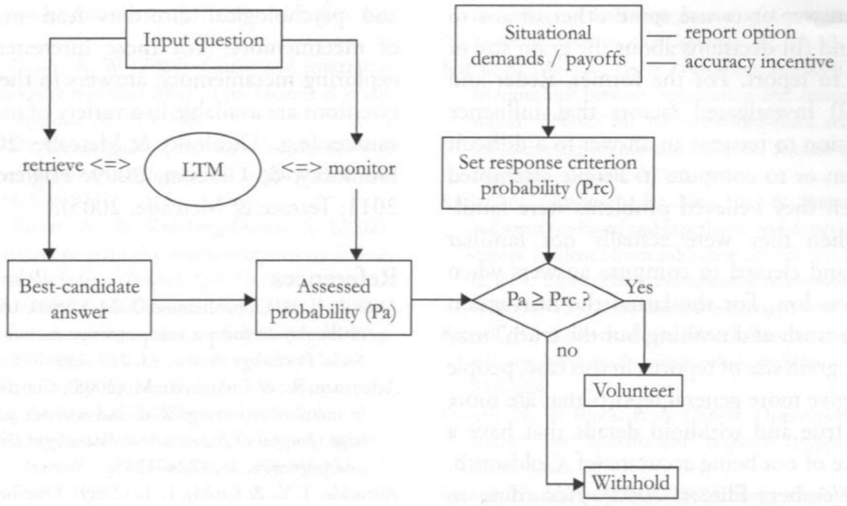

Figure 19.3 Model of self-regulated retrieval that explains whether someone will report an answer that is retrieved when asked a given input question. See text for details. (From Koriat & Goldsmith, 1996.)

such as the current set of payoffs for responses. For instance, you may be taking a test over sports trivia where the reward for a correct response is high and where the penalty for an incorrect response is small. In this case, your response criterion would likely be low. By contrast, if you are on the witness stand and were told not to report any incorrect evidence, you may set a very high response criterion. Let's assume that you set a *P*rc at .40, and hence you would volunteer the response "Brazil," because your assessed probability exceeds your response criterion; that is, *P*a (.70) > *P*rc (.40).

Based on this example, it should be clear why monitoring accuracy is critical for the quality (and quantity) of what responses people will volunteer. Concerning accuracy, if your assessed probabilities do not reflect the actual probabilities of being correct (i.e., poor judgment accuracy), then incorrect answers could be volunteered. This mistake is illustrated in the earlier example: Italy is the correct answer, so the assessed probabilities were not accurate. Thus, your overconfidence in Brazil's team led you to respond incorrectly. If instead, Italy had a much higher *P*a (e.g., close to 1.0), then the enhanced judgment accuracy would have led to the correct response.

To evaluate this implication of the model, Koriat and Goldsmith (1996) used the following method. Participants were forced to answer general-knowledge questions so that the experimenters would know what each participant's best candidate answer was for every question; for each answer, participants provided a confidence judgment, which was used as a measure of *P*a. Next, the participants answered the questions again, but this time, they could withhold answers—that is, they were not forced to respond. As important, two different sets of questions were used. Some were considered standard questions and supported high levels of judgment accuracy and some were deceptive and led to inaccurate judgments. The latter consisted of questions that people often answer incorrectly with high confidence; for instance, "What is the capital of Australia?" is often confidently answered with "Syndey" when the correct answer is "Canberra." During the first round of questions, performance was greater for the standard questions (28%) than for the deceptive ones (12%). This outcome is not surprising, given that it merely demonstrates that the deceptive questions are in fact deceptive. The more important outcomes concern performance for the free-report phase. If participants use their confidence judgments to make decisions about which answers to withhold, their performance should increase for standard questions but may not change for deceptive ones. As expected, the percentage of accurate recall for the free-report phase was 75% for the standard questions, but it was only 21% for the deceptive questions. According to the model, people had difficulties with the deceptive questions because they would have failed to withhold incorrect responses that were inadvertently held with high confidence.

How long one persists in trying to retrieve an item and whether one elects to respond are not the only facets of retrieval that are under people's control. Others include (a) whether to even attempt to retrieve an answer or to use some other means to obtain one and (b) decisions about the grain size of information to report. For the former, Reder and Ritter (1992) investigated factors that influence people's decision to retrieve an answer to a difficult math problem or to compute it. People attempted retrieval when they believed problems were familiar (even when they were actually not familiar with them) and elected to compute answers when familiarity was low. For the latter, the instruction to "speak the truth and nothing but the truth" may influence the grain size of reports; in this case, people will tend to give more general reports that are more likely to be true and withhold details that have a higher chance of not being accurate (cf. Goldsmith, Koriat, & Weinberg-Eliezer, 2002). According to Ackerman and Goldsmith (2008), people control the grain size of their reports (either by providing more precise or more coarse answers) in order to be both informative and confident in their responses. Although we have much to learn about how people control their retrieval and whether such control can be optimal (for discussions of optimality, see Higham, 2007), one conclusion is certain: People do use their monitoring to make decisions about how long to search for answers and whether to volunteer responses. Thus, as with the control of study, accurate monitoring is essential for effective self-regulation.

Conclusion

In the present chapter, we provided an introduction to many of the common measures, methods, and issues that drive research on metamemory. Some major conclusions included that people's monitoring judgments can be accurate, but only when the cues that are used to construct the judgments are diagnostic of criterion performance. As cue diagnosticity increases, then the upper limit on judgment accuracy also increases. Regardless of judgment accuracy, monitoring is also used in the service of controlling memory, whether it be in deciding which items to prioritize for study or when to volunteer (or withhold) a response during retrieval. Our overview, however, did not cover the entire field, which has sought to answer many intriguing questions about metamemory. How does metamemory develop in childhood and does it decline as we age? Do animals have metamemory—do they know when they know and when they do not? How do individual differences in metamemory influence educational outcomes? What are the neurological underpinnings of metamemory, and what kinds of brain insult and psychological disorders lead to disruptions of metamemory? For those interested in further exploring metamemory, answers to these and other questions are available in a variety of more extensive sources (e.g., Dunlosky & Metcalfe, 2009; Hacker, Dunlosky, & Graesser, 2009; Higham & Leboe, 2011; Terrace & Metcalfe, 2005).

References

Alter, A. L., & Oppenheimer, D. M. (2009). Uniting the tribes of fluency to form a metacognitive nation. *Personality and Social Psychology Review, 13,* 219–235.

Ackerman, R., & Goldsmith, M. (2008). Control over grain size in memory reporting: With and without satisficing knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34,* 1224–1245.

Arbuckle, T. Y., & Cuddy, L. L. (1969). Discrimination of item strength at time of presentation. *Journal of Experimental Psychology, 81,* 126–131.

Ariel, R., Dunlosky, J., & Bailey, H. (2009). Agenda-based regulation of study-time allocation: When agendas override item-based monitoring. *Journal of Experimental Psychology: General, 138,* 432–447.

Benjamin, A. S. (2005). Response speeding mediates the contributions of cue familiarity and target retrievability to metamnemonic judgments. *Psychonomic Bulletin and Review, 12,* 874–879.

Benjamin, A. S., & Bird, R. D. (2006). Metacognitive control of the spacing of study repetitions. *Journal of Memory and Language, 55,* 126–137.

Benjamin, A. S., Bjork, R. A., & Schwartz, B. L. (1998). The mismeasure of memory: When retrieval fluency is misleading as a metamnemonic index. *Journal of Experimental Psychology: General, 127,* 55–68.

Brunswik, E. (1956). *Perception and representative design of psychological experiments.* Berkeley: University of California Press.

Carroll, M., Nelson, T. O., & Kirwan, A. (1997). Tradeoff of semantic relatedness and degree of overlearning: Differential effects on metamemory and on long-term retention. *Acta Psychologica, 95,* 239–253.

Castel, A. D. (2008). Metacognition and learning about primacy and recency effects in free recall: The utilization of intrinsic and extrinsic cues when making judgments of learning. *Memory and Cognition, 36,* 429–437.

Dunlosky, J., Ariel, R., & Thiede, K. W. (2011). Agenda-based regulation of study-time allocation. In P. Higham & J. P. Leboe (Eds.), *Constructions of remembering and metacognition* (pp. 182–200). Hampshire, England: Palgrave Macmillan.

Dunlosky, J., & Bjork, R. A. (Eds.). (2008). *A handbook of metamemory and memory.* New York: Psychology Press.

Dunlosky, J., & Matvey, G. (2001). Empirical analysis of the intrinsic-extrinsic distinction of judgments of learning (JOLs): Effects of relatedness and serial position on JOLs. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27,* 1180–1191.

Dunlosky, J., & Metcalfe, J. (2009). *Metacognition.* Los Angeles: Sage.

Dunlosky, J., Serra, M. J., & Baker, J. M. C. (2007). Metamemory. In F. T. Durso, Nickerson, R. S., Dumais, S. T., Lewandowsky, S., & Perfect, T. J. (Eds.), *Handbook* *of applied cognition* (2nd ed., pp. 137–160). Hoboken, NJ: Wiley.

Dunlosky, J., & Thiede, K. W. (2004). Causes and constraints of the shift-to-easier-materials effect in the control of study. *Memory and Cognition, 32*, 779–788.

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American Psychologist, 34*, 906–911.

Goldsmith, M., Koriat, A., & Weinberg-Eliezer, A. (2002). Strategic regulation of grain size memory reporting. *Journal of Experimental Psychology: General, 131*, 73–95.

Hacker, D. J., Dunlosky, J., & Graesser, A. C. (Eds.). (2009). *Handbook of metacognition and self-regulated learning*. Mahwah, NJ: Erlbaum.

Hart, J. T. (1965). Memory and the feeling-of-knowing experience. *Journal of Educational Psychology, 56*, 208–216.

Hart, J. T. (1966). Methodological note on feeling-of-knowing experiments. *Journal of Educational Psychology, 57*, 347–349.

Hertzog, C., Dunlosky, J., Robinson, A. E., & Kidder, D. P. (2003). Encoding fluency is a cue used for judgments about learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*, 22–34.

Hertzog, C., Price, J., & Dunlosky, J. (2008). How is knowledge generated about memory encoding strategy effectiveness? *Learning and Individual Differences, 18*, 430–445.

Higham, P. A. & Leboe, J. P. (Ed.). (2011). *Constructions of remembering and metacognition*. Hampshire, England: Palgrave Macmillan.

Higham, P. A. (2007). No Special K! A signal detection framework for the strategic regulation of memory accuracy. *Journal of Experimental Psychology: General, 136*, 1–22.

Higham, P. A., & Arnold, M. M. (2007). How many questions should I answer? Using bias profiles to estimate optimal bias and maximum score on formula-scored tests. *European Journal of Cognitive Psychology, 19*, 718–742.

Hines, J. C., Touron, D. R., & Hertzog, C. (2009). Metacognitive influences on study time allocation in an associative recognition task: An analysis of adult age differences. *Psychology and Aging, 24*, 462–475.

Humphrey, G. (1951). *Thinking: an introduction to its experimental psychology*. New York: Wiley.

Karpicke, J. D. (2009). Metacognitive control and strategy selection: Deciding to practice retrieval during learning. *Journal of Experimental Psychology: General, 138*, 469–486.

Kolers, P. A., & Palef, S. R. (1976). Knowing not. *Memory and Cognition, 4*, 553–558.

Koriat, A. (1993). How do we know that we know? The accessibility model of the feeling of knowing. *Psychological Review, 100*, 609–639.

Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General, 126*, 349–370.

Koriat, A., Bjork, R. A., Sheffer, L., & Bar, S. K. (2004). Predicting one's own forgetting: The role of experience-based and theory-based processes. *Journal of Experimental Psychology: General, 133*, 643–656

Koriat, A., & Goldsmith, M. (1996). Monitoring and control processes in the strategic regulation of memory accuracy. *Psychological Review, 103*, 490–517.

Koriat, A., & Levy-Sadot, R. (2001). The combined contributions of the cue-familiarity and accessibility heuristics to feelings of knowing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*, 34–53.

Koriat, A., Ma'ayan, H., & Nussinson, R. (2006). The intricate relationships between monitoring and control in metacognition: Lessons for the cause-and-effect relation between subjective experience and behavior. *Journal of Experimental Psychology: General, 135*, 36–69.

Koriat, A., Nussinson, R., Bless, H., & Shaked, N. (2008). Information-based and experience-based metacognitive judgments: Evidence from subjective confidence. In J. Dunlosky & R. A. Bjork (Eds.), *Handbook of metamemory and memory* (pp. 117–135). New York: Psychology Press.

Kornell, N., & Bjork, R. A. (2007). The promise and perils of self regulated study. *Psychonomic Bulletin and Review, 14*, 219–224.

Kornell, N., & Bjork, R. A. (2008). Optimizing self-regulated study: The benefits-and costs-of dropping flashcards. *Memory, 16*, 125–136.

Kornell, N., & Metcalfe, J. (2006). Study efficacy and the region of proximal learning framework. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*, 609–622.

Loftus, E., & Ketchum, K. (1991). *Witness for the defense: The accused, the eyewitness, and the expert who puts memory on trial*. New York: St. Martin's Press.

Matvey, G., Dunlosky, J., & Guttentag, R. (2001). Fluency of retrieval at study affects judgments of learning (JOLs): An anlaytic or nonanalytic basis for JOLs? *Memory and Cognition, 29*, 222–232.

Mazzoni, G., Cornoldi, C., & Marchitelli, G. (1990). Do memorability ratings affect study-time allocation? *Memory and Cognition, 18*, 196–204.

Metcalfe, J. (2002). Is study time allocated selectively to a region of proximal learning? *Journal of Experimental Psychology: General, 131*, 349–363.

Metcalfe, J. (2009). Metacognitive judgments and control of study. *Current Directions in Psychological Science, 18*, 159–163.

Metcalfe, J., & Finn, B. (2008). Evidence that judgments of learning are causally related to study choice. *Psychonomic Bulletin and Review, 15*, 174–179.

Metcalfe, J., Schwartz, B. L., & Joaquim, S. G. (1993). The cue-familiarity heuristic in metacognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*, 851–861.

Miner, A. C., & Reder, L. M. (1994). A new look at feeling of knowing: Its metacognitive role in regulating question answering. In J. Metcalfe & A. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 47–70). Cambridge, MA: MIT Press.

Nelson, T. O., & Dunlosky, J. (1991). When people's judgments of learning (JOLs) are extremely accurate at predicting subsequent recall: The "delayed-JOL effect." *Psychological Science, 2*, 267–270.

Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 26, pp. 125–141). New York: Academic Press.

Nelson, T. O., Gerler, D., & Narens, L. (1984). Accuracy of feeling of knowing judgments for predicting perceptual identification and relearning. *Journal of Experimental Psychology: General, 113*, 282–300.

Nelson, T. O., & Leonesio, R. J. (1988). Allocation of self-paced study time and the "labor-in-vain effect." *Journal of*

Experimental Psychology: Learning, Memory, and Cognition, 14, 676–686.

Pyc, M. A. & Dunlosky, J. (2010). Toward an understanding of students' allocation of study time: Why do they decide to mass or space their practice? *Memory and Cognition, 38, 431–440.*

Rabinowitz, J. C., Ackerman, B. P., Craik, F. I. M., & Hinchley, J. L. (1982). Aging and metamemory: The roles of relatedness and imagery. *Journal of Gerontology, 37*, 688–695.

Reder, L. M., & Ritter, F. E. (1992). What determines initial feeling of knowing? Familiarity with question terms, not with the answer. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*, 435–451.

Rhodes, M. G., & Castel, A. D. (2009). Metacognitive illusions for auditory information: Effects on monitoring and control. *Psychonomic Bulletin and Review, 16*, 550–554.

Roediger, H. L., III, & Karpicke, J. D. (2006). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science, 1*, 181–210.

Serra, M. J., & Dunlosky, J. (2005). Does retrieval fluency contribute to the underconfidence-with-practice effect? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*, 1258–1266.

Shaughnessy, J. J. (1981). Memory monitoring accuracy and modification of rehearsal strategies. *Journal of Verbal Learning and Verbal Behavior, 20*, 216–230.

Singer, M., & Tiede, H. L. (2008). Feeling of knowing and duration of unsuccessful memory search. *Memory and Cognition, 36*, 588–597.

Son, L. K. (2004). Spacing one's study: Evidence for a metacognitive control strategy. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*, 601–604.

Son, L. K., & Kornell, N. (2008). Research on the allocation of study time: Key studies from 1890 to the present (and beyond). In J. Dunlosky & R. A. Bjork (Eds.), *Handbook of metamemory and memory* (pp. 333–351). New York: Psychology Press.

Son, L., & Metcalfe, J. (2005). Judgments of learning: Evidence for a two-stage process. *Memory and Cognition, 33*, 1116–1129.

Son, L. K., & Metcalfe, J. (2000). Metacognitive and control strategies in study-time allocation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 204–221.

Souchay, C., Moulin, C. J. A., Clarys, D., Taconnat, L., & Isingrini, M. (2007). Diminished episodic memory awareness in older adults: Evidence from feeling-of-knowing and recollection. *Consciousness and Cognition, 16*, 769–784.

Terrace, H. S., & Metcalfe, J. (Eds.). (2005). *The missing link in cognition: Origins of self-reflective consciousness.* New York: Oxford University Press.

Thiede, K. W., & Dunlosky, J. (1999). Toward a general model of self-regulated study: An analysis of selection of items for study and self-paced study time. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*, 1024–1037.

Toppino, T. C., Cohen, M. S., Davis, M. L., & Moors, A. C. (2009). Metacognitive control over the distribution of practice: When is spacing preferred? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 5*, 1352–1358.

Winne, P. H., & Hadwin, A. F. (1998). Studying as self-regulated learning. In D. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 277–304). Hillsdale, NJ: Erlbaum.

Woodworth, R. S. (1921). *Psychology: A study of mental life.* New York: Henry Holt.