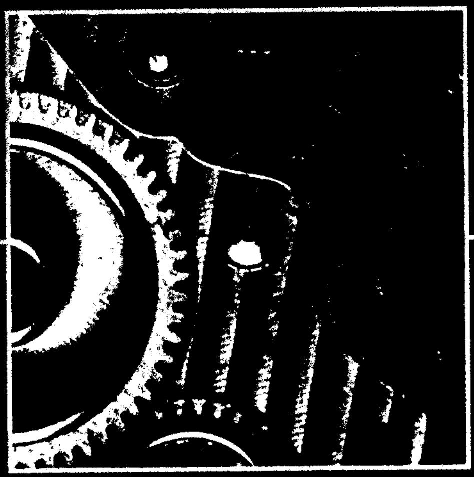
HANDBOOK OF

APPLIED COGNITION

Edited by
FRANCIS T. DURSO

Associate Editors
Raymond S. Nickerson, Susan T. Dumais,
Stephan Lewandowsky and Timothy J. Perfect



SECOND EDITION

Metamemory

John Dunlosky, Michael J. Serra, and Julie M. C. Baker
Kent State University, USA

METAMEMORY APPLIED

Imagine Beth, a sixth-year graduate student, as she prepares to give an important conference talk about her dissertation research. Her PowerPoint slides are all in order, so she proceeds by practicing her talk as she steps from slide to slide. As Beth practices, she realizes that her introduction – where she provides a broader perspective on her field – is quite weak; she notices herself stumbling and searching for words. Because of these problems, she decides to develop more detailed ideas on what to say, so that the beginning of the talk will be polished. Also, Beth is adamant about not using notes or reading from the slides, and she realizes that it is best to straight-out memorize the short list of caveats about her conclusions. To do so, she repeats each caveat over and over. Beth has already given multiple talks, and it is obvious to her that her anxiety can undermine what should otherwise be a smooth presentation. So to help conquer some anxiety, she decides to practice the talk a bit extra this time, and to do so multiple times with a mock audience.

DEFINING METACOGNITION: CONSTRUCTS AND QUESTIONS

Although many of the activities that Beth engaged in – from preparing her slides to practicing her talk – clearly involve cognitive processes, some of the activities recruit metacognitive processes as well. Taken literally, the term *metacognition* means cognition about cognition. Any component of cognition can bear the *meta*-prefix, such as metalinguistics, metaperception, and metamemory. *Metamemory* refers to a person's cognitions about his or her own memory and is the focus of the present chapter. After we define the core constructs and questions that are fundamental to metacognition, we move to a detailed analysis of metamemory and how advances in this area have led to insights into improving human learning and memory. The same definitions, methods, and theory described in these sections can be applied to exploring other real-world problems that involve people's metacognitive abilities. At the end of the chapter, we touch on two of

Handbook of Applied Cognition: Second Edition. Edited by Francis T. Durso.

these problems – improving juror deliberation and people's comprehension of text – to illustrate the breadth and power of the metacognitive framework. Thus, we do not provide a laundry list of all current and possible applications of metamemory. Instead, this chapter is a guide for researchers who may want to explore how metacognitive abilities may contribute to human performance in their applied domain of interest. To best serve this purpose, we first discuss definitions and theory of metacognition relevant to application – which includes real-world examples to illustrate key points – before we consider applied metamemory in detail.

Concerning a more refined conceptualization of metacognition, ever since Flavell (1979) emphasized the importance of metacognition to cognitive development, *metacognition* has been defined by the numerous kinds of activity that are either about one's cognitive processes or work upon them. Although early debates in the area centered on which activities should be included under the rubric of metacognition (Cavanaugh & Perlmutter 1982), current conceptualizations include these three constructs: knowledge about one's cognition, the monitoring of ongoing cognition, and controlling cognition. We consider each construct in turn, as well as how each one arises as Beth prepares for her conference talk. Beth's preparation for her talk, though, is just one example of how metacognition can influence behavior in everyday situations. The three constructs described above – knowledge, monitoring, and control – can be isolated and studied in many situations in which people perform everyday tasks that involve thinking about and controlling their actions.

Metacognitive knowledge, or knowledge about cognition, is declarative knowledge composed of a person's beliefs about any aspect of cognition. In the introductory scenario, Beth believed that anxiety would undermine her performance (which is likely correct; see Matthews *et al.* 2000, chapter 15), and she obviously knew about some rehearsal strategies. Nevertheless, her belief that rote repetition would ensure excellent memory for her conclusions may not be entirely accurate (e.g., Hertzog *et al.* 1998). Initial research on metacognition concerned people's beliefs and knowledge about cognition (and in particular, children's knowledge, e.g., Kreutzer *et al.* 1975), and a great deal of research has been devoted to understanding the development of people's beliefs about cognition and how these beliefs are related to one's wellbeing and cognitive performance.

Almost everyone will accrue extensive knowledge and beliefs about how cognition operates in general. Beginning in early development, children develop a sophisticated theory of mind, which includes an understanding of mental states such as one's desires, intentions, and emotions (for a recent review, see Flavell 2004). Later in life, many people develop a belief that forgetting becomes more of a problem as we age (McDonald-Miszczak *et al.* 1995), and adults of all ages have knowledge and beliefs about how the mind operates in general (Hertzog & Hultsch 2000). The knowledge obtained is not always accurate, and recent research in the domain of psychopathology has explored the degree to which incorrect beliefs about cognition can contribute to mental disorder (e.g., Cartwright-Hatton & Wells 1997). For instance, Emmelkamp and Aardema (1999) report that obsessive-compulsive behavior is related to thought–action fusion, which is the belief that one's intrusive thoughts are "morally equivalent to carrying out a prohibited action" (p. 139, e.g., having a violent thought is almost as inappropriate as violence itself). Likewise, a doctor may hold the belief that patients understand and will remember the instructions they are given, which may often reflect inaccurate knowledge about others' minds that could undermine patients' wellbeing. In these cases, researchers will need to discover the specific knowledge and beliefs people have about cognition and how such metacognitive knowledge influences thought and action.

In the present chapter, we focus on the more process-oriented constructs of metacognition: monitoring and control.[1] *Monitoring* involves assessing the current state or ongoing progress of any aspect of cognition, and *control* involves the regulation of ongoing cognitive processes. Examples of each abound. Beth monitored her ongoing progress toward developing a competent talk by attending to how often she stumbled during practice. She then exerted control by deciding to continue practicing those parts of the talk on which she stumbled the most. In this case, Beth's monitoring (i.e., checking for stumbling) served in the control of how she practiced (e.g., by devising detailed ideas about how to present parts of the talk where she stumbled).

These definitions and examples provide an excellent beginning (and end for many) concerning monitoring and control. Nevertheless, even though the nature of these two constructs in some ways is quite intuitive, their actual conceptualization in the literature is sometimes vague or overextended to include other processes. For instance, consider the following rationale, which is arguably plausible. If monitoring involves assessing an ongoing cognitive state and self-assessments involve conscious activity, it seems reasonable that an activity must be conscious to count as metacognitive. Whether conscious awareness is necessary for metacognitive activities has stimulated research and debates (Reder 1996), but the definitions of monitoring and control should remain agnostic toward the issue of consciousness. This goal was met by Nelson and Narens (1990, 1994), who defined monitoring and control in terms of the flow of information within a dominance relationship between a metalevel and object-level of cognition. As shown in Figure 6.1 (Panel A), Nelson and Narens (1990) defined monitoring and control in terms of the flow of information between these two levels. In particular, monitoring occurs when the metalevel is informed by information from an object-level cognition, and control occurs whenever information from the metalevel modifies the object-level. Nelson and Narens (1994)

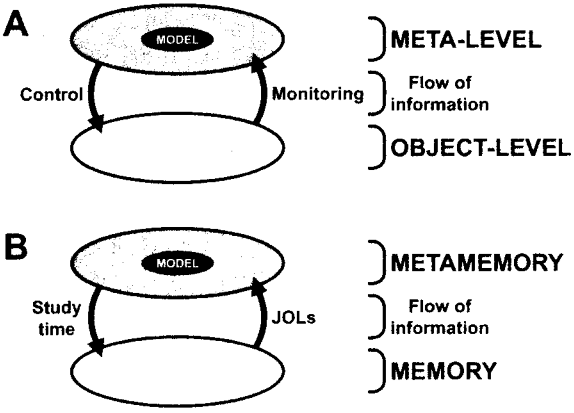


Figure 6.1 Monitoring and control defined as the flow of information between two levels of cognition (Panel A) and a specific instantiation relevant to the allocation of study time (Panel B). Adapted from Nelson and Narens (1994)

described this system in detail and generalized it to more than two levels. Importantly, this model is agnostic as to whether people are conscious of the flow of information or as to whether the two levels of representation are psychologically or physiologically distinct. Both issues will need to be resolved through empirical analyses (for evidence that they are distinct, see Mazzoni & Nelson 1998; Pannu & Kaszniak 2005).

Important advances in our understanding of monitoring and control processes have been achieved over the past three decades. Advances in measurement, methods, and theories have led to a honing of questions relevant to metacognition and to a greater specificity of answers for them. As important, these advances have implications for enhancing human performance in many domains, and some have been validated empirically in the laboratory and in the field.

In the remainder of this chapter, we will discuss monitoring and control mainly as they have been addressed in regards to *metamemory*, which is the facet of metacognition that concerns people's monitoring and control of learning and retrieval processes. We targeted metamemory because progressive – and extensive – research is available concerning both monitoring and control within this domain. We move next to a discussion of standard measures of memory monitoring and control and introduce core questions within this field. This discussion largely comprises a tutorial that may be skipped by more knowledgeable readers. We then review recent research that has sought to answer some of these core questions in one area of metamemory that pertains to how people monitor and control their learning. This review is also meant to showcase how metacognitive methods and theory have been used to investigate metamemory. As important, we describe this research in some detail in hopes that the method and theory can be extended to investigate any real-world task that involves metacognitive processes. After this review, we end with a discussion of how such a metamemory framework has been extended to investigate two applied domains (jury deliberation and text comprehension), so as to showcase the versatility and power of the framework.

MONITORING AND CONTROL: MEASURES AND QUESTIONS

In the domain of metamemory, any memory process can be the object of the metalevel, so that monitoring can potentially be directed toward any one of these processes. A student can monitor his or her ongoing learning of simple materials and can later monitor progress toward retrieving these same materials. To measure monitoring in either case, research participants are often asked to make overt judgments about the particular process under investigation. For instance, to investigate people's monitoring of ongoing learning, participants may judge how well they have learned a particular item by predicting the likelihood of recalling it on a subsequent test. It is important to note that these judgments – or any other kind of metamemory judgment for that matter – are not taken at face value, as if they were a direct and valid measure of on-going monitoring of memory (Nelson 1996a). Instead, researchers obtain objective measures of performance that can be compared to the judgments, so that judgment biases can be described and understood. For instance, to evaluate Beth's metacognitive ability, we might ask her to explicitly judge how well she performed each portion of her talk and then compare those judgments to some objective assessment of her performance, such as ratings obtained from a group of experienced speakers.

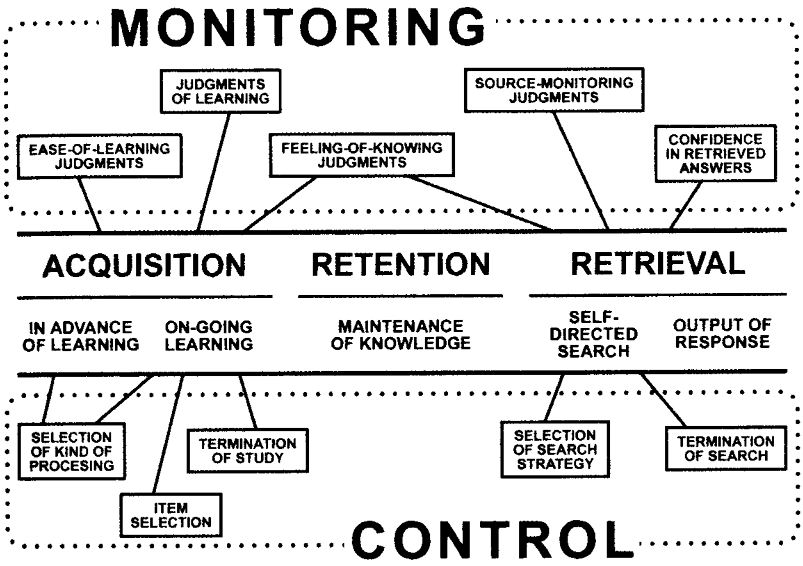


Figure 6.2 Overview of the metacognitive framework first proposed by Nelson and Narens (1990), which now includes judgments that were absent from the original framework

Through the 1970s and 1980s, interest in metamemory grew steadily and investigators explored a variety of monitoring and control processes relevant to learning. Although exceptions exist, research on any one monitoring or control process was largely conducted in isolation from others. Isolationism was curtailed, however, when Nelson and Narens (1990) published a framework that subsequently unified areas of metamemory research. In particular, they organized the different measures of monitoring and control across three stages of learning: acquisition, retention, and retrieval. The measures are illustrated in Figure 6.2 (adapted from Nelson & Narens 1990), and their definitions are presented in Table 6.1.

To make any judgment, a person is asked to evaluate some aspect of learning or test processes. For instance, participants may study paired associates (e.g., dog–spoon), take a test of recall on those items (e.g., recall the correct response to "dog– ?"), and then take a recognition test for all items in which they did not recall the correct response. During these study-test trials, participants may judge several aspects of their progress and performance, and the various metamemory judgments differ with respect to which specific process is being evaluated. For instance, participants may make ease-of-learning judgments prior to learning by judging how easy each item will be to learn during subsequent study trials. For judgments of learning, participants would first study the items and then predict how likely they will be to recall each one on the upcoming criterion test. For feeling-of-knowing judgments, participants first take a recall test over all the items and then judge responses that were not correctly recalled. For these non-recalled items, participants predict the likelihood of recognizing the non-recalled response on a recognition test with multiple alternatives. And for retrospective confidence judgments, participants judge the likelihood that their responses on a test are correct. Each judgment may be

Table 6.1 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 (JOL)	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 to 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.

applied to investigate metamemory in real-world tasks. For instance, when students are preparing for an exam, a teacher may ask them to judge how well they have learned the study materials to identify students who are overconfident about how much they have learned. And during the actual exam, students may be asked to judge the correctness of each response in an attempt to determine how often they guess.

In Figure 6.2, the judgments also differ with respect to which control processes that they may inform. For instance, judgments of learning (JOLs) presumably inform the termination of study during ongoing learning, and confidence in retrieved answers informs the control of search during self-regulated retrieval. Such monitoring-and-control dynamics are addressed more specifically in Figure 6.1 (Panel B), which illustrates the relationship between constructs and measures for the termination of study. Termination of study time is hypothesized to rely on an evaluation (as measured by JOLs) of ongoing learning, which then feeds into a decision about terminating study. For instance, students who are overconfident about how much they have learned may stop studying well before achieving their learning goals. Of course, even the model in Figure 6.1 (Panel B) is not highly refined, because although it predicts that JOLs and study time will be related, it does not specify how (or to what degree) these measures should be related. More refined models of the control functions listed in Figure 6.2 have been proposed and evaluated – some of which will be introduced later in this chapter.

Individual literatures have arisen around almost every one of the monitoring and control processes depicted in Figure 6.2. But even with this diversity and breadth of investigations, the questions that have consistently motivated research have been relatively few in number. Four of these questions are listed in Table 6.2; though not exhaustive, they do represent major foci across the areas. The questions pertain to the constructs of monitoring and

Table 6.2 Core questions about monitoring and control from research about adult human cognition

How do people monitor memory?
How accurate is memory monitoring?
Can monitoring accuracy be improved?
How is monitoring used to control?

Table 6.3 Index to references of research relevant to the dominant questions (Table 6.2) and core metacognitive judgments (Figure 6.2)

Questions	Metacognitive judgments				
	EOL	JOLs	FOK	RC	SM
How do individuals monitor memory?	1	2	3	4	5
How accurate is memory monitoring?	6	7	8	9	10
Can monitoring accuracy be improved?	11	12	13	14	15
How is monitoring used to control?	16	17	18	19	20

Note: EOL, ease of learning judgment; JOL, judgment of learning; FOK, feeling of knowing judgment; RC, retrospective confidence judgment; SM, source monitoring judgment. The number in each cell above corresponds to one or more articles that can be used as an entry point to the literature. The numeral within a cell is used to label the corresponding article in the Appendix.

control, and hence within each literature, they have been pursued by investigating the corresponding measures presented in Figure 6.2. Researchers interested in how people monitor memory investigate how they make the various judgments of memory monitoring, such as JOLs and FOK judgments. Researchers interested in the accuracy (or validity) of people's memory monitoring often compare a person's judgments of monitoring (e.g., JOLs) with the objective target of the particular judgment (e.g., eventual test performance). Similarly, those interested in how monitoring influences control processes often examine the empirical relations between a judgment of monitoring (e.g., JOLs) and a corresponding measure of control (e.g., time devoted to study).

Obviously, to review answers to each question in Table 6.2 for all the literatures, we would need much more room than is provided by a single chapter. Thus, we have limited our review to questions most relevant to JOLs and self-regulated study because this particular area of metamemory is relevant to improving human performance (Bjork 1994). In general, the conclusions offered by research on JOLs and self-regulated study will generalize to other monitoring control functions, but we realize some readers' most avid interests will align with these other functions. For instance, you may be concerned with why some grade school students are better at taking exams than are other students who are presumably as knowledgeable. If your intuition is that the poor test-takers are less skilled at evaluating the quality of their answers, then you may want to begin exploring the literature on retrospective confidence judgments that pertains to the question, "How accurate is memory monitoring?" With this possibility in mind, we have constructed Table 6.3, which can be used to locate a key reference from literatures about each question and function. To use this table, locate the numeral within the cell of interest, and then refer to the Appendix to identify the relevant article that can be used as an entry point to the larger literature pertaining to your question and function of interest. In the example above, the relevant entry of Table 6.3 is "9," which would lead you to Gigerenzer *et al.* (1991).

MONITORING DURING ACQUISITION: JUDGMENTS OF LEARNING

In this section, we consider how people judge their learning of newly studied material, the accuracy of these judgments, and how accuracy can be improved. A motivating principle that has stimulated interest in these issues is quite intuitive: obtaining high levels of judgment accuracy can support efficient learning. In the introductory scenario, Beth judged that the beginning of her talk was not quite coherent. If her judgment was accurate, she could efficiently use her time by practicing this portion of her talk instead of others that she already performed well. If instead she had inaccurately judged that her introduction was incoherent when in fact it was fine, she may inadvertently waste time in practicing this portion of the talk. Beyond this anecdote, this principle also appears to hold for judgments of learning and the effectiveness of study. Namely, people who more accurately judge what they have versus have not learned demonstrate faster learning across multiple trials because they are able to isolate and restudy those items that require it most (for a review, see Dunlosky *et al.* 2005).

How People Judge Memories for Newly Learned Materials

Discovering the sources of judgment (in)accuracy can be informed by understanding how people judge memories for newly learned materials. Put differently, explanations of how people make metamemory judgments often have direct implications about why the judgments are (in)accurate. We explore one of the more thoroughly investigated explanations here and pursue their implications for accuracy in the following section.

To investigate how people make judgments of learning, researchers have adopted a relatively standard method. Namely, participants are instructed to study a list of items, such as paired associates (dog–spoon), which are presented individually for study. Sometime after studying each item, a JOL is made in which a participant rates the likelihood of correctly remembering each item. This rating typically is made on a scale from 0 to 100, where 0 means no chance that the item will be remembered and 100 means that it will absolutely be remembered. After items are studied, a test of memory is administered for all items. In some cases, participants are then asked to select which items that they would like to restudy, and/or asked to pace their restudy of each item. These latter control processes – item selection and self-paced study – are considered in a later section, and we focus here on the issues of how people make JOLs and on JOL accuracy. Hypothetical data relevant to these issues are presented in Table 6.4, which we will return to throughout this discussion.

One relatively successful hypothesis has been proposed that attempts to explain how people make almost every one of the judgments presented in Figure 6.2 (for competing hypotheses that have not faired as well, see Hart 1965; Koriat 1993; Metcalfe 2000). This

Table 6.4 Hypothetical data from a single participant

Item	JOL	Recall	Item selection	Study time (s)
King–Crown	80	1	0	1.2
Paint–Artist	90	0	0	1.0
Leaf–Branch	60	0	1	3.4
Weed–Garden	75	1	1	4.0
Suit–Jacket	90	1	0	0
Paper–Pencil	70	1	0	1.3
Dog–Spoon	35	0	1	4.0
Cotton–Lamp	40	1	1	4.7
Lunch–Motor	55	0	1	3.9
Baby–Forest	50	1	0	1.9
Milk–Clerk	35	0	1	5.2
Snake–Brain	45	0	1	6.0

Dependent measure	Across all items	Related items	Unrelated items
Mean judgment	60.4	77.5	43.3
Relative accuracy	+0.43	+0.14	+0.25
JOL/selection correlation	−0.77	−0.75	−0.60
JOL/study time correlation	−0.68	−0.71	−0.29

Note: Hypothetical data from a single participant along with common descriptive analyses. JOL, judgment of learning. Recall: 1, correct; 0, incorrect. Item selection: 1, selected for restudy; 0, not selected. Gamma correlations were used to derive the measures of relative accuracy and the correlational measures of the allocation of study time (for rationale, see Nelson 1984).

inference-based hypothesis is that people's metamemory judgments are inferential in nature and are based on any number of available cues that can be used to infer the memorability of an item (for a review, see Schwartz *et al.* 1997). For example, consider a list of paired-associate items composed of both related pairs (king–crown) and unrelated pairs (dog–spoon). A participant studies this list and makes JOLs as described earlier. This participant may attend to the cue of relatedness and then infer its meaning with respect to memory. In particular, she may infer that "being related" means that memory will be quite good for that item. In our opening example, Beth presumably inferred how well she was performing her talk by attending to how often she stumbled. In this case, stumbling (the cue) may be fairly diagnostic of the quality of her talk, although she may also use other cues (e.g., dryness of her mouth) that may not be that diagnostic.

Given that inferential accounts have received substantial support (e.g., Koriat 1993, 1997; Schwartz *et al.* 1997; Benjamin *et al.* 1998), one research goal involves identifying cues that people use to make inferences about memory. To discover these cues, investigators have adapted the standard method either by inserting a possible cue within the task (e.g., item relatedness) or measuring cues as they occur naturally. Several outcomes are noteworthy. JOLs are substantially related to the following cues: (1) the semantic association of words within a paired associate, with JOLs being greater for related word pairs (e.g., king–crown) than for unrelated word pairs (e.g., dog–spoon; Koriat & Bjork 2005); (2) the fluency of processing items during study, with JOLs being greater for items in which mediators (e.g., interactive images) are more quickly generated than are more slowly generated (Hertzog *et al.* 2003); (3) whether the expected criterion test is recognition or recall, with JOLs being greater in anticipation of recognition tests regardless of the relative difficulty of the two kinds of test (Thiede 1996); and (4) the fluency of retrieval, with JOLs being greater for items in which responses are retrieved quickly than for those retrieved less quickly (Benjamin *et al.* 1998). By contrast, various encoding strategies that influence performance (e.g., imagery vs. rote repetition) often have a relatively small influence on JOLs (e.g., Shaughnessy 1981; Dunlosky & Nelson 1994). Although a variety of accounts have been offered for these effects (most notably, see Koriat 1997), a general process-oriented explanation is not yet available, which offers one of the most important challenges for theory development in this area. Moreover, much translational research is needed to discover which cues influence people's judgments of their learning in real-world settings, such as when a clinical graduate studies for a licensing exam or a young physician is learning about new procedures for diagnoses.

In summary, much is now known about how people judge their learning of individual items for an upcoming test. JOLs are largely inferential in nature, being constructed from available cues that learners perceive as relevant to memory. As we discuss next, an inferential account of judgments has much heuristic value both for understanding and improving judgment accuracy.

Accuracy of Judgments of Learning

For any metamemory judgment, accuracy is operationalized as the relationship between the judgment and performance on the corresponding criterion test. In the judgmental literature, accuracy has been conceptualized in two ways (Murphy 1973; Lichtenstein *et al.* 1977; Nelson 1996b). First, *relative accuracy* refers to the degree to which judgments discriminate between the test performance of one item relative to another. In our opening example, Beth may have judged that the opening of her talk was great and that her conclusions need work; if that were truly the case, then her judgments demonstrated excellent relative accuracy at discriminating between what she performed well vs. what she had performed less well. Relative accuracy (also known as *resolution*) has been measured in multiple ways in the literature, but the most widely used measures involve computing an intra-individual correlation between each participant's judgments and his or her own test performance. Correlations that are reliably greater than 0 indicate above-chance relative accuracy, and higher correlations (approaching +1.0) indicate higher levels of relative accuracy. Several correlations have been used to estimate relative accuracy, including Pearson r and the Goodman-Kruskal γ correlation (for other measures, see Nelson 1984; Gonzalez & Nelson 1996).

Which correlation provides the best estimate of relative accuracy? Nelson (1984) described six properties that would be desirable for any measure of relative accuracy, including (a) that if two people have equal levels of overall test performance, then the person with better judgment ability should be higher on the measure of relative accuracy; and (b) that the measure of relative accuracy should be independent of overall test performance. Based on his evaluation of prevalent measures, Nelson (1984) argued that these properties were best met by the γ correlation, which is now the most widely used correlation for estimating relative accuracy. Gamma can be computed using common statistical packages (e.g., SPSS and Systat). For the hypothetical data in Table 6.4, relative accuracy of JOLs across all items as computed by γ is +0.43.

Second, *absolute accuracy* refers to the degree to which the absolute values of ratings reflect the actual level of performance obtained. For instance, a person that predicts he will recall 70 per cent of the items correct on a test and then recalls 70 per cent has excellent absolute accuracy, whereas another person who predicts 70 per cent but only recalls 40 per cent shows overconfidence. To estimate absolute accuracy, researchers must have participants make judgments on a scale – e.g., judging the *per cent* likelihood of correct recall – that is comparable to the measure of test performance – e.g., *per cent* correct recall. Absolute accuracy is measured by computing calibration indices between mean ratings and mean performance (Keren 1991). One common technique involves constructing a calibration curve, where per cent correct on the criterion test is plotted (on the ordinate) as a function of increasing levels of judgment rating (on the abscissa). Values from the calibration curve can be used to estimate the degree to which judgments show under-confidence or overconfidence by computing a weighted mean of the differences between the mean judgment and the corresponding level of text performance for each level of judgment (Lichtenstein & Fischhoff 1977). Positive values indicate overconfidence and negative values indicate under-confidence. (For detailed discussion of how to measure absolute accuracy and some difficulties in interpreting these measures, see Keren 1991; Wallsten 1996.)

In the present review, we highlight results pertaining to relative accuracy, which has been consistently reported in the JOL literature. The earliest investigations typically had participants make *immediate* JOLs in which a JOL was made for an item immediately after it had been studied (e.g., Arbuckle & Cuddy 1969; Zechmeister & Shaughnessy 1980). For instance, a nurse may describe a schedule for taking a new medication to a patient and then immediately ask, "Will you remember this schedule?" In this case, the patient would be making an immediate judgment of his or her learning. Although some exceptions exist, the relative accuracy of immediate JOLs tends to be quite low, but above zero (for a review, Sikström & Jönsson 2005) – which is reflected in mean correlations that are reliably greater than zero but below +0.50. Why do these judgments show above-chance accuracy and how might their accuracy be improved? According to inferential-based accounts, judgments will accurately predict performance as long as the cues that are used to make judgments are empirically correlated with test performance. In the case of immediate JOLs, the available cues apparently are somewhat (but not highly) predictive of subsequent performance, so low levels of accuracy should be expected.

Inferential-based accounts provide insight into improving people's judgment accuracy. Namely, to improve accuracy, techniques are needed that produce cues that are highly diagnostic of test performance. One possibility would be to insert diagnostic cues into the learning environment. For instance, mixing related pairs (king–crown) with unrelated word pairs (dog–spoon) on a list improves accuracy because this cue (relatedness) influences JOLs and is also highly diagnostic of test performance. As illustrated in Table 6.4, relative accuracy computed across related and unrelated pairs yields higher levels of accuracy than is found for either subset of item (as in Dunlosky & Matvey 2001). Unfortunately, this solution for improving accuracy is not highly practical, because learners rarely construct the set of materials that they need to learn. Instead, for practical applications, techniques are needed that enhance judgment accuracy regardless of the set of to-be-learned materials.

In the late 1980s, T. O. Nelson devised a technique that would achieve this goal for associative learning of word pairs (e.g., foreign-language translation equivalents). In particular, instead of having participants make a judgment immediately after each stimulus-response pair (e.g., dog–spoon, where "dog" is the stimulus and "spoon" is the response) had been studied, a short delay (about two minutes) filled with the study of other items occurred between the study and judgment of each item. Delayed judgments were collected using only the stimulus of each pair (e.g., if "dog–spoon" had been studied, the JOL prompt would be "dog– ?"). Most important, relative accuracy was substantially greater for delayed JOLs (M = 0.90) than for immediate JOLs (M = 0.38). This delayed JOL effect (Nelson & Dunlosky 1991) has been replicated numerous times and occurs for a variety of subject populations and conditions. Fortunately, delayed judgments can be easily incorporated into real-world settings; for example, after a nurse describes a medication schedule, he may then turn the conversation briefly to another topic before asking the patient if she will remember the schedule.

A general explanation of the delayed JOL effect fits well with inferential accounts of JOLs. Namely, when the stimulus alone (i.e., "dog– ?") is presented for a delayed JOL, a participant may attempt to retrieve the corresponding response and to use the outcome of this retrieval attempt as a cue for the JOL. Because this retrieval outcome would be highly predictive of subsequent test performance, delayed JOLs achieve high levels of accuracy. Put differently, the technique of delaying JOLs using only the stimulus of each pair provides diagnostic cues that can drive highly accurate inferences about memory for paired-associate learning (Nelson *et al.* 2004). Although most agree with this idea in general, there has been ongoing discussion concerning the specific cause of the delayed JOL effect, which includes debates about why retrieval outcomes at the time of delayed JOLs are diagnostic (Nelson & Dunlosky 1992; Spellman & Bjork 1992; Kimball & Metcalfe 2003; Sikström & Jönsson 2005) and whether participants engage in a full-blown retrieval attempt prior to making delayed JOLs (Nelson *et al.* 2004; Son & Metcalfe 2005). Most important, the accuracy of JOLs for predicting associative recall can be dramatically improved, and the technique to do so was in part discovered by considering how to manipulate task constraints in a way that would provide diagnostic cues for learners.

CONTROL DURING ACQUISITION: ALLOCATION OF STUDY TIME

An important reason for understanding how people monitor learning pertains to its function in the control of learning. As illustrated in Figure 6.1, monitoring may serve to control many aspects of learning and retrieval. For instance, students may decide to restudy essential classroom materials they had judged as not well learned. In this way, accurate monitoring can support efficient control, because learners can study again just those items that they (accurately) judged would benefit most from it. This intuitive conjecture was supported by Thiede (1999), who had participants study paired associates during multiple study-test trials. During the first trial, participants made JOLs and selected which items to restudy. The relative accuracy of each participant's JOLs was computed and compared to subsequent learning. As expected, participants who had more accurate JOLs on the first trial performed better on subsequent trials, presumably because accurate monitoring supported more efficient control of learning.

Given the importance of monitoring to efficient control, how monitoring serves to control memory has been scrutinized for each of the various control functions shown in Figure 6.2. Consider the work by Reder and her colleagues (e.g., Reder 1988; Reder & Ritter 1992), who investigated how people select search strategies during retrieval. More specifically, this work has demonstrated how monitoring informs people's decisions about whether to retrieve an answer to a problem (e.g., 12 × 12 = ? or 13 × 23 = ?) or to compute the answer to it. Using a variety of clever procedures, Reder has provided compelling evidence that when a person first reads a problem, the decision to retrieve or to compute an answer is controlled by a quick feeling-of-knowing judgment based on familiarity with the problem itself. If the problem immediately evokes a feeling of familiarity (e.g., 12 × 12 = ?), a person will select the strategy of retrieving the answer. If the problem triggers much less familiarity (e.g., 12 × 23 = ?), then the person would select the strategy of computing the answer. Their research represents one of many programs aimed at understanding metacognitively controlled systems; examples involving other control processes illustrated in Figure 6.2 can be identified using Table 6.3.

We now examine in some detail how monitoring is used to control study. To describe monitoring–control relationships, researchers have used variations on the following procedure. Participants first briefly study each to-be-learned item of a list and then make a JOL for each one. After this preliminary trial, each item is presented again and participants either select which items they would like to restudy (item selection) or spend as much time as they want studying each item (self-paced study time). Although several dependent measures may be computed to explore how people allocate study time, an often reported measure has been the correlation between JOLs and these measures of allocation. In Table 6.4, values are presented for item selection and termination of study, both of which correlated negatively with JOLs made on the initial study trial. A negative relationship has been found in numerous investigations (for a review, see Son & Metcalfe 2000) and indicates that people often spend more time studying (or more often select for restudy) items that are judged as less well learned.

What mechanisms are responsible for this relationship? One of the earliest explanations was inspired by system models of self-regulation that are based on negative feedback (Powers 1973). At the core of this explanation is a discrepancy-reduction mechanism in which a person continues studying an item until the learning goal set for the item has been met. That is, a person studies until the discrepancy has been reduced between the current state of learning and the desired amount of learning. Because relatively difficult items (vs. less difficult ones) will require more study time to meet a given learning goal, this discrepancy-reduction model predicts that JOLs will be negatively related to study time. This model has yielded numerous predictions that have guided research, and in doing so, some of its weaknesses have been discovered and have led to new discoveries about the flexibility and adaptivity of self-regulated study.

Concerning a weakness, the discrepancy-reduction model cannot explain why JOLs and measures of study-time allocation are sometimes positively correlated (Thiede & Dunlosky 1999; Son & Metcalfe 2000). For instance, Son and Metcalfe (2000) reported that participants chose to study items judged as difficult to learn when they had adequate time to study all to-be-learned items. However, when their participants had a limited time for study, they chose to study first those items judged as easier to learn. Similarly, Thiede and Dunlosky (1999) instructed participants that they only needed to obtain a low performance goal (i.e., to learn six items from a 30-item list), and in this case, participants also chose to restudy those items judged as easier to learn and neglected the more difficult ones. Dunlosky and Thiede (2004) dubbed such outcomes as shift-to-easier-materials (STEM) effects, because people apparently shifted from allocating more time to the difficult-to-learn items to allocating more time to the easier items.

One explanation for the STEM effect is that participants attempt to obtain task goals – in this case, a performance goal for recall – in an efficient manner. For Son and Metcalfe (2000), participants under time pressure to learn items presumably interpreted the explicit goal of "total mastery of all items" as "memorize as many items as you can in the brief time allowed." Efficiently obtaining this goal would arguably involve studying the easier items first. Metcalfe and Kornell (2005) have recently proposed that the particular kind of adaptive control reflected by STEM effects results from students' attempts to allocate the most time to those items within a region of proximal learning, which are unlearned items that a learner believes would benefit from study. Focusing one's time on a region of proximal learning may arise in many real-world contexts. For instance, a student may spend little time studying in a physics class (or even drop it) because he feels that the material is too difficult to learn and instead decide to use the time studying for another class in which he feels he could meet his learning goals in a timely fashion. The current support for a region of proximal learning mechanism is extensive (Metcalfe 2002), although it is clear that other processes are also involved in the allocation of study time (Metcalfe & Kornell 2005).

Admittedly, theories of self-regulated study are in their infancy, and it would be premature to rule out any mechanism without further investigation. Nevertheless, some conclusions are being consistently supported. First, one function of monitoring is in the control of study processes. People attempt to use monitoring to control these processes in an efficient manner, but it is also evident that they can have difficulty in efficiently controlling these processes (e.g., Reder & Ritter 1992; Dunlosky & Thiede 2004). Second, because monitoring serves to control study and retrieval, improving the accuracy of monitoring can improve the efficiency of control. Indeed, it seems almost self-evident that accurate monitoring would be critical for efficiently controlling any cognitive process, not just those involved in memory and learning. If so, the research question shifts from "Does improving accuracy support more efficient control?" to "What is the specific relationship between accuracy and efficient control?" Answering the latter question will involve discovering the level of accuracy required to achieve the most efficient control within a given domain. Perhaps only moderate levels of accuracy will be required for maximizing control in some domains, whereas the highest levels of accuracy will be required in others.

Many of the issues and questions raised about self-regulated study speak directly to improving human performance in other domains, and importantly, can be raised for almost any cognitive process. In the next section, we explore this possibility by describing some examples of how metamemory methods and theory have been applied to domains that have relevance to solving real-world problems.

APPLICATIONS OF A METAMEMORY FRAMEWORK

Our discussion of the JOL literature illustrates how a metamemory framework has been used to investigate monitoring and control processes relevant to learning. As alluded to above, such investigations have direct implications for solving real-world problems relevant to learning and memory that are only now being explored systematically. For instance, Kennedy and her colleagues (Kennedy *et al.* 2003; Kennedy & Yorkston 2004) have applied this framework to understand the monitoring and control abilities of people with acquired brain injury (ABI), with one aim being to help them overcome their learning deficits. Although ABI patients show deficits in learning, Kennedy *et al.* (2003) demonstrated that these patients can achieve relatively high levels of monitoring accuracy by using delayed JOLs. Importantly, learning is faster for ABI participants who use their highly accurate delayed JOLs to regulate study as compared to those who use their (less accurate) immediate JOLs. Thus, training ABI patients to make and use delayed JOLs in acquiring new information provides a promising route in helping them to compensate for learning deficits.

The same metamemory framework – focused on memory, learning, and retrieval – has also been used to examine the capabilities of other subject populations who demonstrate learning deficits and who may benefit from accurate monitoring and appropriate self control. These populations include children and older adults, adults with Alzheimer's disease and Korsakoff's disease, adults who are acutely intoxicated, and adults with frontal lobe damage, to name only a few. Not only is the literature too large to review here (for some excellent reviews, see Hertzog & Hultsch 2000; Perfect & Schwartz 2002; Pannu & Kaszniak 2005) but this list of populations is not close to exhaustive. Just think of any population you may be interested in that is not listed (e.g., children with attentional deficits), and this framework can be readily applied to describe (and potentially improve) their metamemory and learning abilities (e.g., Knouse *et al.* in press). Our key point here is simply that the framework outlined above can be readily applied to assessing people's metamemory capabilities and whether they can use them to enhance their learning and memory. Perhaps most encouraging, over a decade of research has demonstrated that many people who have impaired memories have intact metamemory abilities and hence may be able to use them to compensate for their impairments.

Beyond the realm of learning and memory, a metacognitive framework aimed at improving human performance can be applied to any real-world task that relies in part on cognitive processing. To foster such applications, we offer the following suggestions. First, conduct a componential analysis of the task in as much detail as possible. In this task analysis, the idea is to map out the possible components of the task that could contribute to performance. Second, for each *cognitive* component, two empirical questions inspired by the present framework can be posed to drive research that has applied relevance: (1) Can people accurately monitor how well the given component is operating? (2) Can people use monitoring to control the effectiveness of each component process? These two questions pertain readily to applications aimed at improving human performance, because they concern whether people *can* monitor and control a component process. Pursuing these two empirical questions alone within any particular domain will likely comprise a progressive and productive research program with applied merit, because doing so will naturally lead to other, more detailed questions about how metacognitive abilities may enhance performance.

In the next two sections, we explore how the present metacognitive framework can be applied to two real-world domains – jury deliberation and comprehension. Again, our exploration scratches only the surface of all domains that this framework may inform. A longer, albeit still not exhaustive, list of domains would include mathematical and insight problem-solving (e.g., Metcalfe 1998; Desoete *et al.* 2003), psychopathology (e.g., Teasdale 1999), eyewitness memory (Perfect 2002), writing expository texts (Sitko 1998), unconscious plagiarism (Carroll & Perfect 2002), and even artificial intelligence (Cox 2005). In fact, we do not even include exhaustive reviews of the extensive literature on jury deliberation (see Nietzel *et al.* 1999) or on comprehension (see Maki 1998). Instead, we wanted to illustrate how the framework has been applied to two topics with applied relevance. We decided to focus on these topics for the following reason. For jury deliberation, we describe a recent study that incorporates a kind of metacognitive judgment – called an *aggregate or global* judgment – that has been used extensively in the literature to explore people's judgment abilities, but which we did not discuss earlier. For comprehension, we highlight how the use of a metacognitive framework – methods and theory – have supported recent advances that promise to improve people's retention and comprehension of text materials. We hope these additional sections will stimulate researchers to creatively apply some of the same principles to understand more fully the role of metacognition within their domains of interest.

Juror Duties

Certainly, many different cognitive processes are recruited as a person plays his or her civic duty as a juror. Jury selection, ongoing trial activities, closing arguments, and jury deliberation will all challenge cognition as the jury works toward a verdict. Because componential analyses of these juror duties goes well beyond the scope of this chapter, we discuss only a recent example of how a metacognitive approach has been used to investigate a particularly important component of juror duty, namely, jury deliberation.

When jurors are deliberating the innocence or guilt of a defendant, they often must rely on their memories for the trial proceedings. Jurors who are most confident in their memories also tend to have the largest influence during deliberation (Kassin & Wrightsman 1988). Of course, if a juror's confidence reflects his or her actual memory for the trial, then perhaps his or her particular memory for a trial should be heavily weighted during the deliberation process. This possibility leads to the question: How accurate are jurors at judging their memories for trial proceedings?

To answer this question, Pritchard and Keenan (1999) creatively adapted standard metacognitive methods to investigate the judgment accuracy of mock jurors. The mock jurors watched a video tape of an actual murder trial, and after viewing the tape, each made a global judgment of how well he or she would do overall on a test of memory for the trial. These *global* judgments differ somewhat from item-by-item JOLs (Table 6.1) in that they refer to performance across all items. Although item-by-item JOLs could have been made about people's memories for specific aspects of the trial, global judgments not only tap monitoring of individual memories but also each person's overall efficacy for remembering (Hertzog *et al.* 1990; for a general framework that includes both item-by-item judgments and global judgments, see Dunlosky and Hertzog 2000). That is, global judgments tap people's overall self-efficacy at remembering, which may (or may not) be related to actual performance. As pointed out by the authors, "it is reasonable to think that global JOLs may be a primary factor in determining an individual juror's overall level of participation in the deliberative process. For example, a juror who thinks his or her memory is poor is likely to take a backseat in the deliberations" (p. 154). After making the global judgments, the participants then answered 30 questions about the trial.

How accurate were jurors' predictions? To answer this question, Pritchard and Keenan (1999) correlated (across participants) global judgments with mean test performance. Note that this between-person correlation is interpreted differently than the within-participant correlations introduced earlier in this chapter. The latter measure of relative accuracy indicates how accurately a specific person distinguishes between his or her own memory successes and failures. By contrast, the between-person correlation indicates the degree to which people with more confidence also perform better. Across two experiments, this correlation was not reliably different from zero, which indicates that the most confident jurors did not have the best memories for the trial. A somewhat unfortunate implication is that jurors who will likely "have the largest impact on the verdict may not [have] the most accurate [memories]" (Pritchard & Keenan 1999, p. 161).

Pritchard and Keenan's (1999) research demonstrates how metamemory methods can be applied to investigate aspects of juror decision-making: (a) identify a component of cognition that may influence task performance, in this case, memory for the trial may influence the quality of juror deliberations; (b) adopt or develop a metamemory measure that taps people's ability to evaluate that component, in this case, a global judgment about one's memory for the trial; and (c) compare the metamemory judgment to a corresponding objective measure that pertains to what is being judged, in this case, correlating people's global judgments with their actual trial memory. In any given application, researchers may need to develop new kinds of judgment specifically relevant to the particular cognitive component under investigation. For each judgment, it is critical to obtain an objective measure of performance, so that participants' biases can be described, understood, and potentially mended.

Finally, Pritchard and Keenan's (1999) research provided an answer to the first empirical question introduced at the beginning of this section, namely, "How accurately do jurors monitor their memories of a trial?" Given their poor accuracy, the next steps for this line of research may involve identifying cues that could support higher levels of accuracy to evaluate whether jurors can use them when evaluating their memories for a trial. Other empirical questions concern control processes in this area, such as: How strongly does a juror's confidence in his or her own memory influence whether the memory is offered during deliberation? And is one juror's confidence in his or her memory influential in whether other jurors are persuaded? We suspect that answering these questions will lead to prescriptions on how to improve the performance of individual jurors and jury decision making. Importantly, methods to answer these questions (and many other related ones) can be devised from those already developed in the field.

Text Comprehension

Improving people's comprehension while reading has become one of the most sought-after goals of the American educational system, and literacy is being promoted worldwide. A complete task analysis of comprehension would by no means be straightforward because of the numerous components that contribute individually and interact while a person reads. Many of these components are triggered by promiscuous associative processes that serve to construct and integrate representations of the text as a person reads (Kintsch 1998). It seems unlikely that these associative processes are privy to metacognitive monitoring, and even if they are, it is possible that attempting to monitor them would disrupt text processing. Even so, metacognitive monitoring and control processes could play a significant role in other aspects of reading, such as in evaluating one's understanding after reading and in deciding whether to reread sections of a text.

For present purposes, consider the case of monitoring one's understanding of text, which has received substantial interest in the field. In fact, Ellen Markman's (1977) work on error detection during reading was groundbreaking research for the area of metacognition (see also Glenberg *et al.* 1982). Since her seminal research, monitoring of reading and comprehension has been measured in various ways, such as by eye-tracking, think-aloud protocols collected while reading, detection of errors embedded in text, and explicit judgments of text comprehension. Reviews of these areas are available (e.g., Pressley & Afflerbach 1995; Otero 1998), so we consider here only a few outcomes from the literature on explicit judgments of learning and comprehension for text, which in method are most closely aligned with the monitoring judgments presented in Figure 6.2.

Maki and Berry (1984) were the first to investigate how accurately people could judge their learning and comprehension of text. The methods they developed have been adapted over the past two decades, but in general the procedure is as follows. A person reads a series of short texts (typically 200–400 words in length), and sometime after reading each one, a participant predicts how well he or she will perform on a criterion test. Finally, a criterion test over the text materials is administered. Relative accuracy is measured by computing within-participant correlations between judgments and test performance.

For the first two decades of research in this area, the most resounding conclusion about judgment accuracy was fraught with pessimism. Maki (1998) explains that in her own research, the mean across people's correlations for almost 30 different experiments was on average about +0.30, suggesting the possibility that "students cannot predict performance well, and that prediction is not a teachable skill" (p. 142). And although a good deal has been learned about how people *make* these judgments, a defensible claim even now is that no one has discovered a panacea for consistently obtaining high levels of metacomprehension accuracy.

As we argued in our review of JOLs, discovering techniques to boost metacomprehension accuracy may lead to improved learning. To this end, the inferential account of how people judge memory suggests that accuracy could be improved as long as cues were available that accurately predicted eventual test performance. When people judge their learning and comprehension of texts, the cues that are often available are not highly predictive of test performance. For instance, people's judgments of text comprehension and learning may be influenced by how difficult a passage is to read (Rawson & Dunlosky 2002), or by how quickly they can recall any information from the text in the moments prior to judging it (Morris 1990). Unfortunately, both these cues are not necessarily predictive of test performance and may even be misleading in some circumstances. By contrast, perhaps having participants attempt to retrieve the entire text before judging it would produce valid cues that would be predictive of how well each text was learned and comprehended. Consistent with this possibility, Thiede and Anderson (2003) found that having participants summarize a text after it was read (but prior to judging their comprehension) enhanced metacomprehension accuracy, as long as summarization was delayed after reading (cf. the delayed JOL effect, described earlier).

Moreover, Thiede *et al.* (2003) have been working with a new technique – delayed keyword generation – that promises to boost accuracy under many conditions. Keyword generation is based on the same principle as summarization (i.e., more fully retrieve information about each text at a delay prior to making judgments), except that generating keywords takes much less time and may be more practical to boost accuracy in timesensitive situations. For keyword generation, participants first read a text and then are asked to generate five keywords. They were instructed to generate keywords that captured the essence of the to-be-judged text, No other constraints were placed on the generation of keywords, so they could either come directly from the text or just be implied by it. Thiede *et al.* (2003) contrasted two groups: participants in one group generated keywords immediately after reading each text, and those in the other first read all the texts and then generated keywords for each one. Whereas predictive accuracy for the immediate generation group was relatively low, the mean accuracy for participants who predicted performance after delayed keyword generation exceeded +0.70. A take-home message here is that achieving high levels of judgment accuracy is possible even for a task as complex as text learning and comprehension.

This particular technique could be extended to other contexts in which people may need to evaluate their understanding. For instance, a patient who just received instructions from a doctor about how to administer his or her medication may take a moment to generate key ideas from the instructions. Failures to generate them may indicate incomplete understanding, which could be remedied through further discussion. After listening to a lecture, students may be encouraged to generate key ideas from each topic covered in the lecture near the end of the class period. Students may identify topics they do not fully understand, which the instructor could begin to clarify in review. Of course, we do not advocate applying this technique widely without further investigating whether it benefits performance in these or other situations. What we do advocate, however, is using the methods, analyses, and theory that have grown from a metamemory tradition to investigate the possible benefits of using these new techniques. By doing so, their limits will be uncovered as we move toward discovering newer and more potent techniques for enhancing human metacognition and performance.

SUMMARY AND CONCLUSIONS

Our main goals of the present chapter included introducing definitions, methods, and theories that have arisen from research on metamemory. The specific discussion of research on JOLs was intended to highlight how these methods have been used to understand students' self-regulated study, which has straightforward implications for enhancing student scholarship. Other approaches have been brought to bear on this problem, and we certainly are not promoting the current approach as the panacea for understanding and improving self-regulated study. Instead, the metamemory framework presented here should be viewed as one more tool that can complement other approaches for enhancing student scholarship.

Beyond reviewing the literature on JOLs and the allocation of study time, we also described how the present framework has been applied to the domains of juror decision-making and text comprehension. These represent only two of many domains where a metacognitive framework has successfully guided empirical enquiry. After a quick spin through the contents of this book, it is evident that a metacognitive framework could be applied judiciously to many other domains. Such an approach is likely to uncover people's potential at monitoring and controlling real-world tasks in various target domains, and it offers a promissory note for discovering techniques to foster this potential as a means to enhance human performance.

ACKNOWLEDGMENTS

This chapter is dedicated to Thomas O. Nelson and Louis Narens, whose scholarship on metacognition has inspired and guided our research program over the past decade and a half. Much thanks to Susan McDonald for informative discussion about juror duties and behavior.

NOTE

1 It is important to emphasize that the particular perspective on metacognition that we present is based largely on information-processing models of metacognition. And even more specifically, we highlight only monitoring and control. Although we believe the perspective presented here holds much promise for improving human performance, other perspectives and research traditions of metacognition do as well, such as those focused on self-efficacy theory or in training general metacognitive abilities. The various traditions are not mutually exclusive and likely represent different pieces of a larger whole. Given that a theoretical analysis of this larger whole is not currently available and goes beyond the scope of this humble chapter, we ask for the reader's forgiveness in biasing this review toward a single perspective.

REFERENCES

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

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.

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

Carroll, M. & Perfect, T. J. (2002). Students' experiences of unconscious plagiarism: did I beget or forget? In T. J. Perfect & B. Schwartz (eds.), *Applied Metacognition*. (pp. 145–166). New York: Cambridge University Press.

Cartwright-Hatton, S. & Wells, A. (1997). Beliefs about worry and intrusions: the metacognitions questionnaire and its correlates. *Journal of Anxiety Disorders, 11*, 279–296.

Cavanaugh, J. C. & Perlmutter, M. (1982). Metamemory: a critical examination. *Child Development, 53*, 11–28.

Cox, M. (2005). Metacognition in computation: a selective research review. *Artificial Intelligence, 169*, 104–141.

Desoete, A., Roeyers, H. & De Clercq, A. (2003). Can offline metacognition enhance mathematical problem solving? *Journal of Educational Psychology, 95*, 188–200.

Dunlosky, J. & Hertzog, C. (2000). Updating knowledge about encoding strategies: a componential analysis of learning about strategy effectiveness from task experience. *Psychology & Aging, 15*, 462–474.

Dunlosky, J., Hertzog, C., Kennedy, M. R. F. & Thiede, K. W. (2005). The self-monitoring approach for effective learning. *International Journal of Cognitive Technology, 10*, 4–11.

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 & Cognition, 27*, 1180–1191.

Dunlosky, J. & Nelson, T. O. (1994). Does the sensitivity of judgments of learning (JOLs) to the effects of various study activities depend on when the JOLs occur? *Journal of Memory & Language*, *33*, 545–565.

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

Emmelkamp, P. M. G. & Aardema, A. (1999). Metacognition, specific obsessive-compulsive beliefs and obsessive-compulsive behaviour. *Clinical Psychology & Psychotherapy*, *6*, 139–145.

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

Flavell, J. H. (2004). Theory-of-mind development: retrospect and prospect. *Merrill-Palmer Quarterly*, *50*, 274–290.

Gigerenzer, G., Hoffrage, U. & Kleinbolting, H. (1991). Probabilistic mental models: a Brunswikian theory of confidence. *Psychological Review*, *98*, 506–528.

Glenberg, A. M., Wilkinson, A. C. & Epstein, W. (1982). The illusion of knowing: failure in the self-assessment of comprehension. *Memory and Cognition*, *10*, 597–602.

Gonzalez, R. & Nelson, T. O. (1996). Measuring ordinal association in situations that contain tied scores. *Psychological Bulletin*, *119*, 159–165.

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

Hertzog, C., Dixon, R.A. & Hultsch, D. F. (1990). Relationships between metamemory, memory predictions, and memory task performance in adults. *Psychology and Aging*, *5*, 215–227.

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. & Hultsch, D. F. (2000). Metacognition in adulthood and old age. In F. I. M. Craik & T. A. Salthouse (eds.), *The Handbook of Aging and Cognition* (2nd edn, pp. 417–466). Mahwah, NJ: Lawrence Erlbaum Associates.

Hertzog, C., McGuire, C. L. & Lineweaver, T. T. (1998). Aging, attributions, perceived control and strategy use in a free recall task. *Aging, Neuropsychology, and Cognition*, *5*, 85–106.

Kassin, S. M. & Wrightsman, L. S. (1988). *The American Jury on Trial: Psychological Perspectives*. New York: Hemisphere Publishing/Harper and Row.

Kennedy, M. R. T., Carney, E. & Peters, S. M. (2003). Predictions of recall and study strategy decisions after diffuse brain injury. *Brain Injury*, *17*, 1043–1064.

Kennedy, M. R. T. & Yorkston, K. M. (2004). The effects of frontal injury on "on-line" self-monitoring during verbal learning by adults with diffuse brain injury. *Neuropsychological Rehabilitation*, *14*, 449–465.

Keren, G. (1991). Calibration and probability judgments: conceptual and methodological issues. *Acta Psychologica*, *77*, 217–273.

Kimball, D. R. & Metcalfe, J. (2003). Delaying judgments of learning affects memory, not metamemory. *Memory & Cognition*, *31*, 918–929.

Kintsch, W. (1998). *Comprehension: A Paradigm for Cognition*. New York: Cambridge University Press.

Knouse, L., Paradise, M. & Dunlosky, J. (in press). Does ADHD in adults affect the relative accuracy of metamemory judgments? *Journal of Attention Disorders*.

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. (2005). Illusions of competence in monitoring one's knowledge during study. *Journal of Experimental Psychology: Learning, Memory & Cognition*, *31*, 187–194.

Kreutzer, M. A., Leonard, C. & Flavell, J. H. (1975). An interview study of children's knowledge about memory. *Monographs of the Society for Research in Child Development*, *40*, 1–60.

Lichtenstein, S. & Fischhoff, B. (1977). Do those who know more also know more about how much they know? *Organizational Behavior and Human Performance*, *20*, 159–183.

Lichtenstein, S., Fischhoff, B. & Phillips, L. D. (1977). Calibration of probabilities: the state of the art. In H. Jungermann & G. deZeeuw (eds.), *Decision Making and Change in Human Affairs.* Amsterdam: D. Reidel.

McDonald-Miszczak, L., Hertzog, C. & Hultsch, D. F. (1995). Stability and accuracy of metamemory in adulthood and aging: a longitudinal analysis. *Psychology and Aging, 10*, 553–564.

Maki, R. H. (1998). Test predictions over text material. In D.J. Hacker, J. Dunlosky & A. C. Graesser (eds.), *Metacognition in Educational Theory and Practice* (pp. 117–144). Mahwah, NJ: Lawrence Erlbaum Associates.

Maki, R. H. & Berry, S. L. (1984). Metacomprehension of text material. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10*, 663–679.

Markman, E. M. (1977). Realizing that you don't understand: a preliminary investigation. *Child Development, 48*, 986–992.

Matthews, G., Davies, D. R., Westerman, S. J. & Stammers, R. B. (2000). *Human Performance: Cognition, Stress, and Individual Differences*. Philadelphia, PA: Psychology Press/Taylor & Francis.

Mazzoni, G. & Nelson, T. O. (eds.) (1998). *Metacognition and Cognitive Neuropsychology: Monitoring and Control Processes*. Mahwah, NJ: Lawrence Erlbaum Associates.

Metcalfe, J. (1998). Insight and metacognition. In G. Mazzoni & T. O. Nelson (eds.), *Metacognition and Cognitive Neuropsychology: Monitoring and Control Processes* (pp. 181–197). Mahwah, NJ: Lawrence Erlbaum Associates.

Metcalfe, J. (2000). Metamemory: theory and data. In E. Tulving & F. I. M. Craik (eds.), *The Oxford Handbook of Memory* (pp. 197–211). New York: Oxford University Press.

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

Metcalfe, J. & Kornell, N. (2005). A region of proximal learning model of study time allocation. *Journal of Memory and Language, 52*, 463–477.

Morris, C. C. (1990). Retrieval processes underlying confidence in comprehension judgments. *Journal of Experimental Psychology: Learning, Memory & Cognition, 16*, 223–232.

Murphy, A. H. (1973). A new vector partition of the probability score. *Journal of Applied Meteorology, 12*, 595–600.

Nelson, T. O. (1984). A comparison of current measures of the accuracy of feeling-of-knowing predictions. *Psychological Bulletin, 95*, 109–133.

Nelson, T. O. (1996a). Consciousness and metacognition. *American Psychologist, 51*, 102–116.

Nelson, T. O. (1996b). Gamma is a measure of the accuracy of predicting performance on one item relative to another item, not of the absolute performance on an individual item. *Applied Cognitive Psychology, 10*, 257–260.

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. & Dunlosky, J. (1992). How shall we explain the delayed-judgment-of-learning effect? *Psychological Science, 3*, 317–318.

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. 125–173). San Diego, CA: Academic Press.

Nelson, T. O. & Narens, L. (1994). Why investigate metacognition? In J. Metcalfe & A. P. Shimamura (eds.), *Metacognition: Knowing About Knowing* (pp. 1–25). Cambridge, MA: MIT Press.

Nelson, T. O., Narens, L. & Dunlosky, J. (2004). A revised methodology for research on metamemory: pre-judgment recall and monitoring (PRAM). *Psychological Method, 9*, 53–69.

Nietzel, M. T., McCarthy, D. M. & Kern, M. J. (1999). Juries: the current state of the empirical literature. In R. Roesch, S. D. Hart & J. R. P. Ogloff (eds.), *Psychology and Law: The State of the Discipline* (pp. 23–52). Dordrecht: Kluwer.

Otero, J. (1998). Influence of knowledge activation and context on comprehension monitoring of science texts. In D. J. Hacker, J. Dunlosky & A. C. Graesser (eds.), *Metacognition in Educational Theory and Practice* (pp. 145–164). Mahwah, NJ: Lawrence Erlbaum Associates.

Pannu, J. K. & Kaszniak, A. W. (2005). Metamemory experiments in neurological populations: a review. *Neuropsychology Review*, *3*, 105–130.

Perfect, T. J. (2002). When does eyewitness confidence predict performance? In T. J. Perfect & B. Schwartz (eds.), *Applied Metacognition* (pp. 95–120). New York: Cambridge University Press.

Perfect, T. J. & Schwartz, B. (eds.) (2002). *Applied Metacognition*. New York: Cambridge University Press.

Powers, W. T. (1973). *Behavior: The Control of Perception*. Oxford: Aldine.

Pressley, M. & Afflerbach, P. (1995). *Verbal Protocols of Reading: The Nature of Constructively Responsive Reading*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Pritchard, M. F. & Keenan, J. M. (1999). Memory monitoring in mock jurors. *Journal of Experimental Psychology: Applied*, *5*, 152–168.

Rawson, K. A. & Dunlosky, J. (2002). Are performance predictions for text based on ease of processing? *Journal of Experimental Psychology: Learning, Memory & Cognition*, *28*, 69–80.

Reder, L. M. (1988). Strategic control of retrieval strategies. *The Psychology of Learning and Motivation: Advances in Research & Theory*, *22*, 227–259.

Reder, L. M. (ed.) (1996). Implicit memory and metacognition. Mahwah, NJ: Lawrence Erlbaum Associates.

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.

Schwartz, B. L., Benjamin, A. S. & Bjork, R. A. (1997). The inferential and experiential bases of metamemory. *Current Directions in Psychological Science*, *6*, 132–137.

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

Sikström, S. & Jönsson, F. (2005). A model for stochastic drift in memory strength to account for judgments of learning. *Psychological Review*, *112*, 932–950.

Sitko, B. M. (1998). Knowing how to write: metacognition and writing instruction. In D. J. Hacker, J. Dunlosky & A. C. Graesser (eds.), *Metacognition in Educational Theory and Practice* (pp. 93–116). Mahwah, NJ: Lawrence Erlbaum Associates.

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.

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

Spellman, B. A. & Bjork, R. A. (1992). When predictions create reality: judgments of learning may alter what they are intended to assess. *Psychological Science*, *3*, 315–316.

Teasdale, J. D. (1999). Metacognition, mindfulness and the modification of mood disorders. *Clinical Psychology and Psychotherapy*, *6*, 146–155.

Thiede, K. W. (1996). The relative importance of anticipated test format and anticipated test difficulty on performance. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, *49A*, 901–918.

Thiede, K. W. (1999). The importance of monitoring and self-regulation during multi-trial learning. *Psychonomic Bulletin & Review*, *6*, 662–667.

Thiede, K. W. & Anderson, M. C. M. (2003). Summarizing can improve metacomprehension accuracy. *Contemporary Educational Psychology*, *28*, 129–160.

Thiede, K. W., Anderson, M. C. M. & Therriault, D. (2003). Accuracy of metacognitive monitoring affects learning of text. *Journal of Educational Psychology*, *95*, 66–73.

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.

Wallsten, T. S. (1996). An analysis of judgment research analyses. *Organizational Behavior & Human Decision Processes*, *65*, 220–226.

Zechmeister, E. B. & Shaughnessy, J. J. (1980). When you know that you and when you think that you know but you don't. *Bulletin of the Psychonomic Society*, *15*, 41–44.

APPENDIX

See Table 6.3 for instructions on using this Appendix for finding entry points into various areas of enquiry for metacognitive research.

How do Individuals Monitor Memory?

1. Leonesio, R. J. & Nelson, T. O. (1990). Do different metamemory judgments tap the same underlying aspects of memory? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*, 464–470.
2. Koriat (1997). See References.
3. Schwartz, B. L. (1994). Sources of information in metamemory: judgments of learning and feelings of knowing. *Psychonomic Bulletin & Review*, *1*, 357–375.
4. Kelley, C. M. & Lindsay, D. S. (1993). Remembering mistaken for knowing: ease of retrieval as a basis for confidence in answers to general knowledge questions. *Journal of Memory & Language*, *32*, 1–24.
5. Johnson, M. K. & Reeder, J. A. (1997). Consciousness as metaprocessing. In J. D. Cohen & J. W. Schooler (eds.), *Scientific Approaches to Consciousness* (pp. 261–293). Mahwah, NJ: Lawrence Erlbaum Associates.

How Accurate is Memory Monitoring?

6. Underwood, B. J. (1966). Individual and group predictions of item difficulty for free learning. *Journal of Experimental Psychology*, *71*, 673–679.
7. Nelson and Dunlosky (1991). See References.
8. Koriat (1993). See References.
9. Gigerenzer, G., Hoffrage, U. & Kleinbolting, H. (1991). Probabilistic mental models: a Brunswikian theory of confidence. *Psychological Review*, *98*, 506–528.
10. Mitchell, K. J. & Johnson, M. K. (2000). Source monitoring: attributing mental experiences. In E. Tulving & F. I. M. Craik (eds.), *The Oxford Handbook of Memory* (pp. 179–195). New York: Oxford University Press.

Can Monitoring Accuracy be Improved?

11. We are unaware of programmatic research in this area.
12. Weaver, C. A. & Kelemen, W. L. (1997). Judgments of learning at delays: shifts in response patterns or increased metamemory accuracy? *Psychological Science*, *8*, 318–321.
13. Koriat (1993). See References.
14. Arkes, H. R. (1991). Costs and benefits of judgment errors: implications for debiasing. *Psychological Bulletin*, *110*, 486–498.
15. Multhaup, K. S. (1995). Aging, source, and decision criteria: when false fame errors do and do not occur. *Psychology & Aging*, *10*, 492–497.

How is Monitoring Used to Control?

16. Son and Metcalfe (2000). See References.
17. Metcalfe and Kornell (2005). See References.
18. Reder and Ritter (1992). See References.
19. Goldsmith, M., Koriat, A. & Weinberg-Eliezer, A. (2002). Strategic regulation of grain size memory reporting. *Journal of Experimental Psychology: General, 131,* 73–95.
20. See Johnson and Reeder (1997). 5 above.