2.20 Metamemory

J. Metcalfe, Columbia University, New York, NY, USA

J. Dunlosky, Kent State University, Kent, OH, USA

Metamemory refers to the processes and structures whereby people are able to examine the content of their memories, either prospectively or retrospectively, and make judgments or commentaries about them. Thus, metamemory is not memory itself, although it may depend critically upon memory. Rather it is the judgments, assessments, or commentaries that are made about memories or learning. These kinds of self-reflective judgments have a long and controversial history. Presumably, for example, when Descartes was engaged in his famous doubting meditation – musing about how his memories or perceptions could have been different than they were, or how he could have been mistaken about them – he was engaging in metacognition. This kind of reflection was taken by him as the basis of all knowledge and the core of our phenomenological selves. Similarly, the introspectionists (with whom behaviorists later took such exception) were, presumably, engaging in what we would now call metacognition. The lack of reliability of their findings was a shortcoming that proved devastating for their method by opening the door for the behaviorists to oust the study of consciousness, at least temporarily, from the domain of respectable topics in psychology. However, the judgmental biases that were the bane of early twentieth century introspectionism are now being studied under the guise of the biases and framing effects that are both systematic and rampant in metacognitive judgments.

That these metamemory judgments can be studied objectively, and reliably, is now apparent, with many hundreds of studies having been directed at issues of human metacognition. Indeed, growing interest and research from a metacognitive perspective – with its emphasis on people's memory-based attributions – can be considered one of the most significant developments in the science of psychology in this new century. Both the processes that underlie the judgments themselves and the implications that these judgments have for self-guided control of learning are yielding to investigation. Current methods promise both enhanced understanding of impairments in metacognition and also the possibility of remedying certain biases to enable people to better assess and control their own learning.

How these judgments are made has been the focus of much research, and some of these processes are detailed shortly. Classically, three types of judgments

have formed the core of metamemory research: Feeling-of-knowing judgments, tip-of-the-tongue judgments, and judgments of learning. Although there may be some differences between feeling-of-knowing and tip-of-the-tongue judgments (see Schwartz, 2006), their similarities outweigh their differences, and we treat them together. However, the restriction to these so-called classic judgments is arbitrary, because metamemory refers to any judgment that is about a memory. The reflective quality is what is important in the definition. Thus, other judgments such as confidence judgments, source judgments, recognition judgments, and remember/know judgments are also properly considered to be metamemory. Indeed, any attribution about memory is properly considered to be metamemory, and one may even argue that all memory output relies at least partially on metamemory. For instance, if you covertly recall that the word needle was on a list of words that you just tried to memorize or that Dr. Case told me that the medication would have no side effect, you would likely not report the word needle or recommend the medication to a friend if you were not sufficiently confident in your memory (Koriat and Goldsmith, 1996). Accordingly, we also briefly discuss other memory judgments – in particular, source judgments and remember/know judgments – in the same context as the classical metamemory judgments.

Figure 1 provides an illustration of what is meant by metamemory. Nelson and Narens (1990), in a highly influential paper, argued that metacognition entailed two mental levels: an object level and a metalevel. The object level consists of the memories themselves. The metalevel involves monitoring the object level, such as reflecting upon memories and ongoing learning. When the object level is memory, such monitoring is measured by feeling-of-knowing judgments, judgments of learning, source judgments, or judgments about whether the individual remembers the event explicitly or only knows that it must have happened. The requirement is that the object of the judgment be a mental event, rather than something that is present in the environment. Many animals can make judgments about the world, but few are capable of reflecting on the objects of their minds, such as their memories (see Terrace and Metcalfe, 2005). The ability to so reflect indicates a fundamentally different kind of mental life for the animals that have it.

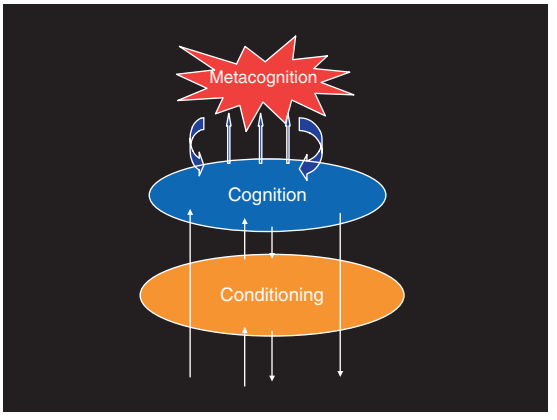

Figure 1 A model of the relations among metacognition, cognition, and conditioning. The model shows that the metacognitive level both monitors (up arrows) and controls (down, thick arrows) the contents of the cognitive level.

As can also be seen from the figure (arrows flowing from metacognition to cognition), metacognition is presumably necessary for high-level control of one's own mental processes and memories. Without knowledge of what one does not know, one could not be expected to take action to remedy the situation by, say, allocating differential study opportunities, rehearsal, or time. The metacognitive individual can choose to mould his or her own mind by self-initiated study processes, thereby learning things under self-control rather than only under stimulus control. To regulate effectively, such self-guided learning requires accurate metacognitions, of course, but it also depends on their appropriate use. If one's metacognitive judgments are inaccurate, self-regulated study could be suboptimal because the person does not know what he or she does not know. Such metacognitive failure could result because of immature metacognition capabilities or because of an impairment due to illness, stroke, or head injury. Distortions in metacognition also occur, even in normal and unimpaired people, because they are blinded by some illusion of metacognition due to the circumstances of the task at hand. Many metacognitive illusions – or biases – have now been documented by researchers (Bjork, 1994), and understanding and finding methods to debias them is fundamental if self-guided study is to succeed (Thiede et al., 2003). However, self-controlled learning and memory processing can also go awry even when a person's metacognitions themselves are excellent, if those metacognitions are not converted into optimal control strategies. One could know what one knows, but still do the wrong thing. Finally, even if one knows what one knows, and one knows what

to do about it, the actual implementation of the control knowledge could be faulty – leaving a fully metacognitive person still unable to effectively control their own learning and memory.

Whereas early research on metacognition and control focused almost exclusively on people's judgments about their memories, with the often-stated hope that this would lead to enhanced learning, recent research is increasingly aimed at the control aspect of meta-level awareness. In the section that follows, we focus on the judgments themselves, first, and on theories about how those judgments are made. We then turn to how those judgments are put to use in controlling learning and memory.

2.20.1 Metamemory Paradigms

2.20.1.1 Feeling-of-Knowing Judgments

The feeling-of-knowing judgment was the first to be systematically explored experimentally, by Joseph Hart, in 1965. Hart gave people a variety of general information questions to answer. When they could not answer a question, he asked them whether they felt they knew the answer anyway. The feeling that they knew it corresponded to their later choosing the correct answer on a recognition memory test. This paradigm posed a puzzle: How is it that people could ostensibly not know, as evidenced by their failure to produce the answer, and yet still be able to predict accurately whether they would know in the future, as evidenced by the correlations that were well above chance between their predictions and subsequent performance?

This finding of above-chance predictive accuracy has been replicated hundreds of times, so there is no doubt as to its reliability. The research in recent years has been directed not at establishing the predictive accuracy of feeling-of-knowing judgments but, rather, at understanding what cues people are using that give rise to it. Several theories have addressed this puzzle of seemingly not knowing and knowing at the same time, that is, how people are able to correctly predict what they will know in the future, when at the moment they are unable to retrieve the correct answer.

2.20.1.1.1 Theories

Whereas some early theories suggested that the person might have direct access to subliminal traces, all modern theories are basically heuristic in nature; that is, they assume that people have explicit access to some information that notably may be correct or incorrect, diagnostic or nondiagnostic, and that their feeling-of-knowing judgments are based on this information. Thus, while all current theories of this metamemory paradigm (and, indeed, of all metamemory paradigms) are heuristic theories, they differ in the exact heuristic that they propose people are using to make metamemory judgments.

2.20.1.1.1.(i) Domain and cue familiarity A logical possibility for the basis of feeling-of-knowing judgments is that people assess the familiarity of the cue (i.e., the question itself) or the domain of the question. Greater familiarity leads to higher judgments, that is, more confidence that a currently unretrieved answer will later be recognized. Concerning domain familiarity, even though people may be unable to immediately answer a question such as "Who painted *The Sunflowers*?," they may be able to assess how much they know about art and make a reasonable judgment on that basis. If they know something about art, they may be able to narrow down the field in a recognition test and eliminate incorrect alternatives. Thus, this kind of familiarity with the domain of the question may both be used to make a feeling-of-knowing judgment and be diagnostic, because, in general, strategic multiple-choice decision making will be better in well-known than in little-known domains. Thus, the person may not know who painted a particular painting but may nevertheless have a quite good idea of who did not do so, and such knowledge will help them on the test.

Glenberg and Epstein (1987) conducted an experiment in which people were selected for participation based on their expertise in various domains. They were then presented with texts to read that were either in their own domain or not. They found that people made higher judgments of knowing on those passages that were within their own area of expertise, thereby indicating that this kind of knowledge about the domain is one of the cues or heuristics that people use in making their judgments. Surprisingly, however, in this particular case, experts were not well calibrated when making judgments within their own domain. The mystery of this unexpected result remains unresolved even today. Finally, because many studies of feeling of knowing have been conducted with general information questions, and there are several domains of knowledge implicated in these questions (e.g., American history, old movies, sports, geology, capitals of various countries, etc.), knowledge of the types of general knowledge

one knows most about could be quite predictive of recognizing the correct answers.

Concerning cue familiarity, Reder and her colleagues (e.g., Reder, 1987; Reder and Ritter, 1992; Miner and Reder, 1994) conducted a series of experiments in which they showed that the familiarity of the cue influenced feeling-of-knowing judgments. For instance, Reder and Ritter (1992) presented participants with math problems (e.g., $113 + 29 = ?$) and had them quickly decide whether they wanted to retrieve or compute answers to each one. Prior to making a decision, the cue item was primed, without altering the target answer. Thus, in a math problem such as $113 + 29$, they would prime the cue by giving another problem such as $113 * 29$. When $113 + 29 = ?$ was presented, people then hit a button if they wanted to retrieve the answer as compared to compute the answer. If they already knew the answer, it would behoove them (because they would gain a greater reward) to hit the retrieve button, indicating that they could quickly retrieve the answer from memory. The interesting finding, from the perspective of the cue-familiarity heuristic, is that when the cues had been primed, people were more likely to indicate that they could retrieve the answer, even though such priming might even have hurt the retrieval of the correct answer. In a similar manner, Metcalfe et al. (1993) found that cue priming of verbal pairs influenced the feeling of knowing without altering target retrievability. In particular, they showed that the crucial factor influencing the magnitude of feeling-of-knowing judgments was the number of repetitions of the cue (which presumably would boost cue familiarity), rather than the retrievability of the sought-after target.

Whereas these and other studies (e.g., Maki, 1999; Eakin, 2005) clearly implicate the familiarity of the cue as one heuristic that people use in making metamemory judgments, evidence also suggests that partial information retrieved about the target is important.

2.20.1.1.1.(ii) Partial target accessibility The other main source of information for making feeling-of-knowing judgments is partial knowledge about the target. Perhaps one recalls that the answer to the sunflower question given earlier is an impressionist, and maybe even that there is a 'G' in the name. Even with this information, the person may be unable to give the answer. However, such partial target information may be sufficient that he or she will assign the item a high feeling of knowing. Such partial information, which is about the target itself, may be insufficient to allow the person to express the target item but may, nevertheless, indicate (often correctly) that he or she will be able to select the target in a multiple-choice test. (The only problem the person might experience in the present case could be in distinguishing Gauguin from Van Gogh, should both be present in the list). Thus, if partial or fragmentary target information is retrieved, it may be used to indicate that people will know the answer (and hence be related to high feeling-of-knowing judgments).

Koriat (1993) conducted experiments in which the to-be-remembered items were four-letter nonword strings. He showed a positive correlation between the number of letters the person could recall and their feeling of knowing rating. Of course, having three letters rather than just one was highly predictive of whether they would be able to pick the right answer from the set of alternatives offered, and so the predictive accuracy of this particular information-based metacognition was extremely high. The experiment was designed such that a 20-questions strategy was highly diagnostic, because one could eliminate half of the multiple-choice test alternatives with every letter correctly remembered. Playing 20 questions, and deliberatively assigning feeling-of-knowing judgments on the basis of the knowledge that partial information would allow them to eliminate alternatives in the test, is a logical possibility, and one that should work fairly well in the world. Phenomenologically, the judgments often feel more intuitive and less deliberative; however, even if people are less analytic about making these judgments than Koriat's experiments would suggest, if one has partial information, such as the first letter of the target, such information may give rise to a diffuse feeling that one knows more than nothing, and in many cases, one would be correct to inflate one's feeling of knowing.

It seems likely that the two mechanisms – cue familiarity and partial target information – account for most of the variability in feeling-of-knowing judgments. If so, hybrid models that describe how both cues combine (e.g., Leibert and Nelson, 1998; Koriat and Levy-Sadot, 2001) will likely fare well and are worthy of further exploration.

2.20.1.2 Tip-of-the-Tongue States

While overlapping in many respects with feeling-of-knowing judgments, tip-of-the-tongue judgments focus more directly on highly accessible partial information, and they appear less inferential in nature (for a general review, *See* Chapter 2.22). Nevertheless,

even if tip-of-the-tongue states merely represent very strong feelings of knowing, tip-of-the-tongue judgments have been investigated extensively (and separately from feeling-of-knowing judgments) because they occur so commonly in everyone's lives (Schwartz, 1999). In fact, well before the term 'metamemory' was coined, and before other metamemory judgments were scrutinized, the tip-of-the-tongue state captured the attention of William James (1890/1981). In his now-famous quote, James wrote: "Suppose we try to recall a forgotten name. The state of our consciousness is peculiar. There is a gap therein; but no mere gap. It is a gap that is intensely active. A sort of wraith of the name is in it, beckoning us in a given direction, making us at moments tingle with the sense of our closeness.… The rhythm of a lost word may be there without a sound to clothe it; or the evanescence sense of something which is the initial vowel or consonant may mock us fitfully, without growing more distinct" (James, 1890/1981: 243–244).

Schwartz (1999) has conducted a survey of 51 language groups and found that in the majority of them, there is an expression for what, in English, is called the tip-of-the-tongue state, though the exact expression varies slightly. In Korean, for example, this state is provocatively called "sparkling at the end of the tongue." This state seems to be almost universally experienced.

2.20.1.2.1 Theories

2.20.1.2.1.(i) Partial target access In a manner that is similar to the target access view of feelings of knowing, the dominant theory of tip of the tongues is that they reflect partial target access. In support of this view, a number of studies have shown that people are able to report the number of syllables in the to-be-retrieved word, some aspects of semantic content, or its first letter (for a review, see Schwartz, 2002).

2.20.1.2.1.(ii) Lexical access without phonological access Burke et al. (1991; see James and Burke, 2000) have proposed that a semantic level of representation of a sought-after word feeds to an articulatory/phonological level, which is necessary for word retrieval and output, and that the two representations can be dissociated. One dissociation is reflected by a tip-of-the-tongue state when the individual has complete or partial access at the semantic or lexical level, without being able to translate that activation into a phonological form that allows retrieval – or output – of the sought-after word. According to this model, the individual really can know an answer without being able to articulate it. Older adults seem to exhibit this phenomenon whereby an impairment occurs in phonological translation, which results in more tip-of-the-tongue states (for a recent review, see Schwartz and Frazier, 2005).

A prediction of this model is also supported by evidence from Metcalfe et al. (1995), who described an anomic patient who had difficulties retrieving words. In particular, this patient (HW), after experiencing a severe stroke, was able to converse intelligently but was unable to articulate the words for nearly all specific nouns, verbs, or adjectives when so requested. Thus, if asked to fill in the correct answer "One _____ the Thanksgiving turkey by brushing butter on while it is roasting," "The precious gem that is red is the _____," "The name of people who explore caves is _____," or even "Sirius is the _____ star in the sky excluding the sun," HW could not say bastes, ruby, spelunker, or brightest. However, he expressed a strong tip-of-the-tongue for these words. When he was later given a recognition test, he was able to pick the correct alternative with an accuracy better than that of Dartmouth College students, indicating that he knew the words he was seeking (i.e., he had semantic knowledge or lexical access) but could not articulate them. Burke et al.'s model eloquently explains HW's deficit.

2.20.1.2.1.(iii) Blocking One phenomenon seen in conjunction with tip-of-the-tongue states is that people often report that an incorrect response persistently comes to mind. This persistent alternative is usually called a blocker. We suspect that what makes tip of the tongues frustrating at times is that people in a blocked tip-of-the-tongue state know perfectly well that what keeps persistently coming to mind is wrong. Blocked tip of the tongues differ from nonblocked tip of the tongues insofar as people's phenomenology is different. In addition, it has been shown that blocked tip of the tongues tend to be more difficult to resolve than tip of the tongues without a blocker (Burke et al., 1991; Reason and Lucas, 1984). Researchers have thought that blockers actively keep people from accessing the correct answer. However, recent research by Kornell and Metcalfe (2007) indicates that this active blocking role of the so-called blockers is incorrect. In particular, they conducted an experiment to investigate the idea that blockers impaired performance, as is assumed both in the tip-of-the-tongue literature (Jones, 1989) and in the insight literature, where a similar phenomenon is thought to occur (Mayer, 1995). Theorists have stated

that people need to incubate (e.g., take a break to think about something unrelated to the problem) in both a problem-solving attempt or when attempting to retrieve a sought-after answer when in tip-of-the-tongue state. If a persistent alternative came to mind originally (which is supposed to be actively interfering with the generation of the correct solution), this break may allow one to forget it. If so, the off time will allow the problem-solver to overcome the harmful blocker and retrieve a correct solution.

To test this idea, Kornell and Metcalfe (2007) asked people to state whether their tip-of-the-tongue states included a blocker or not. The subjects then either continued to try to solve the problem or waited until the end of the experimental session for the additional minutes that they were assigned to attempt to solve the problem. As in the past literature, blocked tip of the tongues were resolved with a frequency that was lower than that of nonblocked tip of the tongues. Furthermore, consistent with the reminiscence literature, people answered more questions correctly at a delay than immediately. However, the delay interval did not particularly help the blocked tip of the tongues, as compared to the nonblocked tip of the tongues, as should have been the case had the blockers themselves kept the correct answer from appearing. Also, the blockers were forgotten over the delay interval. Thus, the delay interval did, effectively, get the blockers out of mind (as presumably should have been necessary to obviate their deleterious effect). But that made no difference for the rate of resolution, indicating that the so-called blockers do not really block. Kornell and Metcalfe (2007) favored a road sign view of blockers; they are in the person's semantic network, and the person might well articulate them in their quest for the correct answer, but they do not actively participate in the process. Whether they are accessed or not has no effect on the probability of retrieving the target.

2.20.1.2.2 Function of feeling-of-knowing and tip-of-the-tongue states

Little emphasis has been placed on the question of why people have feeling-of-knowing states or tip of the tongues. Perhaps the nagging emotional quality of the tip of the tongues is motivational and keeps people seeking an answer that otherwise they would not try to find. Similarly, Reder and Ritter (1992) have suggested that people's feelings of knowing indicate to them that there is something in memory to be found, and hence these feeling states – especially the fast feelings of knowing – provide information that people use to determine whether they will or will not attempt retrieval. Systematic research on whether and how feelings of knowing and tip of the tongues guide decision making and retrieval is needed.

2.20.1.3 Judgments of Learning

Judgments of learning are assessments that people make, either while in the course of learning, or afterwards, about how well they have learned the particular target materials under question. These judgments are thought to be of fundamental importance because the monitoring of study tapped by them is presumably used by a person to determine whether or not to study (e.g., Thiede and Dunlosky, 1999; Son and Metcalfe, 2000). Thus, if the judgments are faulty, so too will be people's subsequent study behavior. It is thought that with biased judgments, ultimately people's learning will be less than optimal.

Judgments of learning can be made in a cumulative manner, whereby the participant is asked to assess the degree of learning over an entire list or session, or they can be made on an item-by-item basis. For instance, when studying a list of 20 paired associates (e.g., dog–spoon), participants may be asked to predict how many out of 20 they will correctly recall when later tested (e.g., dog–?). While studying, they may make item-by-item judgments of learning, where participants are shown either only the cue (e.g., dog–?) or both the cue and response (e.g., dog–spoon) and are asked to predict the likelihood that the correct response (i.e., spoon) will be recalled. Item-by-item judgments of learning can be made either immediately while the person is learning or directly following that learning, or they can be made at a delay. As compared to aggregate judgments, the item-by-item judgments of learning currently have received the most empirical and theoretical attention in the field (for a comparison of the two judgments, see Dunlosky and Hertzog, 2000), so we shall largely restrict our review to them.

Two major findings have held up extremely well over the course of the last decade of research and have become the target of much further investigation. First, delayed, cue-only judgments of learning are highly accurate. The gamma correlations relating people's judgment-of-learning ratings to their later performance are often in the 0.90 range. In contrast, immediate judgments of learning and delayed judgments of learning when the cue and target are also given are often rather inaccurate, and it is not

uncommon to see the analogous gamma correlations being around +0.30. The reasons for these differences, which are tightly related to theories of how people make judgments of learning in these different conditions, are outlined below. The second major finding is that whereas first-trial immediate judgments of learning (and aggregated judgments) are often overconfident (i.e., their mean value is higher than the mean performance that people exhibit when they are tested), judgments of learning made on a second study-test trial over the same items are nearly always underconfident. Again, we discuss the explanations researchers have isolated (and those potential reasons that they have discredited) in the theoretical section that follows, titled 'Theories of the delayed-judgment-of-learning effect.'

2.20.1.3.1 Theories of the delayed judgment-of-learning effect

Four theories have been directed at the issue of why accuracy (as measured by resolution or the correlation relating judgments of learning to subsequent performance) is substantially greater for delayed than immediate judgments of learning, which has been dubbed the delayed judgment-of-learning effect (Nelson and Dunlosky, 1991). The first was the monitoring dual memories hypothesis, and the second is the transfer-appropriate processing framework. The third is the self-fulfilling prophecy hypothesis, whereby the judgment itself alters memory, and this alteration is responsible for the boost in accuracy for delayed judgments of learning. The fourth is a stochastic drift model.

2.20.1.3.1.(i) Monitoring-dual-memories hypothesis Nelson and Dunlosky's (1991) monitoring-dual-memories hypothesis assumes that judgments of learning are made by retrieving information from both short-term memory (STM) and long-term memory (LTM). In the immediate-judgment-of-learning condition, STM information is highly accessible, but it is transient and does not reflect what information will be available at final test. The presence of this STM information during the judgment, therefore, adds nondiagnostic information to the judgment, thereby reducing the accuracy of the judgments of learning. In the delayed-judgment-of-learning case, people are thought to base their judgments primarily on the retrieval of information from LTM. This retrieved information is more accurate in predicting final test performance, which is also based on LTM alone. This first explanation has a basic similarity to the second explanation – the transfer-appropriate processing explanation – insofar as both posit that the information that the person bases the judgment on is more similar to the information at time of test for the delayed than immediate judgments of learning.

2.20.1.3.1.(ii) Transfer-appropriate monitoring hypothesis The second explanation – a transfer-appropriate processing view – proposes that the delayed-judgment-of-learning effect occurs because of differences between the two judgment-of-learning conditions in the degree of contextual match from the time of the judgment to the time of the test (Begg et al., 1989; Dunlosky and Nelson, 1997). Making a judgment of learning in a situation that is as similar as possible to that of the test should maximize its accuracy. Insofar as the retrieval attempt, which is thought to be the critical information on which the judgment of learning is based, is more similar between a delayed test and a delayed judgment of learning than between a delayed test and an immediate judgment of learning, the delayed judgments are predicted to be more accurate.

2.20.1.3.1.(iii) Self-fulfilling prophecy hypothesis The third explanation locates the increase in gamma accuracy between immediate and delayed judgments of learning in a differential change in memory with immediate and delayed judgments of learning that comes with making the judgment itself (Spellman and Bjork, 1992; Kimball and Metcalfe, 2003). This third theory has been called a Heisenberg explanation or the self-fulfilling prophecy hypothesis. An assumption here is that people attempt retrieval to make their judgments of learning but, in the delayed-judgment-of-learning condition, are successful with only some of those attempts. The practice elicited by cue-only delayed judgments of learning enhances memory for retrieved items, but only some items are retrieved at the delay. Moreover, the items that receive this memory boost are not distributed randomly across the judgment of learning range, but rather are those given high judgments of learning, because the basis of the judgment is whether or not the person is able to retrieve. Those items that people fail to retrieve are given low judgments of learning and get no boost in study. Thus, the high-judgment-of-learning items benefit from an extra (spaced) study trial, while the low-judgment-of-learning items receive no additional practice and get no memory boost. This differential study has an effect on memory that bolsters the predictive

value of the ratings only in the delayed-judgment-of-learning condition. In the immediate-judgment-of-learning condition, virtually all items are recalled during the judgment (e.g., Nelson et al., 2004), which occurs immediately after study and has little memorial effect. In addition, being uniform across the entire judgment-of-learning range, this immediate retrieval does not make the high judgments of learning more memorable or the low judgments of learning less memorable.

2.20.1.3.1.(iv) Stochastic drift model Finally, Sikstrom and Jonsson (2005) propose (in a manner related to the monitoring dual memories hypothesis) that the accuracy difference is because memory strength for any given item can be decomposed into exponential functions with slow and fast components. The drift from these decay processes from time of judgments to time of test is large for immediate judgments of learning, resulting in low predictability, but is smaller for the delayed judgments, resulting in high predictability. This model is most welcome in the field for two reasons: First, because it is a much needed formal model of the processes thought to underlie the judgments and their consequences, and second, because it makes new predictions about outcomes.

2.20.1.3.1.(v) Status of theories for the delayed-judgment-of-learning effect Although considerable empirical work has been conducted to evaluate these theories (either in isolation or in competition), it is currently premature to declare one as a clear winner. Nevertheless, albeit intuitive, the transfer-appropriate monitoring hypothesis has been repeatedly disconfirmed (e.g., see Weaver and Kelemen, 2003; Dunlosky et al., 2005b). Moreover, recent modeling of the delayed-judgment-of-learning effect suggests that both a monitoring-dual-memories component and a Heisenberg-style component may be required to fully account for the effect (Jang et al., 2006).

All four of the theories explain the delayed judgment-of-learning effect by assuming that people make their judgments by using the heuristic of trying to retrieve the target, at least in the delayed case. None of these models take into account the possibility that other cues may be used to make the delayed judgments of learning. However, Son and Metcalfe (2005) have shown that people sometimes make very fast delayed judgments of learning and that these fast judgments of learning are probably not based on retrieval or attempted retrieval of the target. They showed that there were notable differences in the results when people were simply asked to make delayed judgments of learning as compared to when they were asked to attempt to retrieve the target immediately prior to making each judgment of learning (e.g., for detailed application of this method, see Nelson et al., 2004). In particular, the very fast judgments of learning drop out in the latter case, suggesting that normally people are doing something to produce these fast judgments of learning that they are not doing when they explicitly try to retrieve the target. They suggested that people are basing these fast low judgments of learning on a lack of familiarity with the cue, and that when the cue is unfamiliar, people do not bother to try to retrieve the target. In this way, they proposed a two-factor hypothesis in which familiarity and retrieval interact to influence people's judgments of learning.

Benjamin (2005) provided support for a two-factor hypothesis by showing that when people are time pressured, factors that affect cue familiarity come into play in their judgments of learning. When they are not time pressured, factors affecting the retrievability of the target are influential. Note that these are the same two cues that people use in making feeling-of-knowing judgments. With delayed judgments of learning, these cues appear to be used in a specific order. First, people assess the familiarity of the cue. If it is unfamiliar, they give a low judgment of learning. If it is familiar, they go on to the second stage, in which they attempt retrieval of the target. If they cannot do so, they give the item a relatively low judgment of learning; if they can do so, they give it a high judgment of learning. Given the evidence for the second factor in delayed judgments of learning, it appears that none of the four theories can fully account for the judgments. Regardless of its ultimate explanation, however, there is general agreement that delayed judgments of learning may be quite valuable in helping people both accurately monitor and effectively control their learning (Bjork, 1994).

The heuristics used when people make immediate judgments of learning are less straightforward than those used in making delayed judgments of learning. Data indicate that a variety of cues may play a role, such as the fluency of processing words during study (Begg et al., 1989), the fluency of generating study strategies (Hertzog et al., 2003), the relatedness of words within paired associates and across individual words (e.g., Koriat, 1997; Dunlosky and Matvey, 2001; Matvey et al., 2006), and memory for the outcome of previous tests (Finn and Metcalfe, 2007, 2008), among many others (for a review, see Koriat,

1997). Whereas some of these cues, clearly, must have some predictive value – the gamma correlations are nearly always greater than zero – they are typically less diagnostic than the cues used in delayed judgments.

2.20.1.3.2 Theories of the underconfidence-with-practice effect

The second major finding within the judgment-of-learning literature is that although people's judgments of learning tend to be overconfident on the first trial, by the second trial, there is a shift to underconfidence that persists on subsequent trials. Much research has focused on this underconfidence-with-practice effect, and a number of efforts to explain it, based on exactly how people make judgments of learning, have been proposed (as shown in **Figure 2**, from Koriat et al., 2002). Besides drawing attention to the underconfidence-with-practice effect, Koriat et al. (2002) demonstrated that it persisted despite a variety of experimental manipulations that might otherwise provide explanations of it. For example, feedback about performance on a prior trial had no effect. Both incorrectly and correctly recalled Trial 1 items showed underconfidence on Trial 2. Although this finding suggests that past test performance may not drive the effect, Finn and Metcalfe (2007) have shown that the underconfidence is significantly larger for items that were incorrect on Trial 1 than for items that were correct on Trial 1, qualifying the earlier conclusion that Trial 1 performance was irrelevant.

One possible explanation for the underconfidence-with-practice effect is that people are underconfident

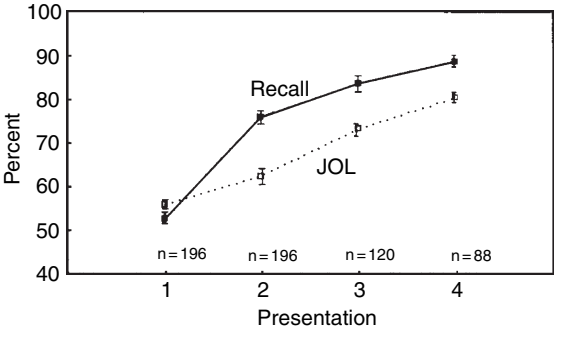

Figure 2 Illustration of the underconfidence-with-practice effect, with judgments of learning (JOL) showing overconfidence on an initial study trial and underconfidence on subsequent trials. From Koriat A, Sheffer L, and Ma'ayan H (2002) Comparing objective and subjective learning curves: Judgments of learning exhibit increased underconfidence with practice. *J. Exp. Psychol. Gen.* 131(2): 147–162.

because they have insufficient control over their own study to learn because the duration of study for each item is typically under experimental control. In contrast to this possibility, the underconfidence-with-practice effect was found when the study time allowed for each item was fixed or when it was self-paced (Koriat et al., 2002). Perhaps people just do not care and make the judgments without due consideration. However, even with incentives given for making accurate judgments – a manipulation that increases Trial 1 judgment of learning accuracy – the underconfidence-with-practice effect persisted. Thus, mere laziness on the part of participants does not appear to be the answer.

Numerous studies have shown that easy materials tend to result in less overconfidence than difficult materials (Lichtenstein and Fischhoff, 1977), so perhaps the underconfidence-with-practice effect is just another manifestation of item effects described in the confidence literature. Although possible, Koriat et al. (2002) reported that both easy and difficult items showed the underconfidence-with-practice effect. And it does not appear to be attributable to the undue effects of retrieval fluency from the first test trials (Serra and Dunlosky, 2005). Their idea was that people might assign low judgments of learning to items that were recalled on Trial 1, slowly or with great difficulty. The data, however, did not support this hypothesis.

One hint about the underconfidence-with-practice effect comes from the finding that immediate judgments of learning show the effect, whereas delayed judgments of learning do not (e.g., Meeter and Nelson, 2003; Koriat and Ma'ayan, 2005; Scheck and Nelson, 2005; Koriat et al., 2006; Finn and Metcalfe, 2007). In fact, early evidence relevant to the underconfidence-with-practice effect involved delayed judgments of learning and did not demonstrate the effect (Dunlosky and Connor, 1997). Meeter and Nelson (2003) showed only a 1% difference between delayed judgments and recall performance on Trial 2. Serra and Dunlosky (2005) showed underconfidence for both delayed and immediate judgments but a much greater shift toward underconfidence across trials for immediate judgments. Koriat et al.'s (2006) data showed overconfidence with delayed judgments of learning, though the difference from calibration was slight. Taken together, these reports suggest that delayed judgments of learning are not underconfident but, rather, are very close to being perfectly calibrated. Immediate judgments of learning, however, are nearly always underconfident after the first study-test trial.

As discussed, one difference between immediate judgments of learning and delayed judgments of learning is that people are very likely to rely on different heuristics in making the two different judgments. In the former case, as described earlier, they rely primarily on retrieval of the target item (with some reliance on familiarity of the cue). In the latter case, though, the heuristics are less clear. Finn and Metcalfe (2007) have proposed that use of the Memory for Past Test heuristic selectively in the immediate judgment of learning case, could account for much of the underconfidence-with-practice effect. The idea is that when people make second-trial judgments of learning they think back to whether they remembered that particular item in the past test. If they did, they give it a high judgment of learning. If they did not, they give it a low judgment of learning. If people were using this heuristic, they would tend to underestimate current trial performance insofar as it ignores the new learning in which the person has just engaged. Thus, they would be underconfident. The relationship between second-trial judgments of learning and Trial 1 performance would be expected to be stronger than the relationship between second-trial judgments of learning and Trial 2 performance, which it is (King et al., 1980). Furthermore, when Trial 1 test was manipulated independently of Trial 2 test, people's judgments of learning gravitated toward their manipulated first trial test performance (Finn and Metcalfe, 2008). And finally, when people were asked to simply report what they did to make the judgment, reliance on first trial test performance was a frequently given reason for the judgment given (Dunlosky and Serra, 2006). Thus, the use of this heuristic appears to be a viable candidate for explanation of the underconfidence-with-practice effect, though there are no doubt other factors that contribute to people's second-trial immediate judgments of learning (e.g., Kelley and Muller, 2006).

2.20.1.3.3 Function of judgments of learning

It is commonly believed that judgments of learning are of critical importance in learning insofar as they determine what people will choose to study and for how long they will persist (e.g., Nelson and Narens, 1990; Nelson and Dunlosky, 1991; Mazzoni and Cornoldi, 1993; Nelson and Narens, 1994; Benjamin et al., 1998; Koriat, 2000; Metcalfe, 2000). If these judgments of learning are accurate, then people will be in a position to choose to study the items that will result in optimal learning. If they are biased, or inaccurate, however, they will be unable to make such optimal choices.

Although the available evidence suggests that judgments of learning in part drive the allocation of study time, this evidence has been largely correlational, so direct experimental evidence is needed to more definitively establish that when metacognitions are manipulated people's study choice follows. Nevertheless, some demonstrations show that when people with inadequate metacognitions have been induced to make more accurate metacognitive judgments, their learning is improved. For instance, Thiede et al. (2003) had students study paragraphs and make a judgment of learning for each. Before making a judgment for a paragraph, participants were asked to generate five keywords about the paragraph that captured its essence. One group generated keywords (and made judgments) immediately after reading each paragraph, whereas another group did so after all the paragraphs were read. After reading and judging the paragraphs, (1) a test was administered about the content for each of the paragraphs, (2) participants were allowed to select paragraphs for restudy, (3) they restudied chosen texts, and (4) a final test was administered.

Several outcomes are notable. First, judgment-of-learning accuracy for predicting first-test performance was substantially greater for the delayed judgment ($+0.70$) than for the immediate group ($< +0.30$). Second, whereas first-test performance did not differ for the groups (both had a mean value a bit greater than $+0.45$ questions correct), the final test performance was much better for the delayed group (approximately 0.65 correct) than for the immediate group (approximately 0.50). Why such a difference? Fine-grained analyses showed that the delayed group, who had much better judgment accuracy, was more likely to choose paragraphs for restudy that they did not know well, and hence they made the greatest gains in learning during restudy. Without the ability to isolate these less well-known items, students' metacognitive judgments simply did not help them effectively regulate their learning. Thus, preliminary evidence is suggestive that people's metacognitions are used to allocate restudy and, more important, that at least one condition that boosts accuracy can also support more effective learning (for other relevant evidence, see Dunlosky et al., 2005a).

2.20.1.4 Source Judgments

Source judgments refer to attributions about the origins of our thoughts and memories (Johnson and Mitchell, 2002; for a review, *See* Chapter 2.19). As such, these judgments are metacognitive, being judgments about other cognitions. Such judgments are targeted when a person is asked who said a particular statement, where they heard something, whether they said something or someone else did, whether they saw the defendant rob the store or only saw him on the sidewalk afterwards, and so on.

Failures of source memory can have profound consequences. One such consequence is unconscious plagiarism. Another is a breakdown in reality monitoring, such as may be seen in psychiatric syndromes such as schizophrenia, in which a person cannot monitor whether the source is internal or external, and in which reality breaks down. Accurate source monitoring is critical for the eyewitness to a crime, but unfortunately, this kind of metacognition can be highly inaccurate.

2.20.1.4.1 Theories of source monitoring

Johnson and Raye (1981; Johnson, 1983; Johnson et al., 1993) have formulated a model, called MEM (for multiple-entry modular memory system framework), which brings together many of the findings from the source literature in a coherent and elegant form. The consensus view, articulated in the MEM model, of the mechanisms underlying source judgments is that they, like other metacognitive judgments, are based on heuristics. When asked to assess a source, people use what information comes to mind to make the judgments, and this information itself can vary radically depending upon a number of factors. For example, if two potential sources are highly similar to one another, the memory will be highly confusable and the resultant judgment will be more difficult and error prone. If they are quite different from one another, the task is easier. So, if one has to say whether Mary or Lynn said a particular sentence, if Mary is female and Lynn is male, the task is much easier than if both are female (Ferguson et al., 1992). If the two sources are spatially discrete, once again the task is easier than if they are overlapping (Ferguson et al., 1992). Physical differences of this sort have been well documented, are systematic, and conform very nicely to one's intuitions.

Interestingly, though, it is not only the conditions in the world that determine how confusable the sources of different events will be but also the individual's mental capabilities and mental operations that play a part. If a person is readily able to construct vivid images – being able to mentally see a turkey when the word turkey is read – and if he or she automatically encodes concrete nouns as images, then the source distinction of whether a word or a picture was presented will be more difficult than for a different person whose imagery capabilities are less well developed (Johnson et al., 1979). If a person is told to imagine words being spoken in a particular person's voice, which is similar to the speaker's, as opposed to imaging in a voice less similar, the source judgments will be affected (Johnson et al., 1979). The vividness of a person's imagination, then, can have a dramatic effect on whether things that actually happened are confused with those that were only imagined.

Since Johnson's seminal research in the field, the literature has grown extensively, with research involving everything from basic cognitive theory to the neurological underpinnings of source memory. Certainly, this literature is too broad to cover here (for a review, see Johnson et al., 1993; Mitchell and Johnson, 2000; Johnson and Mitchell, 2002), but in contrast to many other coverages of metamemory, we wanted to draw some attention here to this very important, and pervasive monitoring skill.

2.20.1.5 Remember/Know Judgments

People can distinguish between events or items that they remember (i.e., for which they have a clear and distinct recollection not only for the target material itself but also for the circumstances of having learned it) versus those that they only know. For example, one might remember one's first iPod, including the circumstances under which one obtained it, and so on, but only have a feeling that they know they saw such-and-such a person some time ago without being able to recall the specific episode. In typical experiments, participants will study a list of words (e.g., pencil, table, football, etc.). After study, the words are presented again mixed with new words, and participants are asked whether each item was originally presented (i.e., a standard recognition judgment), and then whether they recollect that it was presented or merely know that it was presented. In this example, you may state that you recognize that both pencil and football had been presented, but when asked for a remember/know judgment, you may recollect seeing football because you recollected that when it was

originally presented you thought of your favorite football team (e.g., the Denver Broncos), whereas you have no recollections about pencil but just have a diffuse feeling, knowing that it was presented.

Being able to tell the difference between remembering and knowing, that is, the ability to make this particular judgment about a memory, is a category of metacognition that is thought to have significance for our understanding of human consciousness (for a general review on remember-know judgments, *See* Chapter 2.17). Events that are recollected are thought to be true memories and to exemplify a special form of memory and consciousness called autonoetic consciousness (Tulving, 2005) or explicit memory (Graf and Schacter, 1985). Facts that are judged to be only known are thought to require only semantic knowledge or mere familiarity and are thought to require only primed noetic consciousness or implicit memory.

There have been many debates over the past decade about this distinction. People question whether it means that there are different systems of memory, or whether it might be due only to differences in the amount of information stored (e.g., with better-stored memories being judged as remembered and less well-stored memories being judged as merely known). One larger issue here is to whether the phenomenology of recollecting actually contributes to one's recognizing something as being previously studied versus whether this phenomenology is merely epiphenomenal; you have the experience of recollecting (e.g., that you recalled Denver Broncos when football had been presented), but this experience does not contribute to your ability to correctly recognize an item as previously studied. Advocates of dual-process models of recognition – which indicate that both familiarity and recollection influence recognition decisions – state that recollection has a causal influence on our recognition performance, whereas strength theorists claim that a single underlying memory dimension (e.g., familiarity alone) can adequately explain recognition. For the latter group, recollections merely arise from having strong memories, but the phenomenology itself is not important for understanding recognition *per se* (for competing views, see Yonelinas, 1994; Rottello et al., 2005).

Paradigms involving this distinction purportedly allow us to ascertain whether people are consciously aware of the memories. This particular metacognitive judgment, then, is one that has been extensively researched and debated. A detailed discussion of the remember/know literature is given in a separate chapter of this handbook, and so we do not elaborate further on it here. We include this section only to note that this particular judgment, like all of those outlined above, is a kind of metacognition because it involves an attribution about a memory, though one that may have considerable consequence for understanding human memory and consciousness.

2.20.2 Conclusion

Much progress has been made in understanding the mechanisms that underlie the judgments that people can make about their memories. There is considerable agreement that metacognitive judgments are heuristically based. People seem to rely on the information that they have at hand, and usually on a fairly shallow assessment of that information, to make these judgments. Because these judgments are heuristically based, systematic biases are observed. Under some circumstances, people will be underconfident or overconfident; in other situations, they can be misled. However, insofar as research is untangling those systematic biases and the reasons for them, we are increasingly in a position to help students improve their metacognitions, and hence base their learning on a firmer foundation.

References

Begg I, Duft S, Lalonde P, Melnick R, and Sanvito J (1989) Memory predictions are based on ease of processing. *J. Mem. Lang.* 28: 610–632.

Benjamin AS (2005) Response speeding mediates the contribution of cue familiarity and target retrievability to metamnemonic judgments. *Psychon. Bull. Rev.* 12: 874–879.

Benjamin AS, Bjork RA, and Schwartz BL (1998) The mismeasure of memory: When retrieval fluency is misleading as a metamnemonic index. *J. Exp. Psychol. Gen.* 127: 55–68.

Bjork RA (1994) Memory and metamemory considerations in the training of human beings. In: Metcalfe J and Shimamura A (eds.) *Metacognition: Knowing About Knowing*, pp. 185–206. Cambridge, MA: MIT Press.

Burke DM, MacKay DG, Worthley JS, and Wade E (1991) On the tip of the tongue: What causes word finding failures in young and older adults? *J. Mem. Lang.* 30: 542–579.

Dunlosky J and Connor L (1997) Age-related differences in the allocation of study time account for age-related differences in memory performance. *Mem. Cognit.* 25: 691–700.

Dunlosky J and Hertzog C (2000) Updating knowledge about strategy effectiveness: A componential analysis of learning about strategy effectiveness from task experience. *Psychol. Aging* 15: 462–474.

Dunlosky J and Matvey G (2001) Empirical analysis of the intrinsic-extrinsic distinction of judgments of learning

(JOLs): Effects of relatedness and serial position on JOLs. *J. Exp. Psychol. Learn. Mem. Cogn.* 27: 1180–1191.
Dunlosky J and Nelson TO (1997) Similarity between the cue for judgments of learning (JOL) and the cue for test is not the primary determinant of JOL accuracy. *J. Mem. Lang.* 36: 34–49.
Dunlosky J and Serra M (2006) Is the influence of test trials on judgments of learning analytic? Poster presented at the 47th Meeting of the Psychonomics Society. Houston, TX.
Dunlosky J, Hertzog C, Kennedy MRT, and Thiede KW (2005a) The self-monitoring approach for effective learning. *Cogn. Technol.* 10: 4–11.
Dunlosky J, Rawson KA, and Middleton E (2005b) What constrains the accuracy of metacomprehension judgments? Testing the transfer-appropriate-monitoring and accessibility hypotheses. *J. Mem. Lang.* 52: 551–565.
Eakin DK (2005) Illusions of knowing: Metamemory and memory under conditions of retroactive interference. *J. Mem. Lang.* 52: 526–534.
Ferguson SA, Hashtroudi S, and Johnson MK (1992) Age differences in using source-relevant cues. *Psychol. Aging* 7: 443–352.
Finn B and Metcalfe J (2007) The role of memory for past test in the underconfidence with practice effect. *J. Exp. Psychol. Learn. Mem. Cogn.* 33(1): 238–244.
Finn B and Metcalfe J (2008) Judgments of learning are influenced by memory for past test. *J. Mem. Lang.* 58: 19–34.
Glenberg AM and Epstein W (1987) Inexpert calibration of comprehension. *Mem. Cogn.* 15: 84–93.
Graf P and Shacter DL (1985) Implicit and explicit memory for new associations in normal and amnesic subjects. *J. Exp. Psychol. Learn. Mem. Cogn.* 11: 501–508.
Hertzog C, Dunlosky J, Robinson AE, and Kidder DP (2003) Encoding fluency is a cue used for judgments about learning. *J. Exp. Psychol. Learn. Mem. Cogn.* 29: 22–34.
James LE and Burke DM (2000) Phonological priming effects on word retrieval and tip-of-the-tongue experiences in young and older adults. *J. Exp. Psychol. Learn. Mem. Cogn.* 26: 1378–1391.
James W (1890/1981) *The Principles of Psychology*, vol. 1. New York: Henry Holt and Co.
Jang Y, Huber DE, and Wallsten TS (2006) Judgments of learning: Beyond gamma. Poster presented at the 47th Meeting of the Psychonomics Society. Houston, TX.
Johnson MK (1983) A multiple-entry, modular memory system. In: Bower GH (ed.) *The Psychology of Learning and Motivation: Advances in Research and Theory,* vol. 17, pp. 81–123. New York: Academic Press.
Johnson MK and Mitchell KJ (2002) Source monitoring. In: *Macmillan Psychology Reference Series Volume 2: Learning and Memory*, 2nd edn., pp. 628–631. New York: Macmillan Reference.
Johnson MK and Raye LC (1981) Reality monitoring. *Psychol. Rev.* 88: 67–85.
Johnson MK, Raye CL, Wang AY, and Taylor TH (1979) Fact and fantasy: The roles of accuracy and variability in confusing imaginations with perceptual experiences. *J. Exp. Psychol. Hum. Learn. Mem.* 5(3): 229–240.
Johnson MK, Hashtroudi S, and Lindsay DS (1993) Source monitoring. *Psychol. Bull.* 114: 3–28.
Jones GV (1989) Back to Woodworth: Role of interlopers in the tip of the tongue phenomenon. *Mem. Cognit.* 17: 69–76.
Kelley CM and Muller KA (2006) Sources of the underconfidence-with-practice effect. Paper presented at the 47th Meeting of the Psychonomics Society. Houston, TX.
Kimball DR and Metcalfe J (2003) Delaying judgments of learning affects memory, not metamemory. *Mem. Cognit.* 31: 918–929.
King JF, Zechmeister EB, and Shaughnessy JJ (1980) Judgments of knowing: The influence of retrieval practice. *Am. J. Psychol.* 93: 329–343.
Koriat A (1993) How do we know that we know? The accessibility model of the feeling of knowing. *Psychol. Rev.* 100: 609–639.
Koriat A (1997) Monitoring one's knowledge during study: A cue-utilization approach to judgments of learning. *J. Exp. Psychol. Gen.* 126: 349–370.
Koriat A (2000) The feeling of knowing: Some metatheoretical implications for consciousness and control. *Conscious. Cognit.* 9: 149–171.
Koriat A and Goldsmith M (1996) Monitoring and control processes in the strategic regulation of memory accuracy. *Psychol. Rev.* 103: 490–517.
Koriat A and Levy-Sadot R (2001) The combined contributions of the cue-familiarity and accessibility heuristics to feelings of knowing. *J. Exp. Psychol. Learn. Mem. Cogn.* 27: 34–53.
Koriat A and Ma'ayan H (2005) The effects of encoding fluency and retrieval fluency on judgments of learning. *J. Mem. Lang.* 52: 478–492.
Koriat A, Sheffer L, and Ma'ayan H (2002) Comparing objective and subjective learning curves: Judgments of learning exhibit increased underconfidence with practice. *J. Exp. Psychol. Gen.* 131(2): 147–162.
Koriat A, Ma'ayan H, Sheffer L, and Bjork RA (2006) Exploring a mnemonic debiasing account of the underconfidence-with-practice effect. *J. Exp. Psychol. Learn. Mem. Cogn.* 32(3): 595–608.
Kornell N and Metcalfe J (2007) ''Blockers'' Do Not Block Recall During Tip-of-the-Tongue States. *Metacogn. Learn.* 1: 248–261.
Leibert TW and Nelson DL (1998) The roles of cue and target familiarity in making feeling of knowing judgments. *Am. J. Psychol.* 111: 63–75.
Lichtenstein S and Fischhoff B (1977) Do those who know more also know more about how much they know? *Org. Behav. Hum. Perform.* 20: 159–183.
Maki RH (1999) The roles of competition, target accessibility, and cue familiarity in metamemory for word pairs. *J. Exp. Psychol. Learn. Mem. Cogn.* 25: 1011–1023.
Matvey G, Dunlosky J, and Schwartz B (2006) The effects of categorical relatedness on judgments of learning (JOLs). *Memory* 14: 253–261.
Mayer RE (1995) The search for insight: Grappling with Gestalt psychology's unanswered questions. In: Sternberg RJ and Davidson JE (eds.) *The Nature of Insight*, pp. 3–32. Cambridge, MA: MIT Press.
Mazzoni G and Cornoldi C (1993) Strategies in study-time allocation: Why is study time sometimes not effective? *J. Exp. Psychol. Gen.* 122: 47–60.
Meeter M and Nelson TO (2003) Multiple study trials and judgments of learning. *Acta Psychol.* 113: 123–132.
Metcalfe J (2000) Feelings and judgments of knowing: Is there a special noetic state? *Conscious. Cogn.* 9: 178–186.
Metcalfe J, Schwartz BL, and Joaquim SG (1993) The cue-familiarity heuristic in metacognition. *J. Exp. Psychol. Learn. Mem. Cogn.* 19: 851–861.
Metcalfe J, Funnell M, and Gazzaniga MS (1995) Right-hemisphere memory superiority: Studies of a split brain patient. *Psychol. Sci.* 6(3): 157–164.
Miner AC and Reder LM (1994) A new look at feeling of knowing: Its metacognitive role in regulating question answering. In: Metcalfe J and Shimamura A (eds.) *Metacognition: Knowing About Knowing*, pp. 47–70. Cambridge, MA: MIT Press.
Mitchell KJ and Johnson MK (2000) Source monitoring: Attributing mental experiences. In: Tulving E and Craik FIM (eds.) *The Oxford Handbook of Memory*, pp. 179–196. New York: Oxford University Press.
Nelson TO and Dunlosky J (1991) When people's judgments of learning (JOLs) are extremely accurate at predicting

subsequent recall: The ''delayed-JOL effect.'' *Psychol. Sci.* 2: 267–270.
Nelson TO and Narens L (1990) Metamemory: A theoretical framework and new findings. In: Bower GH (ed.) *The Psychology of Learning and Motivation,* vol. 26, pp. 125–173. New York: Academic Press.
Nelson TO and Narens L (1994) Why investigate metacognition? In: Metcalfe J and Shimamura AP (eds.) *Metacognition: Knowing About Knowing*, pp. 1–25. Cambridge, MA: MIT Press.
Nelson TO, Narens L, and Dunlosky J (2004) A revised methodology for research on metamemory: Pre-judgment recall and monitoring (PRAM). *Psychol. Meth.* 9: 53–69.
Reason JT and Lucas D (1984) Using cognitive diaries to investigate naturally occurring memory blocks. In: Harris J and Morris PE (eds.) *Everyday Memory, Actions, and Absent-Mindedness*, pp. 53–70. London: Academic Press.
Reder L (1987) Strategy selection in question answering. *Cognit. Psychol.* 19: 90–138.
Reder LM and Ritter FE (1992) What determines initial feeling of knowing? Familiarity with question terms, not with the answer. *J. Exp. Psycholo. Learn. Mem. Cogn.* 18: 435–451.
Rotello CM, Macmillan NA, Reeder JA, and Wong M (2005) The remember response: Subject to bias, graded, and not a process-pure indicator of recollection. *Psychon. Bull. Rev.* 12: 865–873.
Scheck P and Nelson TO (2005) Lack of pervasiveness of the underconfidence-with-practice effect: Boundary conditions and an explanation via anchoring. *J. Exp. Psychol. Gen.* 134: 124–128.
Schwartz BL (1999) Sparkling at the end of the tongue: The etiology of tip-of-the-tongue phenomenology. *Psychon. Bull. Rev.* 6: 379–393.
Schwartz BL (2002) *Tip-of-the-Tongue States: Phenomenology, Mechanism, and Lexical Retrieval*. Mahwah, NJ: Erlbaum.
Schwartz BL (2006) Tip-of-the-tongue states as metacognition. *Metacogn. Learn.* 1: 149–158.
Schwartz BL and Frazier LD (2005) Tip-of-the-tongue states and aging: Contrasting psycholinguistic and metacognitive perspectives. *J. Gen. Psychol.* 132: 377–391.
Serra MJ and Dunlosky J (2005) Does retrieval fluency contribute to the underconfidence-with-practice effect? *J. Exp. Psycholo. Learn. Mem. Cogn.* 31: 1258–1266.
Sikstrom S and Jonsson F (2005) A model for stochastic drift in memory strength to account for judgments of learning. *Psychol. Rev*. 112(4): 932–950.
Son LK and Metcalfe J (2000) Metacognitive and control strategies in study-time allocation. *J. Exp. Psycholo. Learn. Mem. Cogn.* 26: 204–221.
Son LK and Metcalfe J (2005) Judgments of Learning: Evidence for a Two-Stage Model. *Mem. Cognit.* 33: 1116–1129.
Spellman BA and Bjork RA (1992) Technical commentary: When predictions create reality: Judgments of learning may alter what they are intended to assess. *Psychol. Sci.* 3: 315–316.
Terrace HS and Metcalfe J (2005) *The Missing Link in Cognition: Origins of Self-Reflective Consciousness*. New York: Oxford University Press.
Thiede KW and Dunlosky J (1999) Toward a general model of self-regulated study: An analysis of selection of items for study and self-paced study time. *J. Exp. Psychol. Learn. Mem. Cogn.* 25: 1024–1037.
Thiede KW, Anderson MCM, and Therriault D (2003) Accuracy of metacognitive monitoring affects learning of texts. *J. Educ. Psychol.* 95: 66–73.
Tulving E (2005) Episodic memory and autonoesis: uniquely human? In: Terrace HS and Metcalfe J (eds.) *The Missing Link in Cognition: Origins of Self-Reflective Consciousness*, pp. 3–56. Oxford: Oxford University Press.
Weaver C and Kelemen WL (2003) Processing similarity does not improve metamemory: Evidence against transfer-appropriate monitoring. *J. Exp. Psychol. Learn. Mem. Cogn.* 29: 1058–1065.
Yonelinas AP (1994) Receiver-operating characteristics in recognition memory: Evidence for a dual-process model. *J. Exp. Psychol. Learn. Mem. Cogn.* 20: 1341–1354.