1

Why Investigate Metacognition?

Thomas O. Nelson and Louis Narens

Why should researchers of cognition investigate metacognition? This chapter constitutes one answer to that question.

Metacognition is simultaneously a topic of interest in its own right and a bridge between areas, e.g., between decision making and memory, between learning and motivation, and between learning and cognitive development. Although the focus of this chapter is on the metacognitive aspects of learning and memory — which throughout the chapter will be called *metamemory* — both the overall approach and many of the points apply as well to other aspects of cognition. Emphasis is placed on some shortcomings in previous research on memory that have been commented on by several prominent investigators. It is to those investigators' credit that they stepped back from their specific investigations to take stock of the overall progress in the field and to highlight problems. We believe those problems can be solved, with research on metacognition playing a major role in that solution.

Previous Research

In a well-known book, Kuhn (1962) wrote that science proceeds by alternating between periods of "normal science" (during which investigators do research within a commonly accepted paradigm) and "crises" (during which investigators seek a new paradigm due to problems with the old one). This account of science has been attacked strongly (e.g., Shapere, 1971; Suppe, 1977), but it may never-

theless be useful here as a heuristic conceptualization. Although no single paradigm has completely dominated the research on human learning and memory during the past 50 years, there have been identifiable frameworks that large numbers of researchers have investigated in unison.

Prior to the 1950s, the aim was to unify psychology via a science of all behavior. Learning and motivation were investigated as interconnected phenomena. During subsequent decades, a shift occurred away from animal research and toward research on human memory via information processing; learning became deemphasized, and motivation became "assumed" and was no longer investigated. The next few paragraphs expand some on that shift.

In the 1950s and early 1960s, researchers focused on topics such as multiple-list learning that were important within the framework of interference theory (Underwood & Postman, 1960), but that focus of research waned during the later 1960s. For instance, Postman (1975) concluded that "interference theory today is in a state of ferment if not disarray" (p. 327).

During the 1960s and early 1970s, the emphasis changed from learning to memory, and researchers focused on topics such as serial-position curves in single-trial recall, which were important within the framework of the rehearsal-buffer model of memory (Atkinson & Shiffrin, 1968). That focus was later replaced by investigations of various kinds of orienting tasks during incidental memory. Within the levels-of-processing framework of Craik and Lockhart (1972), memory was construed as a byproduct of perceptual activity rather than as a deliberate consequence of rehearsal. However, by 1980, Wickelgren concluded, "The levels of processing fad is over in the field of learning and memory" (p. 40).

During the 1980s, the field became even more fragmented into isolated pockets of research on various aspects of learning and memory, with no dominating theory or framework that most researchers are working on in unison.[1] Interest increased in taxonomic distinctions (e.g., explicit memory versus implicit memory) and in neuropsychological factors (Shimamura, 1989). There has also been a renewed interest in the topic of consciousness, with an especially compelling case having been made recently by Flanagan (1992, pp. 11–13 ff) for a three-pronged approach to investigating con-

sciousness via phenomenological reports, behavioral data, and research on the brain (e.g., neuropsychological research). However, it is not so much that the substantive problems researched in earlier years have been solved and that their solutions have been integrated into a growing body of knowledge; rather, the previous problems have been left unsolved, and new problems have became the focus of subsequent research.

Thus the net result of 50 years of research on learning and memory has been a particularly rapid series of Kuhnian alternations of "normal science" and "revolutions," with the effects of prior research on subsequent research being remarkably shortlasting. Although this series has produced rich and varied sets of empirical findings, experimental paradigms, and modeling techniques, it has not produced dominant theories or frameworks that expand on their predecessors. This failure to produce theories and frameworks that encompass the findings of prior decades is undoubtedly an important factor for the relatively slow rate of *cumulative* progress[2] in learning and memory when compared to, for example, major subfields of physics, biology, and chemistry. We believe that this failure and the lack of cumulative progress in human learning and memory are due at least partly to the following three shortcomings that have been commented on by several prominent investigators. Those comments are brought together here, and the major goal of the remainder of this chapter is to offer the beginnings of a foundation designed to facilitate cumulative progress.

Three Shortcomings of Previous Research

There are three shortcomings that are from our (and several other investigators') perspective undesirable. These shortcomings are interrelated, and each tends to give rise to the next.

First Shortcoming: Lack of a Target for Research

The bulk of laboratory research on human memory lacks concrete targets. A target for research should be defined in terms of some to-be-explained behavior of a specific category of organism in a specific kind of environmental situation (cf. Neisser, 1976). Scientific fields

typically make the most progress when they have targets outside the laboratory on which to focus (e.g., planetary motion in astronomy, earthquakes in geology, tornadoes in meteorology). Gruneberg, Morris, and Sykes (1991) concluded, "In general terms, it seems to us self-evident that everyday phenomena are the starting point for many questions for all sciences, and that all sciences progress by refining and controlling variables within the laboratory. . . . Compared with the successes of the other sciences, the successes of psychology in general, and memory research in particular, are pretty small beer" (p. 74). Thus the hope is that such naturalistic targets will give the successive programs of research a common goal to continue investigating, so that progress can be cumulative. Although there are exceptions (e.g., Bahrick, 1984), most laboratory research on memory is oriented more toward esoteric laboratory phenomena that are of interest primarily to researchers (i.e., *fachgeist*[3]) rather than toward a concrete target *outside the laboratory* that the laboratory investigations are attempting to illuminate. Similarly in the domain of theoretical models, Morris (1987) concluded, "The choice and development of models of human cognition seems to depend very much upon the personal interests of the modellers and very little upon the empirical and practical demands of the world" (p. xv).

Some people have reacted so strongly against these trends as to suggest that laboratory experiments are no longer appropriate as a research strategy for human memory (e.g., Wertheimer, 1984). By contrast, we believe with Neisser (1976) that our goal should be "to understand cognition in the context of natural purposeful activity. This would not mean an end to laboratory experiments, but a commitment to the study of variables that are ecologically important rather than those that are easily manageable" (p. 7). Similarly, Roediger (1991) wrote, "The traditional role of naturalistic observation is to draw attention to significant phenomena and to suggest interesting ideas. Researchers will then typically create a laboratory analog of the natural situation in which potentially relevant variables can be brought under control and studied" (p. 39). Such laboratory research could then serve as the basis for an integrative theory that has obvious relevance to at least one naturalistic situation.

A similar plea for a naturalistic target has been echoed by Parducci and Sarris (1984):

The desire for ecological validity, expressed in a number of the chapters, cannot be separated from the concern to make psychology more practical. . . . Scientists continue to study psychological problems without apparent concern for practical applications. . . . There do seem to be strong forces pushing even traditional areas of psychological research in practical directions. Granting agencies, particularly in the U.S., have recently been favoring "mission" research. (pp. 10–11)

Although the remarks of these researchers are useful in telling us what we should not be doing (namely, studying a laboratory phenomenon for its own sake), they do not offer a specific suggestion for what we should be doing. Before researchers can focus on specific kinds of ecologically valid situations, it may be desirable to specify the categories of people and the kinds of naturalistic situations that will be the target of the research. This is rarely done, as pointed out by Estes (1975):

The entire array of conceptual systems — association theory, functionalism, and behavior theory — which dominated research on both human and animal learning over the first half of the century had in common a view of a hypothetical ageless organism. . . . The tendency to theorize in terms of an abstract organism may seem unnecessarily sterile, making cognitive psychology both autistic relative to other disciplines and remote from practical affairs. (p. 6)

Estes' opinion was recently echoed by two well-known psychologists this year. Shepard (1992) wrote, "The experimental tasks used in the 1950s by Estes and others (including myself!) continued the existing tradition in American psychology of designing stimuli and tasks on the basis of prevailing theoretical ideas, with little regard to what types of problems the species was adapted to solve in its natural environment" (p. 420). Similarly, the reviewer Boneau (1990) concluded:

Psychological research is too faddish. Movements in research are tied too much to the development of a paradigm or methodology. The problem should be the driving force and the paradigms and methodologies developed for it. The problem should be one that is closely tied to the natural world. Experimental psychology has had too much tendency to go off into the lab and forget all contact with reality. (p. 1594)

Thus there is a need to make explicit both the specific categories of people and the specific environmental situations that are to be

the targets of research on human learning and memory. So what would be a good target on which investigators can focus? Although various targets[4] are possible, the one emphasized in our research is the following: *To explain (and eventually improve) the mnemonic behavior of a college student who is studying for and taking an examination.* We chose this target in part for the following reasons: It is relevant (who spends more time memorizing for and taking examinations than college students?), naturalistic, practical, concrete, and challenging in terms of theory.

Investigators of human memory already do the bulk of their research on college students, but usually only for reasons of convenience (e.g., because such people are easily accessible as subjects for experiments). Rather than trying to understand college students per se, the target of most investigators is, as pointed out in the previous quotation from Estes, vague and typically consists of little more than a hope that the results will generalize to some unspecified target population. By contrast, if we began explicitly to define the population of college students as a target population of interest rather than merely being the population that is handy, the design of our experiments on memory would likely change accordingly (examples are given below). Further, such an approach would help to make explicit some potentially interesting mnemonic processes that previously have been implicit and unexplored.

Second Shortcoming: Overemphasis on a Nonreflective-Organism Approach

In most previous and current research, human memory is conceptualized narrowly, almost in *tabula-rasa* fashion analogous to a computer storing new input on a disk. Although people can be regarded as encoding and retrieving information (perhaps analogously to what occurs in a computer), those activities have been assumed to be *nonreflective.* Indeed, nothing approaching consciousness is evident in any available computer (Searle, 1992). To our knowledge, *none* of the currently available computerized learning/memory algorithms contains a model of itself and its monitoring and control capabilities (ramifications of this can be seen in the discussion of Conant & Ashby below); instead, only the programmer has a model of the comput-

erized learning algorithm and its processes.[5] Ramifications of this point have been elaborated by Searle (1992). Moreover, *computers do not have the imperfect retrieval of stored information that is so characteristic of humans* (Tulving & Pearlstone, 1966; Bahrick, 1970). Whereas current theories of human learning and memory typically construe people as automatic systems, sound theories need to be developed that construe people as *systems containing self-reflective mechanisms for evaluating (and reevaluating) their progress and for changing their on-going processing;* such mechanisms do occur in the domain of metacognition, as discussed below.

One way in which the nonreflective-organism approach manifests itself is exemplified by research on different orienting processes during incidental memory, where the assumption is made that researchers can discover what is automatically stored in memory whenever a subject makes a given orienting response. Although this assumption may sometimes be valid, it certainly cannot capture the fact that a college student studying for an examination is a conscious, self-directed organism who is continually making memory-relevant decisions about how difficult it will be to memorize a given item or set of items, about what kind of processing to employ during that memorization, about how much longer to study this or that item, and so on. No current theory of memory sheds light on (or even attempts to explain) that fact.

Thirty years ago, Miller, Galanter, and Pribram (1960) remarked about the focus of research on human learning and memory: "The usual approach to the study of memorization is to ask how the material is *engraved* on the nervous system . . . an important part of the memorizing process seems to have been largely ignored" (p. 125, italics added). Later, Reitman (1970) expanded on that view, saying, "Memory is not a simple decoupleable system; it is more like a complex interconnected collection of structures, processes, strategies, and controls. Memory behavior does not depend solely upon a memory subsystem; it reflects the activity of the human cognitive system as a whole" (p. 490). Still later, Estes (1975) discussed the importance of "the formulation of the conception of 'control processes' (Atkinson & Shiffrin, 1968) in human memory and the recognition that learned voluntary strategies play a major part in virtually all aspects of human learning" (p. 7).

Viewing people as self-directed seems most compatible with the conception of people as steering their own acquisition and retrieval. We are suggesting not that studies of experimenter-directed learning and memory should cease but rather that substantial research is also warranted on self-directed learning and on the self-reflective mechanisms that people do/could use to facilitate acquisition and retrieval. Some progress has already been made (e.g., Johnson & Hirst, 1991; Nelson & Narens, 1990), and early steps toward such a theory will be discussed below.

Third Shortcoming: Short-Circuiting via Experimental Control

Another potential shortcoming of previous research is a methodological ramification of investigators construing their subjects as nonreflective. Ironically, although the self-directed 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! Two examples serve to illustrate.

First, instead of investigating how and why a subject distributes his study time, most investigators present every item for the same amount of time, typically with instructions to the subject to focus on only the current item. This was noticed by Miller et al. (1960) when they remarked, "People tend to master the material in chunks organized as units. This fact tends to become obscured by the mechanical methods of presentation used in most experiments on rote learning, because such methods do not enable the subject to spend his time as he wishes" (p. 130). Similarly, Carlson (1992) wrote about the "elaborate efforts to hide from subjects such information as that certain items are repeated in studies of memory or learning. Such efforts are, of course, a backhanded acknowledgment of the powerful causal role of consciousness in determining behavior" (p. 599). By contrast, newer approaches have explored how subjects distribute their study time and what the consequences of those activities are (Hall, 1992; Mazzoni & Cornoldi, 1993; Nelson, 1993; Nelson & Leonesio, 1988; Zacks, 1969).

Second, instead of investigating the strategies that a subject spontaneously uses to memorize a given set of items (note: compelling evidence for such strategies comes from the now-classic findings of subjective organization by Bousfield, Bower, Tulving, and others during the 1950s and 1960s), most investigators tell the subject what strategy to use. If the investigator were not concerned that the subject might spontaneously use a strategy different from the instructed one, then the investigator would not bother instructing the subject about which strategy to use!

Others have also noticed this research style of trying to minimize the learner's self-directed processing. For instance, Butterfield and Belmont (1977) concluded:

> In spite of the recent emergence of the executive function as a general theoretical construct, there has been very little effort to study it. Indeed, because of its very complexity, Reitman (1970) advocated a method of minimizing the executive by systematically instructing subjects to use highly specific sequences of control processes. This procedure assigns the executive function to the experimenter, rather than to the subject, in an effort to reduce unexplained variability in dependent measures resulting from spontaneous executive decisions by the subjects. (p. 282)

Thus investigators attempt to eliminate or reduce their subjects' variations in self-directed processing because (1) such processing on the part of the subject is typically construed mainly as a source of noise (as discussed below), and (2) until recently, there have not been theoretical frameworks within which to systematically explore the subjects' self-directed processing. Although the research strategy of attempting to minimize variations in self-directed processing (e.g., via giving instructions to the subject about how to rehearse the items) is legitimate for investigating the main effects of such instructions, there is also a need for a research strategy that investigates self-directed processing.

Sometimes the person's role in directing his or her own processing is not even acknowledged. For instance, "instructions to use imagery" may degenerate into "the person's use of imagery yielded." Moreover, people do not necessarily follow the experimenter's instructions to use a particular encoding strategy. Eagle (1967) found that subsequent recall was uncorrelated with strategy instructions per se but was correlated with people's reported strategies; strategy instructions

served only to shift the number of people who reported using one or another strategy (for additional confirmation, see Paivio & Yuille, 1969). Although the investigator can instruct the person to use imagery, if the person believes that imagery should not be used, the result may be quite different from that of another person who receives the identical instructions but who believes that imagery should be used (cf. MacLeod, Hunt, & Mathews, 1978).

The approach of trying to minimize variations in self-directed processing also prevents the investigator from discovering what kind of strategy the subject would spontaneously use in the situation under investigation. In contrast to the approach to research that minimizes or disregards self-directed processing, research on metacognition emphasizes the potential importance of self-directed processing.

Toward a Theory of Metacognition

Twenty years ago, Tulving and Madigan concluded in the *Annual Review of Psychology:*

> What is the solution to the problem of lack of genuine progress in understanding memory? It is not for us to say because we do not know. But one possibility does suggest itself: why not start looking for ways of experimentally studying, and incorporating into theories and models of memory, one of the truly unique characteristics of human memory: its knowledge of its own knowledge. (1970, p. 477).

Some investigators have begun to explore this possibility under the label of "metacognition" (Flavell, 1979; for prototypes of research on metacognition, see Nelson, 1992). These investigations have been fruitful, indicating that such an approach may indeed yield the kind of progress that Tulving and Madigan called for but could not find in 1970.

Critical Features of Metacognition

Conant and Ashby (1970) proposed and interpreted a theorem that "the living brain, so far as it is to be successful and efficient as a regulator for survival *must* proceed, in learning, by the formation of a model (or models) of its environment" (p. 89, their italics). They

concluded, "There can no longer be any question about *whether* the brain models its environment: it must" (p. 97, their italics). This idea has important implications for psychology (e.g., Yates, 1985; Johnson-Laird, 1983; Rouse & Morris, 1986).

In addition to a model of itself, two additional critical features are needed so as to have a metacognitive system, and they are summarized in figure 1.1. The first is the splitting of cognitive processes into two or more specifically interrelated levels. Figure 1.1 shows a simple metacognitive system containing two interrelated levels that we will call the "meta-level" and the "object-level." (Generalizations to more than two levels are given below.) The second critical feature of a metacognitive system is also a kind of dominance relation, defined in terms of the direction of the flow of information. This flow — analogous to a telephone handset — gives rise to a distinction between what we will call "control" (cf. Miller et al., 1960) versus "monitoring" (cf. Hart, 1965). When taken together with the aforementioned idea that the meta-level contains a model of the object-level, these two abstract features, splitting into two interrelated levels (meta-level versus object-level) and two kinds of dominance relations (control versus monitoring), comprise the core of metacognition as we use the term. These two features are interpreted in the following way.

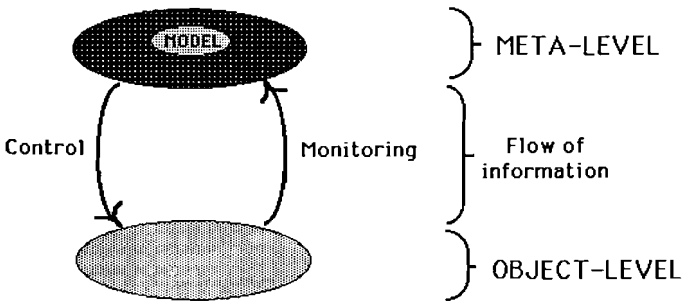

Figure 1.1
Nelson and Narens' (1990) formulation of a meta-level/object-level theoretical mechanism consisting of two structures (meta-level and object-level) and two relations in terms of the direction of the flow of information between the two levels. (Note: The meta-level contains an imperfect model of the object-level.) Adapted from Nelson and Narens (1990).

Control is interpreted as follows:

The fundamental notion underlying control — analogous to speaking into a telephone handset — is that the meta-level *modifies* the object-level, but not vice versa. In particular, the information flowing from the meta-level to the object-level either changes the state of the object-level process or changes the object-level process itself. This produces some kind of action at the object-level, which could be: (1) to initiate an action; (2) to continue an action (not necessarily the same as what had been occurring because time has passed and the total progress has changed, e.g., a game-player missing an easy shot as the pressure increases after a long series of successful shots); or (3) to terminate an action. However, because control per se does not yield any information from the object-level, a monitoring component is needed that is logically (even if not psychologically) independent of the control component. (Nelson & Narens, 1990, p. 127)

Monitoring is interpreted as follows:

The fundamental notion underlying monitoring — analogous to listening to the telephone handset — is that the meta-level *is informed by* the object-level. This changes the state of the meta-level's model of the situation, including "no change in state" (except perhaps for a notation of the time of entry, because the rate of progress may be expected to change as time passes, e.g., positively-accelerated or negatively-accelerated returns). However, the opposite does not occur, i.e., the object-level has no model of the meta-level. (Nelson & Narens, 1990, p. 127)

More Than Two Levels

The distinction between meta-level and object-level can easily be generalized to more than two levels. The development here will be given for monitoring; a similar development, except in the opposite direction (cf. figure 1.1), could occur for control.

During monitoring, the meta-level uses information about the object-level — and perhaps[6] about the relationship between the object-level and still other levels for which that level is in turn a meta-level. This information is used to update the meta-level's model of what is occurring at the object-level. This multilevel idea of processing extends naturally to finitely[7] many levels, $L_0,...L_i,...,L_j,...L_N$, where the first level, L_0, processes information about only the object, and L_j processes information about lower level L_i (where $i < j$) and perhaps about interrelationships between this lower level L_i and other levels

for which L_i is a meta-level (e.g., level L_h for which $h < i$). Then level L_j is acting as a "meta-level" and all the aforementioned levels (i.e., L_0, L_h, L_i) as "object-levels." Simultaneous with that, L_i is acting as an "object-level" for L_j and for perhaps still higher levels.[8]

Thus the critical concern of our analysis of metacognitive monitoring is not the absolute levels in the sequence but rather is the relational aspect, wherein some processes dominate others via control and monitoring. The boundary between object-level versus meta-level (e.g., recalling an answer versus reporting that an answer was recalled, respectively) is sometimes sharp and at other times may be more fuzzy.

The overall system can process information by using all of the various levels, with each level being concerned about different aspects of the situation (cf. Minsky, 1985). In contrast to the view that memory is dissociated from higher level strategies, our view is that almost all memory situations intimately involve some monitoring and control, which are important heuristic categories of organization for our framework. The members of those two categories are defined denotatively (i.e., ostensively), using the general guidelines elaborated earlier (viz. "control" refers to affecting behavior, whereas "monitoring" refers to obtaining information about what is occurring at the lower levels). Next, we will describe some subdivisions of each of these two categories for the area of metamemory.

Control Processes in Metamemory

An early formulation of control processes was illustrated by servomechanisms such as a thermostat that controlled the onset and offset of a furnace so as to yield a desired temperature (Bateson, 1972; Wiener, 1948). Servomechanisms were investigated by psychologists during the 1950s in human-factors research, especially on the role of feedback during motor learning. The formulations of self-directed control in human verbal learning were called "control elements" (Estes, 1972), "control processes" (Atkinson & Shiffrin, 1968), or "executive processes" (Greeno & Bjork, 1973).

However, there are some important differences between those formulations and the one in Nelson and Narens (1990). In the latter, the input stemming from the meta-level is to the object-level mech-

anism itself, such that the meta-level can modify the object-level mechanism. In the aforementioned formulations, by contrast, the control process merely provided input that the object-level mechanism worked on.

This distinction can be illustrated by looking more closely at the thermostat example. In the earlier formulations, the thermostat was conceptualized merely as an on-off switch that provided input to activate or deactivate the furnace; the thermostat never changed the internal workings of the furnace in any way. By contrast, if the Nelson-Narens formulation were applied to a temperature-regulation situation, the control processes could be conceptualized not only in terms of starting and stopping the furnace but also in terms of altering the way in which the furnace worked (e.g., the input might cause the fan belt on the furnace motor to tighten so as to change the speed at which the blower dispenses air into the vents).

Another difference between the Nelson-Narens formulation (versus most previous formulations) is that the meta-level is assumed to be operating simultaneously with the object-level, not sequentially as in most (but of course, not all) computers. The meta-level and object-level processes are assumed to be operating simultaneously on different aspects of the situation and perhaps working at different temporal rates. This departure from previous formulations was made by Broadbent (1977; who in turn cites the earlier views of Kenneth Craik and Bartlett) when he wrote:

> There are two concepts which have been current recently, and which might be used to explain the classic findings. . . . First, there is the notion of separate stages in the nervous system. In this notion, information about an event is processed in one way in one place and then passed on to another place where different operations are performed. The second notion is that of transfer of control, where a single processor is supposed to carry out one operation, store the result, and then carry out a different set of operations, in response to instructions from a different region of the program. These two notions . . . do not include the idea of a simultaneous operation over different time-scales; and above all they do not include the idea of one processor altering the operation of another. . . . When therefore a [production-system] program such as those of Newell and Simon is operating so as to produce hierarchically organized behavior, this does not mean that there is a hierarchy of processes, like the organization chart of the Civil Service. It only means that there is a hierarchy of rules in long-term memory, much

as books in a library are divided into large sections. . . . *The dominant feature is that one process alters the nature of another process, rather than merely supplying it with input . . . the upper level is concerned with modifiability.* . . . To revert to the concepts of Newell and Simon, we do not merely need the processor to manipulate the outside world and its own short-term memory, under the control of various productions; we also need rules in long-term memory for the writing or deletion of rules in long-term memory. (pp. 185–200, italics added)

Thus control processes are not conceptualized as being limited to the starting and stopping of object-level processes, although this is one important function of control processes (e.g., see Logan & Cowan, 1984). Control processes can also modify the object-level processes, e.g., new rehearsal strategies (cf. learning to learn, or, in more computer-oriented jargon, "Higher organisms do not appear to have fixed software — they can implement new programs to meet unexpected contingencies"; Johnson-Laird, 1983, p. 503).

Besides exploring the role of control processes in modifying people's rehearsal strategies, research on metacognition also explores the role of control processes in other aspects of memory performance (e.g., search strategies, the allocation of study time to various items, search termination — see the Framework section below).

Monitoring Processes in Metamemory

For the control processes to regulate the system effectively, information is needed about the current state of the system. The monitoring processes in human memory were initially referred to by Hart (1967) as the "memory monitoring system."

The person's reported monitoring may, on the one hand, miss some aspects of the input and may, on the other hand, add other aspects that are not actually present (cf. Nisbett & Wilson, 1977). Although the accuracy of reported monitoring may vary across different situations, we expect that the reported monitoring seldom gives a veridical (i.e., nothing missing and nothing added) account of the input. This is not unlike one of the traditional views of perception, where what is perceived is different from what is sensed (i.e., perception conceptualized as sensation plus inference), except that

what is analogous to the objects being sensed here is the object-level memory components.

A distinction should be drawn between retrospective monitoring (e.g., a confidence judgment about a *previous* recall response) and prospective monitoring (e.g., a judgment about *future* responding). Prospective monitoring is further subdivided by Nelson and Narens (1990, p. 130) into three categories in terms of the state of the to-be-monitored items:

1. *Ease-of-learning (EOL) judgments* occur *in advance of acquisition,* are largely inferential, and pertain to items that have not yet been learned. These judgments are predictions about what will be easy/difficult to learn, either in terms of which items will be easiest (Underwood, 1966) or in terms of which strategies will make learning easiest (Seamon & Virostek, 1978).

2. *Judgments of learning (JOL)* occur *during or after acquisition* and are predictions about future test performance on *currently recallable* items [but see below].

3. *Feeling-of-knowing (FOK) judgments* occur *during or after acquisition* (e.g., during a retention session) *and* are judgments about whether a given *currently nonrecallable* item is known and/or will be remembered on a subsequent retention test. [Empirical investigations of the accuracy of FOK judgments usually have the subsequent retention test be a recognition test (e.g., Hart, 1965), although several other kinds of retention tests have been used (for reviews, see Nelson, Gerler, & Narens, 1984; Nelson, 1988).]

Perhaps surprisingly, EOL, JOL, and FOK are not themselves highly correlated (Leonesio & Nelson, 1988). Therefore, these three kinds of judgments may be monitoring somewhat different aspects of memory, and whatever structure underlies these monitoring judgments is likely to be multidimensional (speculations about several possible dimensions occur in Krinsky & Nelson, 1985, and Nelson et al., 1984).

We now believe, in contrast to the above, that JOL should be defined as follows:

Judgments of learning (JOL) occur *during or soon after acquisition* and are predictions about future test performance on recently studied items. These recently studied items may be items for which there has not been a recall test or for which a recall test occurred (irrespective of the correctness/incorrectness of answer).

This newer formulation of JOL, although in some cases yielding overlap with the above formulation of FOK, appears to be more

useful (e.g., see Dunlosky & Nelson, 1992; Nelson & Dunlosky, 1991) than the earlier formulation.

There are at least two important questions about a person's reported monitoring. The first question is, *What factors affect the person's judgments* (e.g., what factors increase the degree to which people feel that they will recognize a nonrecalled answer)? For instance, Krinsky and Nelson (1985) found that people report having a greater FOK for items to which they were informed that they had made an incorrect recall response (i.e., commission-error items) than for items to which they had omitted making a recall response (i.e., omission-error items). This question pertains to the basis for the judgment and is not concerned with the accuracy of that judgment (e.g., people may or may not be correct in predicting better subsequent recognition on commission-error items than on omission-error items).

The second question is, *What factors affect the accuracy of the person's judgments* (e.g., when are FOK judgments most accurate)? For instance, FOK accuracy for predicting subsequent recognition of nonrecalled answers is greater for items that previously had been overlearned than for items that previously had been learned to a criterion of only one correct recall (Nelson, Leonesio, Shimamura, Landwehr, & Narens, 1982). Also, the aforementioned variable of type of recall error (commission versus omission) tends to reduce FOK accuracy because subsequent recognition is usually equivalent on those two types of items.

Subjective Reports as a Methodological Tool for Investigating Monitoring and Control Processes

Long ago, William James (1890) emphasized the use of (nonanalytic) introspection:

> *Introspective Observation is what we have to rely on first and foremost and always. . . . I regard this belief as the most fundamental of all the postulates of Psychology.* (p. 185, his italics)

Around the same time, the structuralist psychologists used a form of subjective reports in which trained introspectors (who participated in approximately 10,000 trials before being allowed to contribute data) attempted to discover the elements of the generalized normal human mind. However, because that form of subjective reports

yielded too many unstable empirical generalizations, turn-of-the-century psychologists rejected it (e.g., Watson, 1913). Moreover, the structuralists "had no theory of cognitive development, . . . there was no theory of unconscious processes, . . . there was no serious theory of behavior. Even perception and memory were interpreted in ways that made little contact with everyday experience" (Neisser, 1976, p. 3).

Subjective reports have reemerged in a form that avoids the problems in the version of analytic introspection used by the structuralists. In his state-of-the-field chapter, Estes (1975) concluded,

> Only in the very last few years have we seen a major release from inhibition and the appearance in the experimental literature on a large scale of studies reporting the introspections of subjects undergoing memory searches, manipulations of images, and the like. This disinhibition appears to be a consequence of a combination of factors. Among these are new developments in methodology. (p. 5)

In the new approach, people are construed as *imperfect* measuring devices of their own internal processes:

> This distinction in our use of subjective reports is critical and can be highlighted by noticing an analogy between the use of introspection and the use of a telescope. One use of a telescope (e.g., by early astronomers and analogous to the early use of introspection) is to assume that it yields a perfectly valid view of whatever is being observed. However, another use (e.g., by someone in the field of optics who studies telescopes) is to examine a telescope in an attempt to characterize both its distortions and its valid output. Analogously, introspection can be examined as a type of behavior so as to characterize both its correlations with some objective behavior (e.g., likelihood of being correct on a test) and its systematic deviations — i.e., its distortions. (Nelson & Narens, 1990, p. 128)

As the methodological foundation evolves for determining when the tool (either the telescope or introspection) is or is not accurate, the content-area researchers (either astronomers or investigators of human memory) can use that methodological foundation to improve the accuracy of their conclusions, using the tool where it is accurate and/or adjusting their conclusions to correct for known distortions. For instance, in terms of theoretical formulations, Ericsson and Simon (1980) regard subjective reports as more accurate for short-term memory than for long-term memory (but see the *delayed-JOL*

effect in Nelson & Dunlosky, 1991). In terms of methodology, improvements have been made in the accuracy of conclusions drawn from subjective reports about the FOK, both in terms of new techniques of data collection (for rationale, see Nelson & Narens, 1980, 1990) and in terms of better ways of analyzing FOK data (Nelson, 1984).

Thus the new approach to using subjective reports both recognizes and avoids the potential shortcomings of introspection (e.g., Nisbett & Wilson, 1977) while capitalizing on its strengths (e.g., Ericsson & Simon, 1980, 1984). This view, which is fundamentally different from the ones used at the turn of the century, opens the way for several broad questions that are empirically tractable and that are important both for theory and for practical applications: Can we develop an adequate characterization of introspective distortions? Can anything be done to reduce those distortions (e.g., see Koriat, Lichtenstein, & Fischhoff, 1980)? Can we characterize the way in which introspections — even with their distortions — are used by the person to affect other aspects of the system?

With regard to the last question, even if a person's behavior (e.g., subsequent recognition of nonrecalled items) is predicted no more accurately by the person's own subjective reports than by predictions derived from other people's performance (Nelson et al., 1986a), this does not reduce the importance of our studying the person's subjective reports as related to his or her own control processes (e.g., Nelson & Leonesio, 1988). As long as the person's subjective reports are *reliable* (and the evidence indicates that they are — Nelson et al., 1986a; Butterfield, Nelson, & Peck, 1988), then something is being tapped, and it may be a subsystem that interacts in important ways with other aspects of the system.

Furthermore, monitoring that is less than perfectly accurate is still useful to the individual as an approximation, as pointed out by Fodor (1983): "The world often isn't the way it looks to be or the way that people say it is. But, equally of course, input systems don't have to deliver apodictic truths in order to deliver quite useful information" (p. 46). Although previous writers such as Nisbett and Wilson (1977) have highlighted the possibility of distortions in introspective monitoring, they have not emphasized its potential role in affecting control processes. A system that monitors itself can use its own

introspections as input to alter the system's behavior. One of our primary assumptions is that in spite of its imperfect validity and in spite of its being regarded by some researchers as only an isolated topic of curiosity, introspective monitoring is an important component of the overall memory system, because most memory activities are self-directed on the basis of introspectively obtained information.

Researchers attempting to understand that system can tap the person's introspections so as to have some idea about the input that the person is using. The present chapter attempts to shift the spotlight of researchers' attention toward self-directed memory and attendant processes such as introspection. This should help correct the "drunkard's search" that began when Watson (1913) rightly emphasized investigations of behavior but wrongly asserted that introspection had no critical role to play in those investigations. As Neisser (1976) remarked, "The realistic study of memory is much harder than the work we have been accustomed to . . . the legendary drunk who kept looking for his money under the streetlamp although he had dropped it ten yards away in the dark" (p. 17).

Our Own Approach to Research Metamemory

In our own research on learning and memory, we have striven to avoid the above mentioned three shortcomings and have focused on metacognition.

Framework

Consistent with the idea that "the two great functions of theory are (1) to serve as a tool whereby direct empirical investigation is facilitated and guided, and (2) to organize and order empirical knowledge" (Marx, 1963), we developed a framework that integrated a wide variety of previously isolated findings and that highlighted empirically tractable questions about metamemory for future research to explore.

The master view of our framework is shown in figure 1.2. The three major stages of acquisition, retention, and retrieval (cf. Melton, 1963), along with several substages, are listed between the two horizontal lines. Monitoring processes are listed above the time line, and

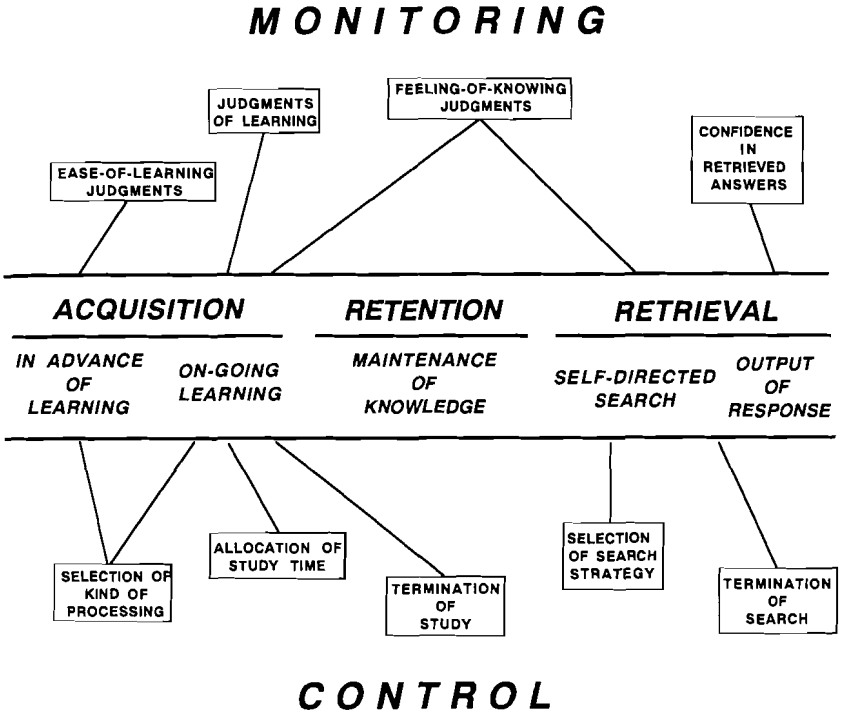

Figure 1.2
Master view of the Nelson-Narens (1990) framework. Memory stages (shown inside the horizontal bars) and some examples of monitoring components (shown above the horizontal bars) and control components (shown below the horizontal bars). Adapted from Nelson and Narens (1990).

control processes are listed below the time line. Figure 1.2 brings those constructs together via a morphological approach (Cummins, 1983). (Note: *Morphological* theories are theories that give a specification of structure — e.g., an explanation of how a cup holds water — in contrast to *systematic* theories, which additionally include the idea of organized interaction; other aspects of our framework not shown in figure 1.2 are systematic, e.g., figures 4 and 5 in Nelson & Narens, 1990.) The details of our framework will not be discussed here, but expositions of it are given in Nelson and Narens (1990; reprinted in part in Nelson, 1992).

Target of the Research

Although the Nelson-Narens framework described above and other theoretical frameworks may try to explain the same sets of data, they do so by emphasizing different aspects of human memory. Many of the important phenomena for metamemory are important to other frameworks only in that they must be neutralized ("short-circuited") experimentally; and many of the important phenomena for the other frameworks are inconsequential for the Nelson-Narens framework because they are not relevant to natural settings and/or do not bear on the concepts used by our framework. This leads to a version of the metaphorical idea of "throwing the baby out with the bathwater" in which one framework's baby is the other framework's bathwater. Because we utilize naturalistic targets, our preference is to let the naturalistic target determine which is the baby and which is the bathwater. We find this approach — letting a target (naturalistic or otherwise) determine basic memory concepts and issues — to be preferable to relying on theorists' guesses about the fundamental mechanisms underlying human memory, because the history of learning/memory research has shown that the overall confidence of one's peers about such guesses reliably fades, and often quickly.

Renewed Emphasis on Learning

We know too little about people's *mastery* of new information during multitrial learning (compared with the kind of memory that remains after a single study trial). In many naturalistic situations, the person's goal is to master a new body of information, e.g., a list of foreign-language vocabulary or new text material. Metacognitive mechanisms can facilitate that goal. The delayed-JOL effect (Nelson & Dunlosky, 1991) illustrates how the accuracy of monitoring one's own learning of new items can be greatly improved. A promising next step is to use the improved monitoring to facilitate mastery through more effective metacognitive control, for example, using delayed JOLs to guide the allocation of study time (Graf & Payne, 1992; Nelson, Dunlosky, & Narens, 1992). Human learning is itself an important topic that has received renewed emphasis recently from the interest

in PDP models and that seems to us to contain an especially rich set of metacognitive components (Vesonder & Voss, 1985). Those components include people's goals, models of how to achieve those goals, and metacognitive monitoring/control mechanisms to be used for that achievement. Although PDP (and other computer-simulation) models focus on object-level memory processes, there is nothing to prevent those models from being conjoined with metacognitive processes. Moreover, the latter may help to solve a formidable shortcoming of computer-simulation models of cognition that has been pointed out by Searle (1992, pp. 212–214 and his summary point no. 7 on p. 226).

Looking Ahead

We envision the end goal of metamemory research to be a system of metamemory that contains a refined account of both how self-directed human memory works and how it can work better. At present there is only a framework of that system, a growing body of experimental findings, and the beginning of a theoretical interplay between models of learning/retrieval and framework mechanisms. Nevertheless, this initial effort is overdue. Almost two decades ago, Skinner (1974) wrote: "there is therefore a useful connection between feelings and behavior. It would be foolish to rule out the knowledge a person has of his current condition or the uses to which it may be put" (p. 209). We believe that the continuation of research on metamemory will result in a scientific understanding of how metacognitive monitoring and control mechanisms are acquired and how they can be employed in naturalistic settings, although perhaps not via explanatory concepts that Skinner would have advocated.[9]

Acknowledgments

This research was partially supported by NIMH grants R01MH32205 and K05MH1075 to the first author. We thank Nancy Alvarado, Harry Bahrick, and William Talbott for comments and suggestions.

Notes

1. By comparison, during the decade 1941–1950 the theoretical framework developed by Clark Hull was cited by 40% of all articles in the *Journal of Experimental Psychology* and the *Journal of Comparative and Physiological Psychology* (and by 70% of the articles on the topic of learning in those two journals).

2. Cumulative progress occurs within a given pocket of research, but the point is that there has been a notable lack of cumulative progress across pockets of research.

3. Whereas *zeitgeist* refers to the spirit of the times, *fachgeist*, which refers to the spirit of the field, seems more appropriate for describing the trends of research in psychology (e.g., see the quotation from Boneau in the text below).

4. For human learning and memory, other acceptable targets could include ones that are non-naturalistic and/or have a large biological/neuropsychological emphasis.

5. But not necessarily all of the products that those processes can produce, which may be one reason that researchers produce computerized learning algorithms.

6. Some aspects of lower level processing may be cognitively impenetrable, not unlike a computer program in which one subroutine may receive input from another subroutine without any direct connection to the internal aspects of that subroutine; other aspects may be monitored by the meta-level.

7. There is no infinite regress here anymore than in, say, the legal system, where, for instance, a trial court can be construed as an object-level court, and an appellate court can be construed as the meta-level court; moreover, the appellate court may be the object-level court for a meta-level decision by a still higher, supreme court.

8. Several points should be made about this analysis. First, "higher" here has no meaning other than as defined above in terms of control and monitoring, similar to an organization such as a business, the military, or a university where the person who is said to be "higher up" is the one who controls and monitors someone else, who in turn may be higher up than yet another person, and so on. Second, it is also possible to have two levels (e.g., L_{h_1} and L_{h_2}) in which neither of them controls or monitors the other; therefore, the aforementioned dominance relation does not apply, and neither of them is a meta-level for the other (e.g., in a university, the chairman of physics versus the chairman of psychology; in memory, rehearsal versus retrieval). In mathematical terminology, the ordering of all the components is transitive (i.e., if P dominates Q, and if Q dominates R, then P dominates R) but need not be connected (i.e., may have distinct components J and K, such that neither J dominates K nor K dominates J). Third, the system described in the text is only one simple instantiation of the

multilevel hierarchical idea; more complex versions are possible (e.g., two sequences designated L_{h_1}, L_{i_1}, and L_{h_2}, L_{i_2} for which L_j is a meta-level for L_{h_1}, L_{i_1} but not for L_{h_2}, L_{i_2}, and so on). Lefebvre (1977, 1992) has used similar ideas about multilevels of meta- and object-level processing in social cognition.

9. Skinner (1974, p. 220 f.) proposed a "consciousness$_2$" that allows people to be self-reflective, but he construed it as only a response (cf. monitoring) and did not allow it to have any causal role in affecting (cf. controlling) external behavior. We believe that such a causal role of metacognitive processing is important for a sound and coherent conception of cognition.