Cognitive Science

An Introduction to Mind and Brain

Daniel Kolak, William Hirstein, Peter Mandik,
Jonathan Waskan

Routledge
Taylor & Francis Group
NEW YORK AND LONDON

© 2006

makes its own unique contribution to the overall goal of understanding of a particular cognitive process.

In this chapter, we'll explore the science of thought by examining, one by one, the findings of each of these interdisciplinary branches of research. As thought and knowledge are intimately related, we'll close out the chapter with an overview of past and current philosophical wisdom concerning what knowledge is and how one attains it – that is, we'll take a brisk jaunt through the field of epistemology.

4.2 THE SCIENCE OF MEMORY

Many of the branches of research alluded to above fall under the broad umbrella of memory research. Memory has traditionally been broken down into two main types, declarative and procedural.

To get a sense for the nature of this distinction, think about what you did last night. In so doing, you are retrieving, and somehow reawakening past experiences. It would seem that in order for these experiences to be reawakened, the relevant information will have had to be stored somewhere until something causes it to be brought on line, or acted out once again in the theatre of conscious awareness. Something similar seems to be taking place when you recall information of a less personal nature. For instance, you can recall the name of the famous individual defeated at the battle of Waterloo. Indeed, in order to understand the previous sentence, you need to recall other kinds of information related to word meanings – like the meanings of famous, individual, Waterloo, and so on. Memory for the kinds of information just described (i.e., facts of a personal or non-personal nature) is known as *declarative memory*. This kind of memory seems to play a central role in your thought processes as you figure out how to build a table. You rely on declarative memory, for instance, in order to remember where you keep your tools, how your father used to work with wood, or the proper procedure for buying things (e.g., lumber and tools).

There is another kind of memory as well. This kind of memory is implicated in your ability to hammer a nail, drive your car to the lumber yard, and perhaps even multiply and divide. These things depend less on *knowing that* as they do on *knowing how*. Philosopher Gilbert Ryle (1949) can be credited with drawing our attention to this important distinction and for realizing that knowing how is as essential for intelligent behavior as is knowing that (though he would surely balk at the non-behavioristic bent of cognitive science). Nowadays this distinction is widely viewed as being more than an intuitive or introspective distinction. Rather, to this intuitive distinction corresponds a real joint in the cognitive system – that is, there really are two different kinds of memory, called declarative and *procedural*, respectively. As we shall see, the latter sometimes takes over for the former, and so the dividing line between the two can become blurred (see box on memory below).

To list further characteristics of each form of memory, declarative memory is usually taken to be attention-demanding, and the kind of information handled by declarative memory systems is easy to verbalize. Procedural memory has the opposite properties. It seldom requires careful attention, and procedural knowledge is often difficult to put into words. Consider, for example, how difficult it is to describe how to keep a bicycle upright.

ENDEL TULVING (1995) ON MEMORY

Memory is many things, even if not everything that has been labeled memory corresponds to what cognitive neuroscientists think of as memory. Memory is a gift of nature, the ability of living organisms to retain and to utilize acquired information or knowledge. The term is closely related to *learning*, in that memory in biological systems always entails learning (the acquisition of information) and in that learning implies retention (memory) of such information.

Memory is a trick that evolution has invented to allow its creatures to compress physical time. Owners of biological memory systems are capable of behaving more appropriately at a later time because of their experiences at an earlier time, a feat not possible for organisms without memory.

Memory is a biological abstraction. There is no place in the brain that one could point at and say, Here is memory. There is no single activity, or class of activities, of that organism that could be identified with the concept that the term denotes. There is no known molecular change that corresponds to memory, no known cellular activity that represents memory, no behavioral response of a living organism that is memory. Yet the term *memory* encompasses all these changes and activities.

4.2.1 Declarative memory

Your ability to remember facts seems to involve the three component processes of encoding, storage, and retrieval. To see why, notice that in order for the information to be *re*awakened, it must have been awake to begin with. The process by which this takes place is known as *encoding*. Subsequent to encoding, the relevant information must be *stored* on either a short-term or long-term basis. If it were not stored, there would be no way in which it could later be *retrieved*, or reawakened. Contemporary research into declarative memory boils down, in large part, to the study of encoding, storage, and retrieval processes.

As suggested in Chapter 2, the length of the storage period will vary depending upon which of the two main declarative memory stores is being utilized. That is to say, there are two forms of declarative memory, short term and long term. This distinction is supported by a variety of findings, several of which have to do with the well-known serial position curve.

4.2.1.1 Short-term memory

If you will recall, short-term memory is the kind of memory implicated in your ability to remember an unfamiliar phone number in the time between looking it up and dialing. There are actually at least two different kinds of short-term memory. When you are remembering a phone number, you are probably relying on a mechanism known as the *phonological loop*. This is a short-term memory system that stores linguistically encoded information for short periods of time (often on the order of seconds). We also have a short-term memory system, the *visuospatial sketchpad*, that holds information about visuospatial relationships for short periods of time. These are usually viewed as the two most basic forms of short-term memory though, as we shall see, there are reasons for thinking that there are others.

4.2.1.1.1 The phonological loop subsystem

A variety of experimental manipulations, usually involving some variation on the error score methodology, have helped to shed light on precisely how the phonological loop subsystem operates. To start with, the kind of information represented by the phonological store concerns either motor commands involved in the production of speech sounds or the acoustic properties of speech sounds (the jury is still out on the matter). At any rate, the encoding of linguistic stimuli is effected on the basis of phonological (sound-based) rather than orthographic (vision-based) properties. One way this has been revealed is by testing how well certain items are stored. It seems that immediate recall of linguistic stimuli (e.g., items from a list) is impaired when, to quote Alan Baddeley (1990, p. 20), "items are similar in sound or articulatory characteristics." This is known as the *phonological similarity effect*. A related finding is that unattended speech stimuli also disrupt the immediate recall of linguistic stimuli. (Some of you may recall teasing a friend or sibling by yelling irrelevant words at them as they try to remember a set of words or numbers.) The magnitude of this *unattended speech effect* seems to be unrelated to the semantic properties of the stimuli (i.e., the meaning of the words) since both nonsense syllables and meaningful stimuli prove to be equally disruptive. The magnitude of the effect does, however, vary with phonological similarity.

Manipulations that affect the shape of the serial position curve have also shed light on the nature of the phonological loop subsystem. As was noted in Chapter 2, when a distracting activity like counting backwards is interposed between the time that the last list item is presented and the start of recall, many items are lost from memory. Indeed, the more time that elapses, the more items are lost (Glanzer and Cunitz 1966). It is thought that the distractor activity prevents the rehearsal of list items (you can get a good sense for what is meant by rehearsal if you try to remember an unfamiliar phone number for about a minute). In other words, a kind of replay loop keeps items from disappearing from short-term phonological storage – hence the name, "phonological loop."

The reason why items are lost from storage is not entirely clear, though there is some evidence to suggest that, rather than simply fading, items are overwritten by other incoming stimuli. One reason for thinking this is that the recency component of the serial position curve seems not to significantly diminish in size when the distraction task involves the repetition of a single, simple stimulus (see Longoni et al. 1993). One reasonable explanation for this finding is that repetition may prevent rehearsal but, given the dissimilarity between the repeated item and most list items, repetition fails to cause stored list items to be overwritten. Thus, instead of serving the function of refreshing fading memory traces, rehearsal may instead prevent other linguistic stimuli (perhaps even the sound of one's thoughts) from overwriting list items. It is also worth noting that a certain kind of Kohonen map (discussed in Chapter 2) exhibits some of these very same properties (Miikkulainen 1993).

There is no requirement that the stimuli encoded in short-term memory be meaningful (as Ebbinghaus [1885] discovered, the same can be said of long-term memory). When stimuli *are* meaningful, however, later recall will be facilitated. This may be due to the influence of long-term memory on short-term memory rather than reflecting a peculiarity of short-term memory. In other words, the best available evidence indicated that linguistic information is encoded in the short term on the basis of phonological or articulatory characteristics.

The system responsible for the recognition of speech sounds may, however, be distinct from the system responsible for short-term phonological storage. A number of patients have been identified who fail to exhibit the standard phonological similarity or recency effects (Martin and Breedin 1992). These patients do, however, show normal performance on phoneme discrimination tasks. Thus, these patients evidence a dissociation of short-term phonological memory and language recognition. In addition, patients have been identified who have impaired phoneme discrimination but normal recency and phonological similarity effects. This *double dissociation* of phoneme identification and phonological short-term memory presents a compelling case that there are two distinct mechanisms involved in carrying out these functions. Functional imaging research corroborates this proposal, as do studies of impairments to phoneme identification ability and immediate recall following direct cortical stimulation (see Chapter 2 for an overview of each of these techniques). Not surprisingly, phonological storage mechanisms have been localized to an area of the temporal lobe that overlaps with an area that has long been thought to be implicated in auditory processing. As you can see, as far as interdisciplinary collaboration goes, the study of the phonological loop subsystem is one of the success stories.

4.2.1.1.2 Short-term visuospatial memory

In Chapter 2, if you will recall, the use of reaction times in psychological investigation was illustrated by considering findings about the nature of mental imagery. Mental imagery is taken by many to depend upon short-term memory mechanisms that are distinct, in certain ways, from those comprising the phonological loop. The main component of short-term visuospatial memory is what Baddeley (1990) calls the *visuospatial sketchpad*.

One of the interesting findings about the visuospatial sketchpad is that when subjects are issued spoken instructions concerning a how to carry out a visuospatial reasoning task, they perform better than when they are issued written instructions. This is taken to indicate that some of the same mechanisms involved in the visual processing of linguistic information are tapped during mental imagery. This finding has been corroborated by functional neuroimaging research which suggests that the sketchpad might be localized to the occipital lobe (the known locus of early visual processing) and areas that are contiguous and anterior (Kosslyn 1994).

Also of interest are findings concerning the role of the frontal lobes in visuospatial short-term memory. Using what is known as a *delayed-response task*, Patricia Goldman-Rakic (1992) found that a certain portion of the prefrontal cortex may be implicated in keeping information on line – that is, active and available for use in behavior guidance. In a typical delayed-response task, a monkey is shown a small object at a certain position in its visual field. The object is then taken away, and after a delay of several seconds, the monkey must indicate where in its visual field the object appeared. This delay is important, because it means that the monkey's response must be guided by an internal representation of the object and its location, as opposed to behaviors which are based on objects present at the time. Lesion data supports Goldman-Rakic's hypothesis about the neural substrate of working memory: when the part of the prefrontal cortex thought to contain this memory system is temporarily deactivated by cooling it, the monkeys are unable to respond correctly.

While clearly relevant to the study of short-term memory, it may be that the role of this area of the frontal cortex is not specific to visuospatial memory. Rather, this region may comprise an important component of an executive network that regulates the activity of slave systems - including the phonological store, the visuospatial sketchpad, and other systems implicated in on-line processing.

4.2.1.2 Long-term declarative memory

On one view of the relationship between short-term and long-term memory (the origins of which can be traced back to Ebbinghaus), information is first entered into short-term storage and, following a period of repetition, the information is transferred into long-term storage. While this proposal seems to provide a reasonable account of many empirical findings (recall the discussion of the serial position curve in Chapter 2), it fails to do justice to many of our pre-theoretical intuitions about how long-term memory systems operate. Imagine being told, for instance, that George W. Bush has converted to Scientology. You would probably be able to retrieve this bit of trivia at some later date, and without ever having taken the time to repeat it to yourself. In fact, there aren't very many cases when repetition is required in order for information to be entered into long-term memory. Instead, the transfer of information into long-term memory seems, in many cases, to be carried out quite automatically (i.e., in a way that requires little or no attention or effort on our part). Cognitive psychologists have devised many ingenious experiments that corroborate this proposal and further refine our understanding of the interplay between long-term encoding, storage, and retrieval processes.

One important set of findings about long-term storage processes suggests that the manner in which information is initially encoded has a very strong influence on whether, and under what conditions, it will later be retrieved. This line of research began with a set of experiments carried out by Hyde and Jenkins (1973), who employed the traditional method of measuring recall for list items - in this case words. They divided their subjects into two groups, only one of which was informed that it would later be tested. Given that the former group expected to be tested, one would expect that they would put more effort into remembering list items. They might, for instance, repeat the words to themselves. Each of these groups was further subdivided (in the very same way) such that one subgroup was asked to perform a task that would require them to pay attention to the outward form of each word (namely, deciding whether or not the word contained a particular letter), while the other subgroup was asked to perform a task that would require them to pay attention to the meaning of each word (namely, rating the pleasantness of the word's referent). As it turns out, having prior knowledge that a test would follow list presentation had no significant effect on subsequent recall. Encoding the meaning rather than the form of the stimulus turned out, on the other hand, to have a profound (and statistically significant) facilitating affect. You might bear this in mind the next time you find yourself studying for an exam. That is, rather than simply repeating facts (e.g., "procedural memories are automatic and difficult to describe") over and over again, you might try instead to concentrate on what the various sentences mean.

The manner in which memory is affected by the depth to which linguistic information is processed (where shallow processing is that concerned with outward form and deeper processing concerns meaning) has since been studied quite extensively. Baddeley (1990), for one, thinks that the best explanation for the above effect might be related to the explanation for why information is lost from short-term memory. When material is processed only in terms of its overt linguistic form, it will have a good chance of being overwritten by stimuli that are overtly similar. This is because there is only so much variation in the realm of words. On the other hand, there is tremendous room for variability in the semantic realm. Thus, phonological or orthographic similarity effects will be much more common than semantic similarity effects - though the latter have been found to occur quite frequently under the right conditions.

One of the most important findings of neuropsychology has been that the seahorse-shaped *hippocampus* (which means "sea monster" in Greek), a structure buried beneath the surface of the temporal lobe, is a mechanism that plays an essential role in the long-term storage of information. Individuals who suffer damage to this area (for reasons ranging from accidental head trauma to Alzheimer's disease) can be stricken by a severe form of amnesia. This disorder has two components to it. On the one hand, these individuals can lose their ability to store new information on a long-term basis (this is the *anterograde* component of amnesia). Accordingly, their window onto the immediate past may extend only as far back as their short-term memory systems will permit - usually just a few moments. (It is worth taking a moment to imagine what that would be like.) On the other hand, these individuals may also lose memory for the events that preceded the time of injury. This *retrograde* component of amnesia can extend as far back as three years. The extent of retrograde amnesia can be determined by asking subjects to answer questions about public or personal events for which dates can be assigned (e.g., by questioning friends or consulting newspapers).

At present, the most popular way of accounting for these findings is to view the hippocampus as a device for organizing long-term memories and, so to speak, filing them away. Where they might be filed *to* is indicated by the extensive anatomical connections (both incoming and outgoing) between the hippocampus and the sensory areas of the cortex. In light of this connectivity, some have speculated that particular memories are ultimately stored (or *consolidated*) in or near the very areas of cortex that were active during their initial encoding. According to this model, the role of the hippocampus is to store memories for as long as it takes (i.e., up to three years) to effect this consolidation process. Thus, any memories that have not yet been filed away at the time of damage to the hippocampus will forever be lost. While quite compelling as an explanatory model, the details of the consolidation process themselves stand in need of explanation. Some have begun to speculate that this process takes place when we are asleep (and is, in fact, the very point of sleep). In addition, connectionist systems have been created that may give us a glimpse into the details of the consolidation process.

4.2.1.2.1 Episodic and semantic memory

There are those who feel that long-term memory mechanisms might be further subdivided on the basis of the kind of information being stored. The distinction at issue was alluded to above when it was pointed out that you can remember facts of a personal nature (e.g., what you did last night) as well those of a non-personal nature (e.g., that Napoleon was defeated at Waterloo, or the meaning of "famous"). Memory for the former sort of information is called episodic memory while memory for the latter is called semantic memory. The driving question for those interested in this distinction is whether or not the two kinds of information at issue are handled by two distinct long-term memory mechanisms.

Some early support for the claim that there are distinct memory mechanisms was offered by informal (or *anecdotal*) evidence concerning individuals with hippocampal damage. Such individuals often have intact memory for non-personal facts and word meanings, but their memory of personal events is severely impaired. It is unclear what should be concluded on this basis, however, because it is very difficult to assess precisely when non-personal information was learned. Try to remember, for instance, when you first learned of Napoleon's defeat at Waterloo. Some recent studies suggest that the hippocampus plays a critical role in both episodic and semantic memory.

In one study, a comparison was made among amnesiacs between their memory for personal events and their memory for events of a more public nature, such as the Iran-Contra hearings (MacKinnon and Squire 1989). For each subject studied, it was found that the onset of retrograde amnesia was the same whether the events at issue were public or private. It is plausible, then, that memory for public events (namely, those that occur during one's lifetime) should be categorized under the heading of *episodic* rather than semantic memory.

There is at least one case study that may shed light on the matter of whether or not distinct mechanisms underwrite episodic and semantic memory and, if so, where, precisely, the boundary between the two should be drawn. De Renzi et al. (1987) describe an Italian woman (known to the scientific community as L.P.) who suffered encephalitis. L.P. was found to have severe semantic memory deficits without any impairment to episodic memory. Specifically, she had no memory for most of the facts that she had learned at school – facts about geography and history for instance (e.g., she had no knowledge of Hitler or of any facts at all concerning World War II). Her ability to remember personal events, on the other hand, was completely intact. In addition to supporting a dissociation between the two forms of memory, what is interesting about L.P.'s case is that her memory for public events that transpired during her lifetime was found to be restricted to facts that had a very *high* personal significance. For instance, though L.P. had no recollection of either the location or nature of the Chernobyl disaster, she did remember that it had caused her plants to suffer. Thus, insofar as there is a distinction between episodic and semantic memory mechanisms, L.P.'s case may indicate that episodic memory handles information only about the most personal aspects of public events while semantic memory mechanisms handle the rest. A single dissociation in one case study is, however, far from conclusive.

4.2.1.2.2 The cellular bases of information storage

Experience changes us, but how exactly does this happen? We are so accustomed to the fact that doing something slowly makes us better at it, or the fact that reading a book leaves us with memories of its contents, that we fail to notice the work which the brain does in making this miracle possible. Donald Hebb is generally credited with being the first to propose the cellular basis by which the brain is changed by experience. What has come to be known as Hebb's rule (discussed briefly in Chapter 1) is still a vital principle in the understanding and design of cognitive systems. Hebb's rule reads as follows:

> When an axon of cell A is near enough to excite cell B and repeatedly or persistently takes part in firing it, some growth process or metabolic change takes place in one or both cells such that A's efficiency, as one of the cells firing B, is increased.
>
> (Hebb 1949, p. 62)

In short, whenever two adjacent neurons tend to fire in synchrony, the connection between them is strengthened, so that firing of the first neuron makes it more likely that the second neuron will fire. A second quotation from Hebb makes this point more clearly: "Two cells or systems of cells that are repeatedly active at the same time will tend to become 'associated,' so that activity in one facilitates activity in the other" (Hebb 1949, p. 70).

The existence of such a mechanism was verified in the simple nervous system of the mollusk *Aplysia*. A process known as long-term potentiation (LTP) may constitute part of the biological basis for Hebb's rule. Researchers in the 1970s found that briefly applying electrical stimulation to the hippocampus of the rabbit made cells adjacent to the stimulated cells more likely to fire, at least for several hours. Later researchers have been able to verify that electrical stimulation can produce LTP lasting for days or even weeks.

It is worth noting, however, that long-term memories seem to be stored in the hippocampus for up to three years. Thus, unless LTP is shown to operate over such lengthy spans of time, scientists will have to begin searching for other cellular mechanisms that can account for long-term memory. One possibility is that LTP is an effect that lasts just long enough for the hormones associated with an emotional response to circulate through the blood to the site of LTP. When these hormones reach sites where LTP has taken place, they set in motion cellular changes that give rise to truly long-term strengthening of inter-neuron connections. According to this model, not just any paired neural firing is worthy of strengthening, and the measure (or one measure) of significance is the associated affective response.

4.2.1.3 Confabulation about memory

Certain types of brain lesion can produce a curious phenomenon known as *confabulation*. Specifically, when asked a question which touches on deficits caused by the injury, rather than simply acknowledging any problems, patients will give a false or irrelevant answer, as if they were attempting to cover up their deficit. Korsakoff's syndrome is a form of amnesia, most often caused by a lifetime of heavy drinking. The locus of lesion is not as clear in the case of Korsakoff's amnesia as in certain other cases, though the most frequent sites of damage seem to be in the lower, more "primitive" parts of the brain. Korsakoff's amnesia is severe enough that patients will typically have no memory at all of the events of the preceding day. But when asked what they did yesterday, Korsakoff's patients will often produce a detailed description of plausible (or not so plausible) events. These events are either entirely made up at the time of utterance, or traceable to some actual but much older memory.

The patients give no sign that they are aware of what they are doing; apparently they are not lying, and genuinely believe their confabulations. They do not give any outward signs of lying, and their demeanor while confabulating is that of any normal person talking about their actual recent experience.

4.2.1.4 Frames and other large knowledge structures

Up until now, we have mainly been concerned with the mechanisms responsible for storing particular pieces of information. Long-term memory mechanisms do more than encode, store, and retrieve a bunch of disconnected facts, however. To get a sense for the kind of global organization that some researchers have been investigating, consider for a moment what you know about the steps required in order to go from not being enrolled in a university to passing your first course. Your ability to envision the appropriate sequence of events suggests that your knowledge of how universities operate is not encoded in the form of a bunch of disconnected facts. Rather, the mechanisms of long-term memory somehow capture the relationships between the individual facts.

What we have been calling *traditional AI* has, for quite some time, taken the lead in investigating the global organization of knowledge. Marvin Minsky (1985) was an early pioneer in the effort to model large-scale knowledge structures. He modeled knowledge for everyday circumstances in terms of a set of schematic (i.e., detail-poor) knowledge structures centered around particular kinds of situations. If the information relevant to filling in the details of these *frames* is not explicitly provided, certain values are simply assumed by default. For instance, a system might be equipped with a frame associated with how to ride public transportation. It might assume, unless instructed otherwise, that public transportation always requires a fee. A further frame, with its own default assumptions, might be called upon in order to supply information about the kind of token-selling mechanism, and so on. Frames are thus said to be organized in a *recursive* manner (i.e., there can be frames within frames).

The ability to organize bits of knowledge into large-scale structures seems to play a prominent role in text comprehension. A given passage of prose will often leave out a great deal of information, the implicit assumption being that the reader will fill in the relevant details. Should, for instance, the main character in some story buy a hotdog from a street vendor, you might take for granted that the vendor has a cart with heated containers and condiments, that money is exchanged, and so on.

The task of modeling this kind of filling was one of the goals set for Schank and Abelson's (1977) SAM (see Chapter 1). SAM could read passages of text from a newspaper and generate sensible answers to questions about the text by filling in the information that was only implicit. Evidence that this kind of filling in actually does takes place has been gathered on the basis of behavioral research concerning the time it takes to recognize that a given stimulus is a word. There are certain conditions, for instance, under which the time it takes to recognize that an item is a word (as opposed to a meaningless collection of letters) is reduced. One such case is when the word to be recognized concerns the implicit details of a passage of text read prior to the word recognition task. For instance, after reading a sentence about the collapse of a building, it was found that subjects were faster at recognizing "earthquake" as a word than they were with other words (Kintsch 1998). This and other evidence suggests that we quite naturally draw upon our background knowledge of particular kinds of situations in order to fill in details that were never mentioned.

4.2.2 Procedural memory

Procedural memory, or *know-how*, has not always been viewed as a key player in the mental life of the individual. Procedural memory is often thought of as the kind of memory that underwrites particular motor skills, such as riding a bike. To see why it might play a more prominent role in our thought processes, the consideration of basic motor skills is actually a good place to start.

Can you recall your first clumsy attempts at driving a car? When you first sat behind the wheel of a car, it probably took all of your concentration to heed traffic signs and signals, work each of the pedals and the gearshift, and keep the car moving smoothly and in a straight direction. If you were anything like the rest of us, when you first started driving you were probably quite awful. Eventually, however, you got to the point you are now at. Now, when you drive home from work or school you are probably scarcely even aware of what you are doing. Many researchers now believe that you have your cerebellum to thank for this, though a set of nuclei buried underneath the cortex (the *basal ganglia*) also seem to play an important role. One of the interesting anatomical features of the cerebellum is that it is wired up to the rest of the brain in such a way that it has many incoming connections from the sensory areas of the cortex and many outgoing connections to the motor areas. Moreover, a look at the low-level anatomical details of the cerebellum reveals a structure that is configured in a manner remarkably reminiscent of a connectionist perceptron (see Chapter 1). In other words, every indication is that this structure is ideally suited to the task of learning how to pick up on input/output regularities. Many researchers thus view the cerebellum as a system that monitors the cortex for input/output patterns. For example, when the sensory system delivers information about sitting behind a wheel and seeing a red hexagon with the word STOP on it, the motor system responds with a characteristic motion of the right foot. When you first learned to drive, this characteristic motion may have only occurred after the most deliberate and attention-demanding effort. Eventually, however, your cerebellum (and basal ganglia) picked up on this input/output pattern and took over in such a way that you no longer had to think about it. It became automatic.

There are lots of sensory-motor input/output patterns for the cerebellum to pick up on. For instance, if you do a lot of pencil-and-paper arithmetic, you'll find that this process is also highly automated. The same can be said in the case of formal logic. Insofar as the cerebellum is responsible for automating these activities, it should be viewed as a crucial player in our everyday thought processes.

Some support for this proposal comes in the form of connectionist modeling research. Bechtel and Abrahamsen (1991), for instance, designed a pair of networks which learned to accomplish such tasks as assessing the validity of formal logic proofs and filling in missing information. On the basis of this and other data, they propose:

> logic problems, when viewed as pattern recognition tasks, can be solved by networks which, like humans, seem to be capable of learning from errors and tuning their performance . . . The ability to reason using logical principles may not need to be grounded on proposition-like rules, but rather reflect a kind of *knowing how*.
>
> (Bechtel and Abrahamsen 1991, p. 208)

Indeed, it may be that cognitive automation is not restricted to perceptuo-motor patterns. It may be that the patterns associated with everyday, deliberate reasoning may themselves become automated (see Thach et al. 1992). The attempt to determine what the precise role of procedural knowledge is in our everyday thought processes is thus one of the exciting new areas of interdisciplinary investigation.

4.3 REASONING

Assume, for a moment, that the moon is made of green cheese. Assume, furthermore, that all green cheese comes from Wisconsin. On the basis of these assumptions, what can you conclude about the origin of the moon? More importantly, what memory system do you utilize to reach this conclusion? The process of reasoning seems to be an active, on-line affair (see also section 1.2.1 in Chapter 1). Perhaps, as just suggested, the cerebellum plays an important role in this process as well. Of course, long-term declarative memory is also a key player. The study of reasoning is thus intimately related to the study of memory. Indeed, perhaps a full understanding of the memory systems and their interactions will exhaust what there is to know about reasoning. While this is a reasonable long-range plan for cognitive science, researchers have found ways to investigate reasoning in its own right.

Crudely put, reasoning is a process whereby one derives certain information (e.g., facts, knowledge, sentences, etc.) from other information, as in the following:

(1) Lulu is a cow.
(2) All cows chew cud.
(3) Therefore, Lulu chews cud.

There are two broad ways in which one might undertake the study of reasoning. On the one hand, there is the study of how one ought to reason, or the *norms* of reasoning. Philosophers have always been the experts in this area. Because logical argumentation is the philosophers' primary research tool, they have always been both cognizant and critical of the methods by which people are led to conclusions. Alternatively, though, one can investigate how we actually *do* reason under everyday circumstances. One characteristic method for investigating reasoning on an empirical basis is to determine the extent to which humans depart from the norms established by philosophers and other cognitive scientists. One thing (perhaps the only thing) that philosophers generally agree upon is the basic taxonomy of reasoning methods. The central division in this taxonomy is between monotonic and non-monotonic forms of inference.

4.3.1 Monotonic reasoning

There are many who feel that "monotonic inference" and "deduction" are co-extensive (i.e., the two labels pick out all and only the same things). Good (or *valid*) deductive inferences are such that *if* the premises (the reasons given) are true, then the conclusion *must* be true as well. The inference to (3) above has this property.

4.3.1.1 Deduction

Because deductive inferences supply this kind of certainty, and the philosopher lusts after absolute certainty, the study of deductive reasoning has been a mainstay of philosophical investigation. If you will recall from Chapter 1, it was the formalization of the rules for deduction by philosophers that eventually led to the creation of mechanical deduction devices (i.e., electronic computers) and, ultimately, to the advent of the field of artificial intelligence.

Introspectively and intuitively, it seems as though we do engage in deductive reasoning. As much as possible, however, cognitive scientists have aimed to do away with introspection in favor of replicable and objective empirical data. Much effort has accordingly been expended on the task of generating data that will help to clarify whether or not human reasoning actually accords with prescribed norms; if so, what the precise mental operations are by which these reasoning activities are carried out; and if not, what are the hidden causes of these observed shortcomings.

One of the most important research programs concerning the nature of deductive reasoning began in the very early days of cognitive science. In the 1960s, Allen Newell and Herbert Simon began an investigation into the procedures by which humans solve formal reasoning problems. They utilized and helped to pioneer the now-popular technique of protocol analysis – whereby subjects are presented with a task and are asked to verbalize their thought processes as they go about solving that task. Newell and Simon (1972) recorded and analyzed the reports issued by subjects who were asked to solve tasks that seem to have a substantial deductive reasoning component. Their ultimate goal was the creation of a computational system capable of implementing these very procedures. The pay-off was the creation of the production system approach to cognitive modeling described in Chapter 2.

This line of research continues, unabated, to the present day. Lance Rips, for instance, utilizes production systems in order to model the reasoning processes (revealed by protocol analysis) that subjects engage in when solving problems such as this one (Rips 1994):

> Each of A, B, and C is either a knight or a knave.
> Knights always tell the truth and knaves always lie.
> A, B, and C make the following statements:
>
> > A: Both B and C are knaves.
> > B: C is a knave.
> > C: A is a knight.
>
> Question: Is A a knight or a knave?

Not only do subjects seem to draw upon a competence at deduction (e.g., their ability to infer *q* from "If *p* then *q*" and *p*), they also use a number of *metalogical* strategies – that is, strategies for deploying their core deductive capacities. For instance, one strategy for solving problems like the one above is to make a provisional assumption about whether a particular character is a liar or truth-teller. Subjects follow through on these provisional assumptions by examining their implications in terms of the lying or truth-telling status of the other characters.