11 Metacognitive aspects of memory

Asher Koriat and Tore Helstrup

COGNITIVE AND METACOGNITIVE PROCESSES

People routinely engage in a variety of metacognitive processes when they learn new material or when they receive information that they are likely to use in the future. The more complex the material studied or the information received the more elaborate are the monitoring and control processes in which they engage. Even a simple prospective memory task in which we have to carry out a few errands may require some planning that takes into account many cognitive considerations. For example, we might decide to begin with the chore that, according to our judgment, is the one we are most likely to forget. As we complete these errands, we must “cross them out” from our mind, and of course, we might change our original plan when we are suddenly reminded of an extra errand. All these activities require not only planning our behaviour but also managing and orchestrating many cognitive operations along the way. Deficient monitoring may result either in omission errors (e.g., missing a chore or an appointment), repetition (e.g., telling the same story once again; Koriat, Ben-Zur, & Sheffer, 1988) or over-checking to make sure that a planned action has already been performed (e.g., checking that we have locked the door, see Koriat & Ben-Zur, 1988).

Successful learning (as well as teaching) requires a great deal of knowledge about the capacities and limitations of the cognitive system, about the effectiveness of different learning strategies and the effort that they require, and so on. In particular, a learner must have a realistic assessment of his or her abilities and competence. The work in developmental psychology has indicated that not only is it important that learners know the benefits of different learning strategies, but it is also crucial that they can and do implement them (Dufresne & Kobasigawa, 1989).

The effective self-management of learning and remembering requires the on-line monitoring of one’s knowledge during different phases of the process and the adaptive regulation of various cognitive operations. Thus, in everyday life we learn and rehearse new information, we retrieve information from memory and make use of it, we consult our knowledge as we try to solve problems and to plan our activities, we consider alternative courses of actions
in terms of their benefits and costs, and finally we choose to behave in one way or another. While engaging in these various cognitive processes, however, we also observe ourselves as we do so, inspecting the course of these processes, and regulating our thoughts and actions accordingly.

Thus, a distinction may be drawn between cognitive and metacognitive processes. In the terminology of Nelson and Narens (1990, 1994) this distinction amounts to that between an object level and a metalevel. The object level includes the processes that are traditionally subsumed under the rubric of information processing – encoding, rehearsing, retrieving and so on. The metalevel is assumed to supervise the processes that take place at the object level, to regulate them and navigate them towards one’s own goals.

We shall clarify this distinction using the example of a student preparing for an upcoming examination. As the student prepares for the exam, he or she makes use of a variety of cognitive processes: reading and text processing, making sense of the material by relating it to information retrieved from memory, engaging in inferential reconstructive processes designed to fill in some of the gaps encountered, organizing the materials in his or her mind, and trying to memorize it. In parallel, however, he or she engages in metacognitive processes, monitoring these cognitive processes and regulating their course according to a variety of considerations.

Metacognitive processes include two general functions – monitoring and control (Nelson & Narens, 1990). The monitoring function refers to the reflective processes involved when we observe and supervise our cognitive processes on-line and evaluate their ease, progress and success. For example, the student must assess on-line the degree to which he or she has mastered the material in order to decide whether he or she needs to continue studying or “knows” the material and is ready for the exam.

The control function refers to the regulation of the cognitive processes, and includes a variety of higher-order operations that initiate, modify, and regulate the course of basic processes. Thus, in the course of learning the student needs to choose what learning strategy to use, how much time to allocate to different parts of the materials, which parts of the materials to restudy, and when to stop studying and move on to other activities. Such regulatory control operations are normally guided by the output of the monitoring operations. However, they are also based on the students’ goals and on their beliefs about cognition. For example, as will be discussed later, learners normally allocate more study time to the items that are judged to be difficult than to those that are judged to be easy. Of course, this strategy is adaptive if the goal is to achieve a homogenous level of competence across all items (Nelson & Leonesio, 1988). However, when the goal set is relatively easy (e.g., to recall only a few of the items), participants focus on the easier items (Thiede & Dunlosky, 1999). The same is true when learning occurs under time pressure (Son & Metcalfe, 2000). Thus, the allocation of time and effort during study is affected not only by feedback from on-line monitoring (e.g., realizing that a piece of material is difficult to comprehend) but also by goals and situational
constraints. In addition, control processes are also guided by prior knowledge and beliefs: In preparing for an exam, students would choose a learning strategy that they believe to be the most effective for study under the particular conditions (e.g., expecting open-ended questions vs. expecting forced-choice questions in the test (Mazzoni & Cornoldi, 1993). But they may relinquish a particular strategy if they find it to be ineffective.

Of course, monitoring and control processes would also be involved during the exam itself. In choosing which questions to answer (when the student has that option), in allocating time between different questions, and in the “self-scoring” of one’s performance before handing in the test (Koriat & Goldsmith, 1998).

The example of the student preparing for (and taking) the exam also illustrates an important principle: The student’s success in the exam will depend not only on cognitive skills but also on metacognitive skills, that is, on the ability to monitor one’s degree of comprehension and competence during study, to allocate study time and effort effectively to different parts of the materials, and to chose and implement useful learning strategies that take into account both the qualities of the material and the nature of the expected exam. Of course, it also depends on the ability to engage in effective monitoring and control processes during the exam itself. For example, when monitoring is deficient, students may experience an “illusion of knowing”, and may stop studying prematurely (and later be surprised to receive a low grade on the test). They might also choose or write down an answer in the exam that is clearly wrong (even though they might “know” the correct answer, but either do not know that they know it or do not try hard enough to look for it). Thus, metacognitive skills are no less important than cognitive skills in determining actual performance (Bjork, 1999; Koriat & Goldsmith, 1998).

We should stress that the distinction between cognitive and metacognitive processes is not sharp. First, the same type of process may sometimes subserve a cognitive function while in other times it may be used in the context of a supervisory, metacognitive level. Second, in many cases it is difficult to specify whether a particular process belongs to the “object level” or to the “meta level”. However, the rough distinction between cognitive and metacognitive levels is useful to retain. We shall now examine some of the work on metacognitive processes.

THE INTEREST IN METACOGNITIVE PROCESSES

There has been an upsurge of research on metacognition in recent years. This interest derives from the recognition of the role that metacognitive processes play in many aspects of behavior. For example, extensive work in developmental psychology, stimulated by Flavell (1979), supports the idea that developmental changes in memory performance are due in part to the development of metacognitive knowledge and metacognitive skills.
Discussions of intelligence assume that metacognitive skills, such as the ability to plan how to perform a certain task and to monitor one’s own success in doing so, are central to intelligence (Sternberg, 1986). In the area of forensic psychology there have been many issues concerning metacognition, such as the reliability of witnesses’ reports and the extent to which witnesses can monitor the accuracy of their reports (Perfect, 2002). Of course, in memory research, questions about metacognition emerge in many different contexts such as the determinants of memory accuracy (Koriat, Goldsmith, & Pansky, 2000), the processes underlying source monitoring and source confusions (Johnson, 1997), fluency attributions and misattributions (Kelley & Jacoby, 1998), false memories and misinformation effects and how they can be escaped (Israel & Schacter, 1997), and so forth. There has also been increased interest in the neuropsychological study of brain-damaged patients demonstrating intrusions, false recognitions and confabulations (Buckner, 2003; Burgess & Shallice, 1996; Schacter, Norman, & Koutstaal, 1998). In parallel, there has also been a great deal of research in recent years on monitoring and control processes in old age (Hertzog, Kidder, Powell-Moman, & Dunlosky, 2002; Kelley & Sahakyan, 2003).

In addition, however, the interest in metacognition may also reflect a shift from the traditional behaviouristic view in which people’s behaviour is assumed to be driven by the impinging stimuli towards a more active view in which the person is assumed to have some degree of control over cognitive processes and behaviour, and to regulate his or her behaviour towards particular goals (see Koriat, 2000a, b).

This view presents a methodological dilemma. For example, we know that self-initiated monitoring and control processes take place during memory testing, and that these processes affect memory performance. Thus, the accuracy of what is reported from memory is in part under the person’s control (Koriat & Goldsmith, 1996a, b). How should such intervening processes on the part of the subject be handled? Traditionally, memory researchers tended to treat subject control as a nuisance factor that should be either eliminated (e.g., by using forced-choice tests in order to minimize subjects’ decision to volunteer or withhold an answer) or partialled out (e.g., by using a correction for guessing). Indeed, Nelson and Narens (1994) noted that although subject-controlled processes are not explicitly acknowledged in most theories of memory, “there is an implicit acknowledgment on the part of investigators concerning the importance of such processes. The evidence for this is that investigators go to such great lengths to design experiments that eliminate or hold those self-directed processes constant via experimental control!” (p. 8). Thus, the implicit assumption underlying traditional memory research is that subject-controlled processes conflict with the desire for experimental control.

However, in contrast to the tendency of laboratory studies of memory to exert strict experimental control and to minimize the contribution of self-regulation, in everyday life, people typically have great freedom in controlling
and regulating various aspects of learning and remembering: During learning they are free to choose which encoding strategies to use, how to allocate their learning resources and when to terminate study. Similarly, when attempting to retrieve a piece of information from memory, they are free to decide whether to continue searching for that information or to give up, and whether to volunteer a candidate answer that comes to mind or withhold it lest it might be wrong. Metacognitive researchers share the assumption that such self-controlled processes constitute an integral part of memory functioning (Barnes, Nelson, Dunlosky, Mazzoni, & Narens, 1999; Goldsmith & Koriat, 1999), and should be incorporated in the experimental study of cognition rather than being eliminated or partialled out.

Of course, there are many processes that occur automatically, outside the person’s consciousness and control. Evidence for the occurrence of such processes has accumulated in recent years (see Bargh, 1997). Some of these seem to involve the kind of monitoring and control operations that have been discussed by researchers in metacognition (e.g., Reder, 1988). There is still disagreement whether such processes that occur without conscious control should also be subsumed under the topic of metacognition (see Spehn & Reder, 2000).

In what follows we shall examine some of the findings on monitoring and control processes in learning and remembering. Research in this area has focused on five different questions. First, what are the bases of metacognitive judgments? Second, how accurate are these judgments? Third, what are the factors that are responsible for the accuracy and inaccuracy of monitoring? Fourth, what are the principles governing the link between monitoring and control? Finally, what are the consequences of monitoring-based regulation on actual memory performance? We shall begin by discussing some of these questions with regard to monitoring and control processes during learning.

MONITORING AND CONTROL PROCESSES DURING STUDY

We normally engage in a variety of monitoring processes in the course of learning and remembering. Consider a prospective memory situation in which we have to remember to perform some act in the future. For example, my wife reminds me “don’t forget, we have an appointment tomorrow at 8 with the insurance person”. What do I do then? Typically, I may try to assess the probability that I will remember or forget the appointment. On the basis of that assessment (which takes into account my schedule on that day, my beliefs about how good I am at remembering appointments, etc.), I may decide either to take some special measure so as not to forget, or to simply do nothing about it, being sure that I will remember it anyway. Of course, whether I eventually show up for the appointment depends not only on my memory, but also on my metamemory, that is, on the ability to correctly
assess the probability of future recall and to allocate encoding resources accordingly.

The bases of judgments of learning

How do we assess the likelihood of recalling a piece of information in the future? How do students monitor their degree of comprehension and mastery as they study new material? One common and straightforward theory – the direct-access theory – is that learners can access the memory trace that is formed during study, and can make judgments of learning (JOLs) by taking a reading of the strength of the memory trace (Cohen, Sandler, & Keglevich, 1991). For example, in studying a list of words, a learner is assumed to detect the increase in encoding strength that occurs as more time is spent studying each word. In fact, the learner can then stop studying when a desired strength has been reached. Of course, this direct-access model can also explain the accuracy of JOLs: If JOLs monitor encoding strength, they should be accurate in predicting future recall because recall also varies with memory strength.

In contrast to this model, several authors subscribe to the cue-utilization view of JOLs (e.g., Begg, Dulf, Lalonde, Melnick, & Sanvito, 1989; Benjamin & Bjork, 1996; Koriat, 1997). According to this view, JOLs are inferential in nature: Learners have no way of monitoring the strength of the memory directly but must utilize a variety of cues and beliefs to reach a reasonable assessment of future recall. Thus, they may take into account the perceived difficulty of the study items, the ease with which they come to mind during study, the number of study repetitions and the encoding strategies used, the type of memory test expected, one’s beliefs about one’s own memory efficacy, and so on.

An important distinction that has been proposed in discussing the bases of JOLs is between experience-based and theory-based JOLs (Koriat, 1997). Experience-based JOLs are assumed to rely on mnemonic cues that derive from the on-line processing of the studied items. These cues, such as encoding and retrieval fluency, give rise to a sheer experience of knowing, which can serve as a basis for the reported JOLs. Indeed, evidence has accumulated suggesting that JOLs reflect the learner’s monitoring of the ease with which studied items are processed during encoding (Begg et al., 1989; Koriat, 1997; Matvey, Dunlosky, & Guttentag, 2001). Begg et al., for example, reported results suggesting that the effects of several attributes of words (e.g., concreteness-abstractness) on JOLs are mediated by their effects on ease of processing. Other researchers have emphasized retrieval fluency rather than encoding fluency, arguing that JOLs are based on the ease and probability with which the to-be-remembered items are retrieved during learning (Benjamin & Bjork, 1996). Using cue-target paired-associates, Matvey et al. (2001) found that JOLs increased with increasing speed of generating the targets to the cues at study, and Hertzog, Dunlosky, Robinson, and Kidder (2003) also found that JOLs increased with the success and speed of forming
an interactive image between the cue and the target. Taken together, these results support the view that JOLs are based on the fluency of perceiving or retrieving targets at study.

Turning next to theory-based judgments, there is little doubt that people make use of their a-priori theories about memory in making JOLs. Theory-based JOLs rely on the deliberate application of metacognitive beliefs or theories about one’s memory skills and about the way in which various factors can affect memory performance (see Dunlosky & Nelson, 1994; Koriat, 1997; Mazzoni & Kirsch, 2002). For example, in making JOLs, people may take into account beliefs about how good they are in retaining certain kinds of information (e.g., names, faces) and how different learning strategies may affect memory performance. For example, JOLs appear to draw on the belief that generating a word is better for memory than reading it (Begg et al., 1991; Matvey et al., 2001). The contribution of metacognitive beliefs has been spelled out most clearly by developmental psychologists (e.g., Flavell, 1979; see Koriat, 2002) in the context of children’s memory functioning, but such beliefs clearly influence adults metacognitive judgments as well (see Koriat, 1997).

It should be stressed that unlike theory-based JOLs, which rely on an analytic, deliberate inference, experience-based JOLs are based on non-analytic, contentless cues such as encoding and retrieval fluency. These cues are typically used unconsciously, and their effects are automatic. The non-analytic basis of metacognitive judgments is responsible for the phenomenal quality of the feeling of knowing as an immediate, unexplained intuition, similar to that which is associated with the experience of perceiving (see Kahneman, 2003).

**Dissociations between predicted and actual recall**

The clearest evidence in support of the idea that JOLs are based on inference from cues comes from observed dissociations between JOLs and actual recall performance. Benjamin, Bjork, and Schwartz (1998) had participants answer several questions and then assess the likelihood that they would be able to recall the answer in a free-recall test. They found that the more rapidly participants retrieved an answer to a question the higher was their estimate that they would be able to recall that answer at a later time. In reality, however, the opposite was the case. These results imply that the accuracy of JOLs is not guaranteed (as might have been the case if JOLs were to monitor memory strength), but depends on the validity of the cues on which these judgments are based.

Another type of dissociation has been recently observed by Koriat, Bjork, Sheffer, and Bar (2004). They had participants study a list of paired associates and make JOLs for tests that were expected either immediately after study, a day after study or a week after study. Assuming that JOLs monitor processing fluency during study, then they should be expected to exhibit
insensitivity to the expected time of testing. Indeed, whereas actual recall dropped considerably with retention interval, JOLs were entirely indifferent to the expected retention interval. The result showed dissociation between JOLs and recall such that JOLs matched closely actual recall for immediate testing, whereas for a week’s delay they were considerably inflated.

It is interesting to note that when a new group of participants was asked to estimate how many words learners would recall after each of the three retention intervals, their estimates matched very closely the first group’s actual recall, exhibiting a clear forgetting function. This finding suggests that in making theory-based predictions, participants draw upon their beliefs about forgetting, but do not do so when their predictions rely on their immediate subjective experience.

The validity of JOLs in predicting recall

Although, as noted above, dissociations have been observed between JOLs and recall under some circumscribed conditions, these are the exceptions rather than the rule. By and large learners are moderately accurate in predicting recall success. As early as 1966, Underwood showed that participants can estimate with some accuracy which items should be easier to learn and which should be difficult to learn. In addition, learners can monitor their degree of mastery of studied material on line (e.g., Dunlosky & Nelson, 1994; Mazzoni & Nelson, 1995): They can estimate roughly the percentage of items that they will recall (absolute accuracy), and can also say which items they will recall and which they will not (relative accuracy). In most studies, relative accuracy (or “resolution”) has been indexed by the within-person gamma correlation between JOLs and recall (Nelson, 1984). This correlation reflects the degree to which a learner can discriminate between what he will recall and what he will not.

However, there are situations in which monitoring is particularly poor. An example is when monitoring concerns one’s own performed actions (e.g., Cohen et al., 1991; Koriat, Ben-Zur, & Druch, 1991). Thus, when participants perform a series of mini-tasks (so called Self-Performed Tasks, or SPTs; see Chapter 3) and asked to indicate the likelihood of recalling these tasks in the future, the accuracy of their predictions tends to be much poorer than that of monitoring the recallability of different words in a studied list.

Evidently, it is important to seek procedures that can improve JOL accuracy. Two such procedures have been found to be effective across several experiments. The first is repeated practice studying the same list of items. Several experiments confirmed that the accuracy of JOLs in predicting future recall improves with repeated study–test cycles of the same list of items (King, Zechmeister, & Shaughnessy, 1980; Koriat, Sheffer, & Ma’ayan, 2002; Mazzoni, Cornoldi, & Marchitelli, 1990). Koriat (1997) proposed that this improvement occurs because (1) with increased practice studying a list of items, learners shift from basing JOLs on the pre-experimental attributes of
the items towards greater reliance on mnemonic cues (e.g., processing fluency) associated with the study and retrieval of these items, and (2) mnemonic cues tend to have greater validity in predicting recall than pre-experimental cues, being sensitive to the actual processing of the items.

The second procedure that was found to improve JOL accuracy is that of soliciting JOLs not immediately after studying each item, but a few trials later. In paired-associate learning, delayed JOLs, prompted by the cue alone, have been found to be considerably more accurate than immediate JOLs or delayed JOLs prompted by the entire cue-target pair (Dunlosky & Nelson, 1992; Nelson & Dunlosky, 1991). Presumably, the condition in which JOLs are delayed and cued by the stimulus alone approximates the eventual cued-recall test. Indeed, Nelson, Narens, and Dunlosky (2004) reported evidence that in making delayed JOLs, participants rely heavily on the accessibility of the target. When JOLs are solicited immediately after study, the target is practically always retrievable, and hence its accessibility has little diagnostic value. Koriat and Ma’ayan (2005) also observed that whereas immediate JOLs rely primarily on the ease with which the item is encoded, as JOLs are further delayed they tend to be based primarily on the ease with which the target comes to mind, which is a better predictor of later cued recall.

**Illusions of knowing during learning**

Everyday experience suggests that students sometimes exhibit an illusion of competence, holding unduly high expectations about their future performance (see Dunning, Johnson, Ehrlinger, & Kruger, 2003; Metcalfe, 1998). What are the mechanisms that can instil such illusions?

Bjork (1999) discussed several conditions of learning that tend to enhance performance during learning but impair long-term retention. These conditions, according to Bjork and Bjork (1992), facilitate “retrieval strength” but not “storage strength”. As a result learners may experience an illusion of competence, resulting in inflated predictions about their future performance. For example, whereas massed practice generally results in better performance than spaced practice on the short term, spaced practice yields considerably better performance on the long term (e.g., Bahrick, 1979). Therefore massed practice causes learners to overestimate their future performance (see Zechmeister & Shaughnessy, 1980). Indeed, in Simon and Bjork’s (2001) study, massed (blocked) practice inflated participants’ predictions of their future performance: Participants asked to learn each of several movement patterns under blocked conditions predicted better performance than they did when those patterns were learned under random (interleaved) conditions, whereas actual performance exhibited the opposite pattern.

Koriat and Bjork (2005) also described a condition that has the potential of creating an illusion of competence during learning: Because JOLs are made in the presence of information that is absent but solicited during testing, the failure to discount the effects of that information when making JOLs
can instil an illusion of competence. Koriat and Bjork presented evidence suggesting that such illusions occur when the target (or answer) presented during study activates aspects of the cue (or question) that are not likely to come forward during testing when the cue (or question) appears alone.

In sum, the cue-utilization approach to JOLs has the advantage of explaining the conditions that lead to accurate and inaccurate JOLs. The results obtained thus far suggest that JOLs are sensitive to mnemonic cues that are revealed on-line during encoding. The advantage of cues such as encoding and retrieval fluency is that they are generally sensitive to a variety of factors that affect actual recall, such as level of processing, prior presentation, and exposure duration. Hence they are generally diagnostic of future memory performance. Under some conditions, however, these cues can also mislead metacognitive judgments. In general, JOLs will be accurate to the extent that processing fluency at the time of making JOLs incorporates the same demands as later recall (see Benjamin et al., 1998; Groninger, 1979). Conditions that produce an illusion of competence tend to involve different demands at study and test.

**On-line control processes during learning**

As noted earlier, much of the work in metacognition assumes a causal effect of monitoring on control. Thus, we might expect learners to use their monitoring judgments during study as a basis for the controlled, strategic regulation of learning. A classic demonstration of this idea is the relationship between JOLs and study time in self-paced learning: When learners are allowed to control the amount of time spent on each item, they generally allocate more time to items that are judged to be difficult to learn than to those that are judged to be easy to learn (for a review see Son & Metcalfe, 2000). This observation has been taken to indicate that learners use their JOLs as a basis for regulating the allocation of study time, investing more effort in the study of difficult items in order to compensate for their difficulty (Nelson & Leonesio, 1988).

Dunlosky and Hertzog (1998) proposed a discrepancy-reduction model according to which learners specify a desired level of memory strength that they wish to reach – a level that was referred to as the “norm of study” (Le Ny, Denhiere, & Le Taillanter, 1972). As they study the material, they monitor continuously the increase in memory strength that occurs as more time is spent studying each item, and cease study when the pre-set norm of study has been reached.

Son and Metcalfe (2000), who reviewed the literature regarding the relationship between item difficulty and self-paced study time, indeed found that in 35 out of 46 published experimental conditions, learners exhibited a preference for studying the more difficult materials. However, as noted earlier, there are exceptions to this rule. For example, Thiede and Dunlosky (1999) presented participants with an easy goal: to learn a list of 30 paired-associates
with the aim of recalling at least 10 of those. When participants indicated which items they wished to study, they were more likely to choose the easier items rather than the more difficult items. Thiede and Dunlosky proposed a model in which there is a superordinate level of control, which concerns pre-study and planning decisions that are made in order to maximize the efficiency of study and minimize the effort invested in it. Son and Metcalfe also showed that under high time pressure, participants tend to invest more study time in items that are judged as easy rather than on those that are judged as more difficult.

These results suggest that people adopt an adaptive strategy, choosing to focus on the easier items when time pressure is strong or when they have an easy goal. Presumably students preparing for an exam will do the same when they do not have enough time to spend studying or when they only want to pass the exam rather than receiving a high grade. The adaptive, goal-driven nature of study time allocation is also revealed by studies indicating that learners invest more study time when they expect a recall test than when they expect a recognition test (Mazzoni & Cornoldi, 1993), and more time when the instructions stress memory accuracy than when they stress speed of learning (Nelson & Leonesio, 1988).

An important question that emerges concerns the effectiveness of the policy of study time allocation for enhancing memory performance. Unlike expectations from the discrepancy-reduction model (Dunlosky & Hertzog, 1998), Metcalfe and her associates (Metcalfe, 2002; Metcalfe & Kornell, 2003) observed that learners allocated most time to medium-difficulty items and studied the easiest items first. In parallel, they observed that when study time was manipulated by the experimenter rather than self-paced, medium-difficulty items benefited more from increased presentation duration than did easy or difficult items. These results were taken to suggest that learners adopt an effective strategy of study time allocation when allowed to pace their study.

MONITORING PROCESSES DURING REMEMBERING

We shall turn next to the monitoring and control processes that occur during retrieval. As early as 1970, Tulving and Madigan claimed that one of the truly unique characters of human memory is its knowledge of its own knowledge. They argued that genuine progress in memory research depends on understanding how the memory system not only can produce a learned response or retrieve an image but also can, rather accurately, estimate the likelihood of its success in doing it.

Ever since this statement was made, a great deal of work has been conducted on the feeling of knowing (FOK) that sometimes accompanies the search for a memory item. William James has provided a poetic description of the feeling that accompanies the tip-of-the-tongue (TOT) state, when we
struggle to retrieve an elusive name or word from memory. The TOT state is interesting because it combines two seemingly inconsistent features: The person is unable to retrieve the sought target, but at the same time has a strong feeling of knowing, and can sometimes monitor the emergence of the elusive target into consciousness.

The bases of feelings of knowing when recall fails

The discrepancy between subjective and objective indices of knowledge naturally raises the question: How do we know that we know? When we fail to retrieve a name from memory, how do we know that it is “there”? As with JOLs solicited during study, we can distinguish between two general explanations for the basis of the FOK that is sometimes experienced when recall fails. One explanation is based on the idea of direct access. Hart (1965) proposed that FOK judgments are based on accessing a special memory-monitoring module that can directly inspect the information stored in memory to determine whether the solicited target is stored in memory or not. Thus, whenever a person is required to recall a target, the monitoring module is activated to make sure that the target is present in store before attempting to retrieve it. Such a monitor, then, can save the time and effort looking for a target that is not in store. The important feature of the direct-access model is that it also offers a straightforward explanation for the accuracy of the FOK: If the FOK directly monitors the presence of the target in memory, then it ought to serve as a valid predictor of actual memory performance. In fact, if this view is endorsed, it should be the inaccuracy of the FOK that would need explanation.

More recent approaches, however, assume that FOK judgments are inferential in nature. Two types of inferential processes have been assumed to underlie FOK judgments. First, these judgments may be based on beliefs and information retrieved from memory. For example, a person may remember an episode in which she has used the word or name that she now fails to recall. Such information-based judgment typically involves a conscious and deliberate inference. Second, FOK judgments may be based on a sheer subjective feeling, as when a person “senses” that a name is on the tip of the tongue. It has been argued that FOK judgments that are based on subjective experience are also inferential in nature because the feeling that one knows is itself a product of implicit heuristics (Koriat & Levy-Sadot, 1999). These heuristics, like those underlying JOLs, may occur below full consciousness to influence and shape subjective experience and the feeling of knowing. Thus, it has been proposed that rememberers have no privileged access to information that they fail to retrieve, but must infer the presence of that information in memory on the basis of what they can retrieve (Koriat, 1993).

Two heuristic-based accounts have been proposed to underlie experience-based FOK judgments: the cue familiarity and accessibility accounts. According to the cue familiarity hypothesis, FOK is based on the familiarity
of the pointer that serves to probe memory, not on the retrievability of the target itself (Metcalfe, Schwartz, & Joaquim, 1993; Reder, 1988). Thus, a rapid preliminary FOK is routinely and automatically elicited by the familiarity of the terms of a memory question, and this FOK governs question-answering strategy. Indeed, in several studies, the advance priming of the terms of a question (assumed to enhance the familiarity of the question) was found to enhance speeded, preliminary FOK judgments without correspondingly raising the probability of recall or recognition of the answer (Reder, 1988; Schwartz & Metcalfe, 1992). Additional support for the cue-familiarity account comes from studies using a proactive-interference paradigm (Metcalfe, Schwartz, & Joaquim, 1993), also from and arithmetic problems (Reder & Ritter, 1992; Schunn, Reder, Nhouyvanisvong, Richards, & Stroffolino, 1997). Consistent with this account is also the finding of Glucksberg and McCloskey (1981) that increasing the familiarity of questions for which participants do not know the answer increases the latency of “don’t know” responses to these questions.

According to the accessibility account, in contrast, FOK is based on the overall accessibility of pertinent information regarding the solicited target (Koriat, 1993, 1994). Even when retrieval fails, people may still retrieve a variety of partial clues and activations, such as fragments of the target, semantic and episodic attributes, and so on. These partial clues may induce the subjective feeling that the target is stored in memory, and that it will be recalled or recognized in the future. An important assumption of the accessibility account is that participants have no direct access to the accuracy of the partial clues that come to mind, and therefore utilize the accessibility of correct and wrong partial clues indistinguishably.

Support for the accessibility account comes from a study that examined the nature of word definitions that consistently induce a TOT state (Koriat & Lieblich, 1977). The results suggested that the critical factor is the overall amount of partial information they tend to precipitate, regardless of whether that information is correct or not.

Koriat (1993) had participants study a nonsense string and then attempt to recall as many of the letters as they could, and make FOK judgments regarding the probability of recognizing the correct string among lures. The results indicated that FOK judgments increased with the number of letters that participants reported regardless of the accuracy of these letters. Thus, FOK increased as a function of the number of correct letters and also as a function of the number of wrong letters reported. When the number of letters reported was held constant, FOK judgments also increased with the ease with which information came to mind, as reflected in the latency to initiate recall.

If FOK judgments increase with the accessibility of both correct and incorrect partial information, why are they nevertheless accurate in predicting correct recall or recognition of the target? Koriat (1993) argued that this is because much of the information that comes spontaneously to mind is
correct. That is, when a piece of information comes spontaneously to mind during remembering, it is much more likely to be correct than incorrect. Therefore, the total amount of partial information accessible is a good cue for recalling or recognizing the correct target.

**Dissociations between knowing and the feeling of knowing**

The assumption that FOK judgments are inferential in nature, being based on a variety of cues, implies that they need not be always accurate. Indeed, the findings supporting the cue-familiarity account of the FOK demonstrate a dissociation such that the advance priming of the cue enhances FOK judgments without correspondingly affecting recall. A similar dissociation, consistent with the accessibility account, was reported by Koriat (1995) using different classes of general-information questions. Such questions typically bring to mind more correct (partial or complete) information than incorrect information. Hence, FOK judgments based on that information are expected to be correct by and large, as was found to be the case. However, for a minority of questions – so-called deceptive questions (Fischhoff, Slovic, & Lichtenstein, 1977) – people tend to produce predominantly incorrect information (e.g., “What is the capital of Australia?”, “In which US state is Yale University located?”). For such deceptive questions FOK judgments made following recall failure were found to be negatively correlated with subsequent recognition memory performance, presumably because these questions bring to mind partial clues that are predominantly wrong. Thus, FOK judgments tend to be accurate as long as the questions bring to mind more correct than incorrect partial information.

In sum, there is sufficient support for the idea that FOK judgments are based on mnemonic cues such as cue familiarity, partial information about the target, and the ease with which information comes to mind. In fact, there is evidence for a two-stage model in which both cue familiarity and accessibility are assumed to contribute to the FOK, but whereas the effects of familiarity occur early, those of accessibility occur later, and only when cue familiarity is sufficiently high to drive the interrogation of memory for potential answers (Koriat & Levy-Sadot, 2001; Vernon & Usher, 2003).

As far as the accuracy of FOK judgments is concerned, these judgments are accurate by and large in predicting future recall and recognition (Schwartz & Metcalfe, 1994). However, their accuracy varies with the validity of the cues on which they rest.

**The control effects of feelings of knowing**

Positive feelings of knowing generally drive memory search: When people feel that they know the answer to a question, they try harder to look for it than when they feel that they do not know the answer (Barnes et al., 1999; Costermans, Lories, & Ansay, 1992; Gruneberg, Monks, & Sykes, 1977).
People also spend more time searching for a solicited target when they experience a TOT feeling than when they do not (Schwartz, 2001).

More generally, Reder (1988) argued that preliminary FOK judgments guide the selection of strategies that people use to answer questions or solve problems. For example, Reder and Ritter (1992) had participants make fast judgments whether they knew the answer to an arithmetic problem and could retrieve it directly, or whether they had to compute it. “Know” judgments were found to increase with increasing frequency of previous exposures to the same parts of the problem, not with availability of the answer. Thus misled FOK judgments can misguide the decision to retrieve or compute the answer.

The regulation of memory retrieval is affected not only by FOK judgments but also by other considerations. For example, when participants are penalized for slow responding, they retrieve answers faster but produce more incorrect answers (Barnes et al., 1999).

RETROSPECTIVE CONFIDENCE IN ONE’S MEMORY PRODUCTS

Even after retrieving an answer from memory or choosing an answer from among distractors, we can generally monitor the likelihood that that answer is correct.

The bases of subjective confidence and its accuracy

Much of the work in this area has been conducted in the framework of judgment and decision making using forced-choice questions. The typical finding is that people are generally overconfident in the correctness of the answers that they choose (Lichtenstein, Fischhoff, & Phillips, 1982). Part of this overconfidence is possibly due to a confirmation bias: In making their confidence judgments, people selectively review the evidence that entered into making the choice, focusing on the evidence that favours the chosen answer and discounting the evidence against it (Koriat, Lichtenstein, & Fischhoff, 1980). However, this bias may also stem in part from a biased selection of items by experimenters, with an overrepresentation of challenging “deceptive” items (Gigerenzer, Hoffrage, & Kleinbölting, 1991).

While subjective confidence may be based on an analytic process in which the overall support for the chosen or produced answer is consciously evaluated, there is evidence that it may also rest on such mnemonic cues as the ease with which the answer has been reached. Indeed, it has been observed that the more effort and the longer the deliberation needed to reach an answer, the lower is the confidence in that answer (e.g., Costermans et al., 1992; Nelson & Narens, 1990; Robinson, Johnson, & Herndon, 1997). Kelley and Lindsay (1993) specifically showed that when priming speeds up the emergence of an
answer, confidence judgments also increase accordingly. This effect occurred even for incorrect answers. Typically, however, correct answers are associated with shorter latencies than incorrect answers, so that latency of responding is generally a valid diagnostic cue for the correctness of the answer.

The strategic regulation of memory accuracy

How does subjective confidence affect behaviour? Clearly, the more confident a person is in the correctness of an answer or a decision, the more he or she is likely to commit self to it. A good example is the case of a person on the witness stand. According the conceptual framework proposed by Koriat and Goldsmith (1994, 1996a, b), an eyewitness who is sworn to tell the truth and nothing but the truth, must monitor the subjective likelihood that a memory response that comes to mind is correct, and then determine whether to volunteer that response or not. The decision to volunteer or withhold a candidate answer is assumed to depend on the confidence associated with that response relative to a control threshold that is pre-set on the basis of the relative utility of providing as complete a report as possible versus as accurate a report as possible. A response is provided if its associated confidence exceeds the threshold but is withheld otherwise. The results on the whole supported this model. First, several results suggest that the decision to volunteer or withhold an answer is based almost entirely on the subjective confidence in the correctness of that answer, when other factors are held constant. Thus, the within-subject correlation between confidence and volunteering averaged 0.95 or more. Furthermore, rememberers were found to rely heavily on their subjective confidence even when the accuracy of subjective confidence was very limited. Thus, monitoring is a critical determinant of control.

Second, when rememberers were allowed freedom to choose whether to volunteer an answer or not, their memory accuracy was much higher than when they were forced to answer each and every question. This finding can explain the impression that peoples’ reports are more accurate under everyday, naturalistic conditions than in the laboratory (e.g., Neisser, 1988). Clearly, in everyday life people are generally allowed much more freedom in reporting information from memory than is the case in the laboratory. How do rememberers enhance the accuracy of their report? They do so by screening out answers that are associated with low confidence, and to the extent that confidence is diagnostic of accuracy, they can thereby enhance the accuracy of what they report. However, because monitoring effectiveness is typically not perfect, the enhanced accuracy comes at the expense of memory quantity performance, because participants also sacrifice some of the correct answers. The implication is that as long as monitoring effectiveness is not perfect, eyewitnesses cannot both “tell the whole truth” and “tell nothing but the truth”. Only when monitoring is perfect can a person volunteer all correct responses that come to mind and withhold all incorrect responses. When monitoring is not perfect a quantity–accuracy tradeoff would be
observed, the magnitude of which should decrease with increasing monitoring effectiveness.

Third, under free-report conditions, participants’ screening policy was found to vary with the incentives for accuracy. When participants were given high incentives for accuracy (e.g., a high penalty for wrong answers) they were more conservative in their reporting, and in fact achieved a higher memory accuracy performance but at a greater expense in memory quantity. Thus, memory accuracy (unlike memory quantity, see Nilsson, 1987) seems to be under the control of the rememberer.

Koriat and Goldsmith’s model has been found to apply to adults as well as to children. In a study that included school-age children (Koriat, Goldsmith, Schneider, & Nakash-Dura, 2001) it was found that even second-to third-grade children used the option of free report effectively to enhance the accuracy of what they reported. Furthermore, they exhibited sensitivity to accuracy incentive, achieving higher memory accuracy performance when accuracy incentive was high than when it was low. These results have implications for the dependability of children’s testimony in legal settings.

The conceptual framework of Koriat and Goldsmith (1996b) was also extended to incorporate another means by which people normally regulate the accuracy of what they report: the control over grain size (Goldsmith, Koriat, & Weinberg-Eliezer, 2002). Thus, when not completely certain about the time of an event, a person may simply report that it occurred “early in the morning” rather than “at 7:30 am”. Neisser (1988) pointed out that when he solicited responses to open-ended questions the participants tended to provide answers at a level of generality at which they were not likely to be mistaken. Indeed, Goldsmith et al. (2002) observed that when participants are given the option to control grain size, they may choose to sacrifice informativeness (degree of precision) for the sake of accuracy, and will tend to do so when their subjective confidence is low. By regulating the grain size of their answers people may be able to achieve a relatively high level of accuracy even when a great deal of information has been forgotten.

In sum, in everyday life people have great freedom in reporting an event from memory: they can choose what perspective to adopt, what to emphasize and what to skip, how much detail to provide, and so on. The results obtained so far suggest that they regulate their reporting flexibly and effectively to achieve certain goals. It is our view that such strategic regulation processes are part and parcel of memory, and must be incorporated into memory research.

In concluding this chapter we should note that there are many other aspects of metacognition that have not been covered here and that are quite common in everyday life. One is source monitoring and reality monitoring (Johnson, 1997). We can remember when and where we have last met a certain person. Sometimes reality monitoring is difficult: I may wonder whether I actually performed a certain action or only planned to do it. Obsessive-compulsive people may go back several times to check whether they have
locked the door because by the third or fourth time they are not certain any longer whether they checked that the door was locked or only intended to do so (Reed, 1985). The confusion between reality and imagination is probably responsible for the imagination inflation effect (Garry & Polaschek, 2000): Imagination sometimes leads to memories for events that have not happened, and increases confidence that these events have actually taken place.

Metacognitive processes also occupy an important role in problem solving. For example, in the course of attempting to solve a problem people can sometimes judge whether they are on the right track to the solution (Carlson, 1997). Metcalfe and Wiebe (1987), for example, distinguished between incremental problems and insight problems. They found that feelings of warmth increased gradually as the problem neared completion. This, however, was only true for incremental problems, whereas insight problems are usually solved suddenly, without any subjective warning signals.

To sum up, learning and remembering in everyday life typically entail many metacognitive processes that are used by people to optimize their performance and to adapt to a variety of circumstances. Children gradually learn more about their memory skills and memory limitations; they acquire new strategies and learn about the usefulness of these under different conditions. Their success in school and their ability to solve problems depend heavily on the efficient evaluation of their performance and on the self-management of strategies of learning and remembering. However, such skills are also critical for adult performance, and many problems in real life derive from deficient metacognitive skills rather than from deficient cognitive skills. Thus, deficient monitoring can be hazardous in many real-life situations, as when a driver overestimates his ability to overtake a car. Similarly, illusions of knowing and overconfidence can result in failures and frustrations. Thus, the study of metacognition can have important theoretical as well as practical implications.

CONCLUSION

In this chapter we have outlined several threads of research in metacognition that are concerned with the processes that occur during encoding, during retrieval and during memory reporting. We have shown how these lines of research bring to the fore the importance of self-assessment and self-regulation during various stages of information processing. While demonstrating the operation of these processes we have also attempted to stress their contribution to effective memory functioning. Effective monitoring of one’s knowledge and effective regulation of one’s cognitive processes represent an essential component of adaptive functioning, and affect one’s memory performance.

Clearly, however, the experimental study of metacognition has so far incorporated only a fraction of the complexity of metacognitive processes
that occur constantly in everyday life: when we plan and carry out a series of errands, monitoring their completion as we go; when we solve simple problems, such as a crossword, deciding where to start, what strategy to adopt, and how to probe our memory for the solution; when we prepare a talk, trying to adopt the perspective of the audience, and so on and so forth. Clearly people in everyday life have a much greater freedom in regulating their learning and memory processes than participants have in the metacognitive experiments described in this chapter. If we are to understand the complex dynamics of memory processes, we must allow investigation of the broad variety of monitoring and control processes that take place in everyday life.

REFERENCES


Everyday Memory
Everyday Memory

Edited by Svein Magnussen and Tore Helstrup