15 A taxonomy of thinking

P. N. Johnson-Laird

Introduction

Imagine visiting a large group of islands and exploring each of them thoroughly. You learn how to find your way around each, yet you lack a complete understanding of the overall topography of the islands because you have no idea of the relations among them. You do not know how they were shaped into an archipelago. Indeed, you do not even know whether you have visited them all. When I tell you that the names of these islands are "Induction," "Deduction," "Problem solving," and so on, you will realize that I have in mind your predicament as a diligent reader of this book. Each of its chapters is a guide to some domain of thought, but there has so far been no account of the relations among these domains. Hence, you may wonder how many sorts of thinking there are and how they are related to one another.

To answer these questions it is helpful to bear in mind the distinction between the function of thinking and the underlying procedures that it relies on. The late David Marr (1982) referred to this distinction as one between the computational level (what the mind is doing) and the algorithmic level (how it is doing it). Many of the chapters in this book have described thought processes at the algorithmic level. My task in this final chapter is to establish the outlines of a taxonomy that will help you to fit the domains of thought into a single framework – a single map of the mind. To establish this taxonomy, I shall be concerned mainly with what the mind accomplishes at the computational level.

Thinking without a goal

The ability to think is crucial to human life, but it is difficult to study because we are not aware of how we do it. We can observe only its consequences in our conscious thoughts, in our behavior, and in our speech. It also occurs in such dazzling varieties that some cognitive scientists despair of our ever understanding it completely (see, e.g., Fodor, 1983). There is, at one extreme, the free flow of ideas in daydreams. James Joyce re-created this variety of thought in the final pages of his great novel *Ulysses*. Molly Bloom, the wife of the novel's protagonist, lies in bed thinking about her husband:

Yes because he never did a thing like that before as ask to get his breakfast in bed

with a couple of eggs since the *City Arms* hotel when he used to be pretending to be laid up with a sick voice doing his highness to make himself interesting to that old faggot Mrs Riordan that he thought he had a great leg of and she never left us a farthing all for masses for herself and her soul greatest miser ever was actually afraid to lay out 4d for her methylated spirit telling me all her ailments she had too much old chat in her about politics and earthquakes and the end of the world let us have a bit of fun first . . .

Joyce chose not to punctuate Molly Bloom's soliloquy, perhaps to catch its fleeting, inchoate nature, but if you read it aloud, it makes excellent sense. The process generating such a daydream is rapid, involuntary, and, apart from its results, outside conscious awareness. You recall an episode, "she never left us a farthing [in her will]," and the memory triggers a judgment: "greatest miser ever," which, in turn, reminds you of something else: "was actually afraid to lay out 4d for her methylated spirit," and so on and on.

William James (1890), who may well have influenced Joyce's writing, likened the stream of consciousness to the trajectory of a bird – a sequence of alternating flights and perchings. Dreams depend on the same sort of thinking, but their narrative is often compelling (see Gerrig's remarks on narrative structure in Chapter 9). Indeed, more than one short story has been dreamed in its entirety (Brook, 1983). Perhaps the most important feature of this sort of thinking is that it has no goal. It is not directed toward solving any problem or reaching any conclusion.

There is a long tradition in psychology from Aristotle onward of explaining the flight of ideas in terms of associations: One thought triggers another in a chain of linked ideas, which often affect our emotions. A classic study of associations was carried out by the nineteenth-century English scientist Francis Galton (1883). He compiled lists of words, which he put away in a desk drawer and forgot. Later, he went through the list, making free associations to each word; that is, he read a word and responded with the first word that it called to mind. Sometimes his responses shocked him so much that in his report of the study he passed briefly over them, saying only that they revealed the otherwise "hidden plumbing" of the mind. (Molly Bloom was not so easily shocked.) In another classic study of free association, Carl Jung (1919) observed that it takes longer to respond to emotionally laden words than to emotionally neutral words. He was thereby able to unmask a thief from her slow responses to words relating to the details of the crime.

Goalless thinking is evidently important to our emotional lives, but it is also important theoretically because it illustrates a particular mechanism of thought. The traditional theory held that one thought A elicits another thought B because the two have somehow become linked, or associated, in memory. Of course, there may be associated with thought A a number of alternative thoughts, B, C, D, and so on, and the strength of the links may vary. Which thought emerges on any particular occasion, according to this theory, is a matter of chance, though it is biased according to the strength of

the associations. The difficulty with this theory is that it is hard to imagine that all the links in the flight of ideas have been forged already. It may never have occurred to Molly Bloom before that Mrs. Riordan was "the greatest miser ever." Indeed, she may never have previously called to mind that Mrs. Riordan left her nothing in her will. There is always a first time for a memory to be recalled, and there is always a first time for a particular judgment to be made. Conversely, if the flight of ideas always depended on preexisting links, it would never lead to any novel thoughts. The notion of preformed links is perhaps feasible for associations between individual words (although even here it runs into difficulties; see Johnson-Laird, Herrmann, & Chaffin, 1984), but it is not feasible as a general account of the flight of ideas. I will return to the problem of creativity later. For the time being, the point to bear in mind is that thinking in certain circumstances may lack a goal; it may appear to be outside voluntary control and to have no particular destination. It throws up ideas that are related to one another but that, like clouds, have no overall structure.

Calculation and the "problem space"

Perhaps the antithesis of the flight of ideas is the thinking that occurs in mental arithmetic. If I ask you, "What is twenty times thirteen?" you deliberate in an explicit, voluntary, and consciously controlled way. You may say to yourself, for instance, "Two times thirteen is twenty-six" and "Ten times twenty-six is two hundred and sixty." You are not aware of the way you retrieve these arithmetical facts, and you are not aware of the way they are represented in your mind. You just happen to know them and can recall them as you need them. Likewise, your plan for dealing with the problem derives from a knowledge of how to multiply by 10, and it comes to mind almost without thought (see also the discussion of mental arithmetic in Chapters 10 and 14). Although some of the processes occur outside consciousness, you are nevertheless totally aware of the overall plan that you are following. You can choose how to do the calculation (or whether or not to carry it out at all), but once you have chosen a plan, you have no freedom about what to do to obtain the right answer. Your thinking has a single precise starting point, a single precise goal, and it unwinds like clockwork.

There was a time when psychologists believed that all behavior was controlled by external events. Karl Lashley (1951) pointed out that this hypothesis could not explain the rapid execution of skills, such as mental arithmetic, which call for an internal hierarchical organization. George Miller, Eugene Galanter, and Karl Pribram called these organizations "plans," and the publication of their book *Plans and the Structure of Behavior* (1960) sealed the demise of behaviorism. They defined a plan as "any hierarchical process in the organism that can control the order in which a sequence of

operations is to be performed," and they demonstrated convincingly that planning is a major part of thought.

Allen Newell and Herbert Simon (1972), who pioneered the computational analysis of problem solving, provided a unifying framework for planning and thinking, which Lesgold has described in Chapter 7. In all cases of a problem, there is a *starting point* – the initial conditions – and a set of mental *operations* that must be carried out in an appropriate way so as to reach a *goal* – the solution to the problem. There is a "space" of all possible sequences of operations, and what has to be worked out is a sequence, if there is one, that forms a route through the space from initial state to goal:

state 1→ state 2→ . . . →goal

A successful plan generates a route that solves the problem.

Not all thinking depends on a goal, but the bulk of it does, and much of the taxonomy of thought can be based on characteristics of the problem space. Thus, mental arithmetic is deterministic; that is, at each point, the next step in the calculation is determined wholly by its current state. There is only one route through the problem space from one state to the next, and your knowledge enables you to follow it with little difficulty.

Nondeterminism

Perhaps most thinking lies between the two extremes of daydreaming and calculating – between the clouds and the clocks of the mind. Unlike a daydream it has a goal and thus a global structure, and unlike a calculation it does not unwind in a strictly determined way. When you are trying to solve the missionaries and cannibals problem (see Chapter 7), for example, you do not have a "sure-fire" procedure, like a procedure for multiplication. Different people tackle the problem in different ways; you yourself, if you could step backward in time and make another attempt in ignorance of the first, might take a different path. Nothing constrains you to one inevitable choice at each step in the problem space. Your choice is not deterministic.

In computational theory, a device that can yield different outcomes from the same input and internal state is known as "nondeterministic" (see Hopcroft & Ullman, 1979). Imagine, for instance, a computational device that generates sentences according to the grammar of English. According to one rule of English syntax, a verb phrase can consist of a transitive verb followed by a noun phrase, as in the sentence "John told a joke." According to another rule, a verb phrase can consist solely of an intransitive verb, as in the sentence "Mary laughed." Which rule should be used to produce a sentence? A device that followed some principle in making the choice would be deterministic; one that followed no principle would be nondeterministic. Real computers are deterministic, but they can easily be made to simulate nondeterministic behavior.

There are different ways to interpret nondeterminism. If you are trying to solve the missionaries and cannibals problem for the first time, then at many points in the problem space there will be several possible moves that could be made. Sometimes you may be guided by a hunch or an intuition or some miniscule aspect of your environment, in which case your choice will have been determined by some principle even though you may not have been aware of it. A causal explanation of how the choice was determined by, say, some fleeting memory of another puzzle would amount to a deterministic theory, but at present we have no such explanation. Hence, according to this interpretation, your thinking *is* deterministic, but our ignorance forces us to treat it as nondeterministic. On other occasions, however, you may make a purely arbitrary choice. Experiments have shown that people do not perform in a truly random way (e.g., Baddeley, 1966), but it does not follow that they have no machinery for making arbitrary choices. Indeed, the experimental results suggest that people can make arbitrary choices by means that are not available to introspection. The method is the mental equivalent of spinning a coin, albeit one that is biased and that does not yield independent results from one spin to the next. Still another interpretation of nondeterminism is that your choice depends on the state of your brain, your brain is a physical device assembled out of fundamental particles, and fundamental particles behave according to the indeterminacy of quantum mechanics. This last interpretation seems to be ruled out by the poor performance of people in tasks that call for random behavior. The other interpretations, however, seem plausible: There may be occasions when nondeterminism is a label for our ignorance and other occasions when it characterizes an arbitrary choice.

Types of search in the problem space

Thinking without a goal wanders around a hypothetical problem space defined by the set of possible mental operations that move from one thought to another. The traditional probabilistic account of associations assumes, in effect, that the mechanism is nondeterministic. The choice of the next step in the problem space is not completely determined by psychological factors. Any sort of thought directed toward a goal calls for a sequence of choices that leads from the initial state to the goal. Sometimes the goal is precise, and sometimes you have a procedure that enables you to proceed in the right direction without ever erring; that is, you have a successful deterministic plan. Sometimes, however, the task of finding your way to the solution may be difficult. Here, in theory, you could explore a single route at a time in a "depth-first" search, just as you do in trying to find a way through a maze. When you come to a choice of routes, you select one on the basis of whatever means are at your disposal. In trying to solve a problem, you might choose from your assessment of the potential values of the alterna-

tives. Of course, if you had an absolutely certain method of assessing these values, there would be no difficulty: You would choose the best option at each point and thereby arrive at the destination without ever exploring any blind alleys. Your search would be governed by a deterministic plan.

Many problems, alas, are like the Hampton Court maze. You are forced to make a choice with only an uncertain guide to the value of any alternative. You must therefore simulate a nondeterministic procedure that would always yield the correct choice. "Solving a problem nondeterministically" is in these circumstances just a fancy way of saying "solving a problem by magic." A less magical simulation of nondeterminism is to proceed through the maze until you reach either the goal or a dead end. In the latter case, you can go back to the last point of choice and try a different tack. If you exhaust all the options at this point to no avail, you can go back another step, and so on. If you exhaust all possibilities at all choice points, the problem is insoluble. This procedure of working back through the choice points is called *backtracking*. Anyone who uses it must be as prudent as Theseus, who unwound a ball of thread given to him by Ariadne as he made his way into the Minotaur's labyrinth so that he could be sure of retracing his steps. It is also important to keep a record of each choice made at each choice point. "Those who know no history," it is said, "are doomed to repeat its mistakes." It is the same with simple backtracking, because it fails to take into account the *reason* that a particular choice failed. If you pick up a red hot poker with one hand, backtracking would lead you to try the other hand.

Another method of simulating nondeterminism is to pursue all possible routes in parallel. You start at the initial state, apply all feasible mental operations to it to yield a set of alternative second states, and then do the same to each of these states, and so on. Sooner or later, this so-called breadth-first search leads to the goal if there is at least one route to it.

There are still other methods of search, such as means–ends analysis, which the reader will find described in Chapter 7. But any plan for searching for a route may run into insuperable difficulties. The logician Alonzo Church (1936) proved that there can be no procedure that is guaranteed to determine the status of an argument in the predicate calculus of formal logic (see Chapter 5). If an argument in this calculus has a proof, there are procedures that are guaranteed to find the proof. But if an argument has no proof, there can be no procedure guaranteed to reveal this fact: Any procedure may get lost in the problem space of possibilities, wandering around for an eternity. Computer programs for proving theorems are accordingly designed to minimize the time taken to search for a route that constitutes a proof, because as they grind away there is no way of knowing whether they will ultimately yield a decision or go on computing forever. If a problem is equivalent to an argument in the predicate calculus, there is never any guarantee that one can discover that it is insoluble.

Even in domains that have a guaranteed search procedure, the number of routes to be explored will grow exponentially with each step of the search if there is more than one possible operation at each point. As in the game of chess, it will soon cease to be practicable to explore all possible routes. Hence, no matter what procedure is used, problems for which there is no deterministic procedure almost always require constraints of some sort to keep the search to a manageable size. Very often, the mark of an expert is precisely the ability to explore only fruitful paths. The expert has a knowledge of a domain – often a tacit knowledge – that constrains the search process (see Chapter 7).

So far, I have discussed the ways of finding a path in the problem space from the initial state to the goal, but I have said little about the mental operations that lead from one state to the next. The nature of these operations is, as we shall see, a major factor in distinguishing one sort of thinking from another.

Semantic information and deduction

Consider the following passage in a newspaper:

The victim was stabbed to death in a cinema. The suspect was on an express train to Edinburgh when the murder occurred.

You would probably conclude that the suspect was innocent. This example illustrates a number of phenomena that are typical of everyday reasoning.

First, the inference leads from several verbally expressed propositions to a single verbally expressible conclusion. Even when inferences are based on thoughts rather than words, these thoughts, as Rips argues in Chapter 5, are typically beliefs, that is, entities that may be true or false.

Second, your inference depends both on your understanding of the premises and on your general knowledge. You know, for example, that one person cannot be in two places at the same time and that there are no cinemas on express trains to Edinburgh. You use this knowledge to forge links in the inferential chain so rapidly and automatically that you are hardly aware of them. They play an important part in your comprehension of discourse and in your comprehension of events in the world. Cognitive scientists have proposed a variety of theories about the representation of knowledge in schemata, scripts, and other such structures (see Chapter 9).

Third, you drew an informative conclusion. The concept of semantic informativeness is important, particularly because it has often been overlooked by students of reasoning. Philosophers define semantic information in terms of the possible situations that a proposition eliminates from consideration. The more situations that a proposition eliminates, the more information it contains (see Bar-Hillel & Carnap, 1952; Johnson-Laird, 1983, chap. 2). For example, the assertion "It is freezing but there is no fog" excludes more

situations than the assertion "It is freezing," because the former rules out the presence of fog, whereas the latter leaves the possibility open.

Whenever thinking leads from one state to another in a problem space, one can ask, Does the second state (the conclusion) contain more semantic information that the first state (the premises)? More precisely, does the conclusion rule out some additional situations over and above those ruled out by the premises? If not, the conclusion is a valid deduction. But if it does rule out some additional state of affairs, it is not a valid deduction. This definition is equivalent to Rips's definition: A valid deduction has a conclusion that is true in any state of affairs in which the premises are true. But the concept of semantic information, as we shall see, has some additional uses.

The reader should note that literally an infinite number of valid conclusions follow from any set of premises. Most of them are totally trivial. Consider the following inference:

It is freezing.
Therefore, it is freezing or it is foggy (or both).

It is deductively valid, but no sensible person would draw such a conclusion spontaneously. The conclusion contains *less* semantic information than the premises, and even when people reason validly, they do not throw semantic information away for no good reason. It follows that they must be guided by at least some principle altogether outside logic, because logic sanctions any valid inference including one with a conclusion containing less information than its premises. This consideration, of course, rules out any theory that bases all reasoning on logic alone, for example, the theory proposed by Inhelder and Piaget (1958).

There is a further observation to be made about your inference concerning the stabbing: It is not valid. The conclusion that the suspect is innocent, although plausible, is not necessarily true. Indeed, if you were challenged about it, you would test its validity. When Tony Anderson and I questioned our subjects about such conclusions in some unpublished experiments, they searched for alternatives and often produced scenarios in which the suspect is guilty. For example, he may have had an accomplice, he may have used a spring-loaded knife or a radio-controlled robot, or he may have used a post-hypnotic suggestion that the suspect stab himself during a certain climactic scene in the movie.

Deductive inference should depend on mental operations that do not increase semantic information. For a long time, these operations were assumed to be based on the formal rules of inference of a logical calculus. But as Rips describes in Chapter 5, there are some problems for this doctrine, notably that the content of premises can exert a marked effect on what inference is drawn. Likewise, as Holyoak and Nisbett observe in Chapter 3, the failure of a lengthy course on logic to improve inferential performance casts further doubt on purely formal theories.

Another school of cognitive scientists favors rules containing specific

knowledge. Such systems have been developed in computer programs that function as "expert systems," that is, programs that capture aspects of human expertise and that enable the user of the program to obtain advice about such matters as medical diagnosis, the molecular structure of compounds, or the proper place to drill for oil. The programs rely on conditional rules with specific contents that have been extracted by interrogation of human experts. Although current expert systems differ strikingly from human experts – humans, for example, are rather better at making excuses for their mistakes, there are psychologists who propose that the mind contains content-specific rules of inference in the form of production systems (as outlined in Chapter 7). The conjecture explains the effects of content on reasoning, but it cannot be a complete explanation of human reasoning. It provides no machinery for general deductive ability. It swings too far away from formal procedures.

What we need is the best of both worlds: general inferential ability coupled with sensitivity to content. Another school of thought, which Rips describes, aims to meet this requirement. Its adherents argue that deductive reasoning depends on three processes: (a) imagining the state of affairs described by the premises, (b) using this "mental model" to formulate a conclusion about something that was not explicit in the premises, and (c) attempting to test the validity of the conclusion by searching for an alternative model of the premises in which it is false (see Johnson-Laird, 1983). The process of constructing a model based on the premises takes into account any relevant general knowledge; and the process of searching for alternative models is affected by the apparent truth or falsity of the premises. Although there is evidence supporting both these hypotheses, I will say no more about the possible mechanisms of deduction. The fundamental issue is whether they are, in essence, syntactic manipulations of strings of uninterpreted symbols or semantic manipulations of mental representations of situations. An analogous issue has arisen over language and thought. As Glucksberg points out in Chapter 8, it is now generally accepted that there can be thought without language. To most psychologists, in contrast, the controversy about deductive reasoning is a long way from being settled.

Induction

The discovery of penicillin began with a single observation. Sir Alexander Fleming noticed that areas of bacteria had been destroyed on a culture plate that had been sitting on his desk for a couple of weeks. In fact, a chain of coincidences had led to their destruction. "Chance," Pasteur is supposed to have said, "favors the prepared mind." Fleming was prepared. He knew that the bacteria were hardy, and so he reasoned that something must have destroyed them:

Events of this type do not normally happen.

An event of this type has happened.
Therefore, there is some agent that caused the event.

Making this inference depends on noticing something unusual, a factor that Holyoak and Nisbett note in Chapter 3, and it leads to an increase in semantic information: its conclusion rules out more states of affairs than its premises do. There are indeed systematic processes of reasoning that lead to such conclusions, and these processes are *inductions*. The invocation of a causal agent is an explanatory conjecture of the sort that the American philosopher C. S. Peirce (1931) called an "abduction." One cannot get something for nothing, and the price of trying to expand knowledge, (i.e., increasing semantic information) is the possibility that the step is unwarranted. The conclusion may be false even though its premises are true. Induction should come with a government health warning.

Let us suppose that, under the tutelage of a helpful doctor, you study some cases of smallpox. You note that each patient had prior contact with someone suffering from the disease. You reason thus:

Patient A was in contact with a case of smallpox and A has smallpox.
Patient B was in contact with a case of smallpox and B has smallpox.
. . .
Therefore, if anyone is in contact with a case of smallpox, he or she is likely to catch the disease.

The inference is an induction; it goes from a finite number of instances to a conclusion about every member of a class. It is an example of what Holyoak and Nisbett refer to as an "instance-based" generalization. The resulting conjecture about smallpox seems reasonable, but, to borrow an argument from Nelson Goodman (1955), the evidence also supports the following conclusion:

If anyone is in contact with a case of smallpox, then until the year 2001 he or she is likely to catch the disease and thereafter is likely to catch measles.

Obviously this inference is silly, but why? You might say, "Because we know that diseases no more change their spots than leopards do." But how do we know that? If you are not careful, you may reply, "Because all our observations support this claim." Alas, all our observations are equally consistent with the claim that smallpox will remain smallpox until the year 2001, when it will become measles.

One reaction to this problem is to reject induction altogether. Sir Karl Popper (1972) argues that science is based not on induction but on explanatory conjectures that are open to empirical falsification. And where do conjectures come from? Popper says it does not matter; they can come from anywhere. However, since not all conjectures are equally sensible, and since many of them appear to be based on systematic processes of thought, the problem does not go away. Induction cannot be swept under the cognitive carpet. Its basic operations have indeed been studied in the psychological

laboratory, and inductive computer programs have been implemented in a variety of forms.

Although induction is a way of trying to solve a problem, it too can be treated as a problem in its own right. There is a problem space of possible inductive conjectures, and the goal is to move from the initial state of knowledge to the correct inductive hypothesis. In essence, it calls for a test-operate-test-exit (TOTE) procedure of the sort proposed by Miller et al. (1960). If a test reveals a problem to be solved, such as explaining an unusual event, an inductive operation leads to a hypothetical explanation. A rational individual, however, will not be satisfied with such a hypothesis until it has withstood empirical testing. If a test fails, the cycle may continue with further inductive operations. If the hypothesis withstands testing, it will be accepted, at least provisionally, as the solution to the problem. The concept of semantic information provides a framework for clarifying induction, and, as we shall see, it also suggests a general constraint that people may use in generating inductive hypotheses.

The formulation of inductive hypotheses

There are many potential inductive operations, and their basis can be traced back to John Stuart Mill's (1847) canons of induction (see Chapter 4), which in turn go back to Sir Francis Bacon's (1620/1889) formulation. They boil down to two main ideas. First, if positive instances of a phenomenon have only one characteristic in common, it may play a crucial role. Second, if positive and negative instances differ in only one characteristic, it is critical.

An inductive conjecture may be remote from the truth because it is not even based on appropriate notions (e.g., "smallpox is a punishment for blasphemy"). As Schustack points out in Chapter 4, the most difficult problem is to identify what is relevant. This problem is cracked when the relevant notions are among those available for formulating a hypothesis. The hypothesis should be general enough to include all positive instances of the phenomenon in question but specific enough to exclude all negative instances. There are accordingly, as Holyoak and Nisbett point out in Chapter 3, two main ways in which an inductive hypothesis may have to be revised. On the one hand, it may be too specific and exclude some positive instances: It must be generalized. On the other hand, it may be too general and include some negative instances: It must be specialized. Hence, induction calls for both generalization and its converse, specialization.

One form of generalization, which Holyoak and Nisbett describe, drops part of a conjunction. Thus, the conjecture:

If anyone is in contact with a case of smallpox *and* is elderly, he or she is likely to catch the disease

becomes:

If anyone is in contact with a case of smallpox, he or she is likely to catch the disease.

Another form of generalization adds a disjunction. Thus the previous hypothesis becomes:

If anyone is in contact with a case of smallpox *or* with infected clothes, he or she is likely to catch the disease.

When these changes proceed in the opposite direction, they produce more specific hypotheses.

There are two outstanding questions. First, what is the underlying nature of generalization (and specialization) and, second, how many distinct operations of generalization (and specialization) are there? The answers to both questions can be derived from the concept of semantic information.

The greater the number of possible states of affairs that a hypothesis eliminates from consideration, the greater is its semantic information. Generalization, which has been defined in several ways in the literature (as Holyoak and Nisbett remark), can be analyzed in a simple, uniform way. It is any operation that increases the semantic information of a hypothesis by ruling out at least some additional state of affairs. Specialization has the converse effect; it admits some additional state of affairs. In other words, specialization is a valid inference but one that reduces semantic information for good reason – for example, the step from:

If anyone is in contact with a case of smallpox *or* with infected clothes, he or she is likely to catch the disease.

to:

If anyone is in contact with a case of smallpox, he or she is likely to catch the disease.

The fact that a specialization is always a valid inference does not mean that it necessarily yields a true conclusion; the hypothesis that serves as its premise may be false. Moreover, even if the conclusion is true, it may be less than the whole truth. The premise above is a better explanation of the cause of smallpox than is the specialization.

Holyoak and Nisbett observe that there are many possible generalizations of any hypothesis. Indeed, unless the hypothesis has a very high semantic information content, the number of possible generalizations increases exponentially with the number of simple propositions that may be relevant to the formulation of a generalization (see Johnson-Laird, 1986). An important but unfortunate consequence of this fact is that any procedure based on eliminating putative hypotheses will be unable to examine them exhaustively in a reasonable amount of time. There are so many possible inductions that one cannot examine them all.

Although there are many possible generalizations, there is no need for a

corresponding number of distinct inductive *operations*. Consider, for instance, the generalization that leads from two hypotheses of the form

If p and q, then s,
If p and r, then s,

to one of the form

If p, then s,

This operation is used in the computer model of human inductive reasoning that Holyoak and Nisbett and their colleagues have devised. But it does not require a separate operation of its own. Its premises validly imply

If p and (q or r), then s

and the generalization of this conjecture to

If p, then s

is just a case of dropping part of a conjunction. In fact, it turns out that only three operations are needed for any generalization in the ordinary predicate calculus (see Johnson-Laird, 1986). The first operation consists in conjoining the negation of the description of a situation to the original hypothesis. The second consists in moving from a finite number of observations to a universal claim, as in the earlier inference that contact with smallpox is sufficient for catching the disease. The third operation, yet to be exploited in any theory of induction or by any inductive program (as far as I know), is exemplified by the step from:

Any type of smallpox is cured by some drug.

to:

There is some drug that cures any type of smallpox.

Even though there are only three basic forms of generalization, it remains wholly impracticable to examine all their possible uses in generalizing a hypothesis. One moral to be drawn from this observation, and from Goodman's argument, which I presented earlier, is that induction cannot be a matter of manipulating symbols according to purely formal or syntactic rules. A hypothesis of a particular form may have one appropriate generalization in one domain and quite a different appropriate generalization in another domain. Another moral, which is drawn by Holyoak and Nisbett as well, is that the search for the appropriate generalization (within a vast problem space) must be constrained in some way. They describe a number of constraints to which I shall add a further candidate based on semantic information.

When human beings try to induce a novel hypothesis, they concentrate on positive exemplars of it (see the "confirmation" bias referred to in Chapter

4 and Peter Wason's 1977 study of the failure to examine disconfirming evidence). Thus, they concentrate on the people with smallpox rather than healthy individuals. In such circumstances, it is important not to formulate a hypothesis that contains too little semantic information. For example, in order to teach you to identify a particular disease, I show you a patient who has a fever, a rapid pulse, and a backache. You should therefore hypothesize that the disease has the following symptoms: fever, rapid pulse, and backache. This conjecture contains the largest amount of semantic information based directly on the evidence. If I now present a patient with a fever, backache, and a *slow* pulse, you will realize at once that your previous hypothesis eliminates too much. You will modify it to the maximally informative one based on the evidence: fever and backache. Suppose, however, that you had started off with the following conjecture: fever or rapid pulse or backache. It fits the facts, but it contains much less semantic information. Moreover, it remains unaffected by the evidence from the second patient. You will not home in on the real disease from positive exemplars alone, because your initial hypothesis will always accommodate them. Hence, when you are trying to formulate a hypothesis from positive instances, you must advance the most semantically informative hypothesis based on the data. It may rule out too much, but if so, sooner or later you will encounter a positive instance that allows you to correct it.

When children develop their taxonomies of the world, they appear to be guided by this principle. Frank Keil (1979) has shown that they organize their concepts in hierarchies, as in Figure 15.1. Overlapping arrangements like the one in Figure 15.2 are rare and sometimes arise from ambiguities. Keil derives the children's classifications from the pattern of their answers to such questions as: Does it make sense to say that a tree is an hour long? A child may have the following taxonomic rule:

If something is living, then it is a person.

An older child, however, distinguishes two classes:

If something is living, then it is a person or a plant (but not both).

This way of refining a taxonomy suggests that children are sensitive to semantic information. If a category is to be divided, the division that creates the most semantic information is one that yields two mutually exclusive subcategories; that is, no entity can belong to both. Perhaps it is this semantic principle that leads children to avoid overlapping taxonomies.

Knowledge as a constraint on inductive thinking

Some theorists, notably the linguist Noam Chomsky (1980), have suggested that there may be no general inductive procedures, only specific procedures based on innate knowledge of particular domains. The claim is debatable,

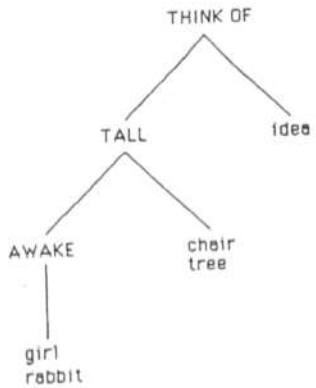

Figure 15.1. A tree representing the typical judgments of a five-year-old in Keil's study.

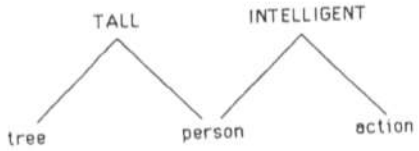

Figure 15.2. An artificial tree showing the sort of overlap that does not occur in children's judgments. The overlap occurs because of an ambiguity: An intelligent action does not possess intelligence as a person does.

but certainly a major constraint on induction is knowledge of a domain. Most of the inferences we make go beyond the information we are given and, like Fleming's inference about the destruction of the bacteria, they do so because they rely on our knowledge. We avoid any inductive conjecture that goes against our knowledge of the way the world works; we are biased toward conjectures that are compatible with our knowledge. It is no coincidence that knowledge is stressed as an important factor in so many chapters in this book: by Schustack, by Holyoak and Nisbett, by Perkins, by Sternberg, by Smith, Sera, and Guttuso, and in all the schemes for teaching thinking reviewed by Vye, Delclos, Burns, and Bransford.

The importance of knowledge is easily demonstrated by asking people to make judgments about matters on which they know little. For example, I attach a ball to a piece of string and spin it horizontally in a circle. The string snaps. What is the horizontal trajectory of the ball afterwards? People who do not know Newton's laws often say that it moves off in a spiral that gradually staightens out. They appeal to an almost medieval notion of impetus: something that a moving object is supposed to have (McCloskey, 1983). The correct answer, ignoring gravity, is that the ball moves in a straight line; the centripetal force pulling it toward the center of the circle ceases when

the string breaks, and an object on which no forces are exerted moves uniformly in a straight line (or remains in a state of rest).

Knowledge of the variability of instances is critical when people make an inductive generalization. As Holyoak and Nisbett report in Chapter 3, a single exemplar suffices for most of us to conclude that a rare element conducts electricity, but a single instance of an obese member of an exotic tribe leads to no generalization. We know that people's properties vary much more widely than those of a chemical element.

A full grasp of variation calls for knowledge of the theory of probability. The theory is so different from its intuitive precursors that at least one commentator, Ian Hacking (1975), has remarked that anyone who had played dice in ancient times armed with the modern calculus of probabilities would soon have won the whole of Gaul. Numerous studies have shown that people make egregious errors of judgment because of their ignorance of the workings of probability (see Chapter 6). The burden of these findings is that the errors of the naive are reproduced at a sophisticated level by the experts. Amos Tversky and Daniel Kahneman have proposed an account of the phenomena, which assumes that people use various heuristics (rules of thumb) in order to estimate the probabilities of events and that they are particularly affected by the relative availability of information and by its seeming representativeness (see Chapter 6). Tversky and Kahneman's (1973) explanation dovetails neatly with the account of reasoning by mental models. They write:

> Some events are perceived as so unique that past history does not seem relevant to the evaluation of their likelihood. In thinking of such events we often construct *scenarios*, i.e. stories that lead from the present situation to the target event. The plausibility of the scenarios that come to mind, or the difficulty of producing them, then serve as a clue to the likelihood of the event. If no reasonable scenario comes to mind, the event is deemed impossible or highly unlikely. If many scenarios come to mind, or if the one scenario that comes to mind is particularly compelling, the event in question appears probable. (p. 229)

Concepts and theories

A baby girl at 16 months hears the word *snow* used to refer to snow. Over the next months, as Melissa Bowerman (1977) has observed, the infant uses the word to refer to snow, the white tail of a horse, the white part of a toy boat, a white flannel bed pad, and a puddle of milk on the floor. She is forming the impression that *snow* refers to things that are white or to horizontal areas of whiteness, and she will gradually refine her concept so that it tallies with the adult one. The underlying procedure is inductive. We all continue to make inductions throughout our lives as we form impressions about classes of people, events, and the meanings of expressions.

Psychologists have often studied induction in terms of the means by which

people acquire concepts. Following Mill and other Empiricist philosophers, they assumed until recently that a process of abstraction drops idiosyncratic details that differ from one exemplar to another and leaves behind only what they hold in common. But as Smith emphasizes in Chapter 2, this classical view of concepts does not hold for all concepts. Kenneth Smoke had this worry in the 1930s: "As one learns more about dogs, his concept of 'dog' becomes increasingly rich, not a closer approximation to some bare 'element' . . . . No learner of 'dog' ever found a 'common element' running through the stimulus patterns through which he learned" (Smoke, 1932).

Everyday concepts are not isolated, independent entities; they are related to one another. This idea goes back to the *structuralism* of the Swiss linguist Ferdinand de Saussure (1960). The perceptual boundaries of entities are set, in part, by the taxonomy in which they occur. Whether something is perceived as a dog depends on its similarity to typical dogs, typical cats, typical wolves, and so on. Granted the complex structures of certain domains, much of the development of knowledge, as Smith, Sera, and Gattuso argue in Chapter 13, depends on creating mental representations from which the appropriate relations among concepts can be recovered. It is natural to assume that these relations are represented explicitly, but this assumption should not be taken for granted. One of the interesting features of the current work on parallel distributed processing is that relations among concepts may not be explicitly represented at all, but may merely be an emergent property of an implicit representation that is distributed over many parallel processors (see Rumelhart, Smolensky, & McClelland, 1986).

Analogy

Lying behind the perceptual characteristics of concepts are schemata that relate form to function; and lying behind such schemata is a conceptual *core* – some kind of everyday "theory," that plays a role in the ordinary use of language that is analogous to the role of theories in scientific discourse (Miller & Johnson-Laird, 1976, Sec. 4.4.4). The development of a theory, however, may call for the discovery of the relevance of certain new ideas, or new combinations of existing concepts (see Smith's description of some of the principles governing conceptual combinations in Chapter 2). It is seldom merely a question of making inductive generalizations based on a given set of ideas.

New ideas come from mental operations (other than induction) that lead to an increase in semantic information. One such source is analogical thinking. When you realize that a problem (the target domain) is analogous to another more familiar topic (the source domain), you may be able to import new ideas into the target domain from the source domain; see, for example, Holyoak and Nisbett's account of the way an analogy can help you to solve

the celebrated X-ray problem. Thus, analogies are important because they can provide the novel ideas necessary for the development of a new theory. (Generalization and specialization work only if the relevant ideas are already available.)

Could there be a purely formal theory of analogy? The answer appears to be negative for the same sort of reasons that formal theories of induction are impossible. Consider, for example, Rutherford's elucidation of the structure of an atom by analogy to the solar system. As Dedre Gentner (1983) has pointed out, this analogy maps the sun onto the nucleus of the atom and maps the planets onto the electrons. The properties of the sun, such as its color, are dropped, but the higher-order semantic relations are carried over. The sun's attraction of the planets causes them to revolve around it. Hence, it is inferred that the nucleus's attraction of the electrons causes them to revolve around it. If you ask people in what way a clock is analogous to the solar system, from my anecdotal observations, they are likely to respond: "Both involve a revolution: the hands of the clock go round just as the planets go round the sun." This answer and Rutherford's analogy depend on mappings from the same objects in the source domain:

Atom target		Source		Clock target
nucleus	←	sun	→	center
electrons	←	planets	→	hands

A purely formal theory would therefore lead to the transfer of the same information in both analogies. But unlike the case of atomic structure, the causal relation should not be carried over in the analogy with clocks: "The center's attraction of the hands causes them to revolve around it." This conclusion is obviously false, but matters of fact are precisely what formal theories must *not* depend on. Current theories of analogy accordingly rely on semantics and matters of fact (see the theories discussed in Chapters 3 and 10). Once again, knowledge is at the heart of the matter.

Creativity

Innovations in science and art often arise as a result of analogical thinking (see Hesse, 1966). Such analogies, however, call for genuinely creative thought. David Perkins argues in Chapter 11 that creativity calls for results that are both original and appropriate. I have similarly suggested that an act of creation yields a product that is novel (at least for the individual who created it) and that satisfies some existing criteria or constraints: One creates pictures, poems, stories, sonatas, theories, principles, games, and so on, and anything that lies outside the criteria of any domain is likely to be deemed uncategorizable rather than creative (Johnson-Laird, in press). Of course, the process does not occur in a vacuum; one cannot construct new ideas out of nothing. There must be mental elements that already exist –

concepts, images, principles, and so forth – that provide the raw materials for the process and the constraints on it. Yet there *are* new things under the sun; a combination, or modification, of existing elements can indeed be novel. Highly original works of art and science are constructed out of existing languages, such as English and mathematics. The criteria that I have referred to are not necessarily the sorts of explicit principles that are found in theoretical treatises on aesthetics and scientific method. They are any principles that an individual uses in order to constrain the processes by which elements are combined, modified, or refined within a particular domain. Most criteria will probably be implicit principles that are not available to introspection. Some of them may be common to many creators; they specify the genre or paradigm. Others may be unique to individuals; they constitute the idiosyncracies of individual style within the genre or paradigm. In short, the criteria are constraints on the mental operations available to the creator.

Unlike a reasoning problem, there is no clear and explicit starting point in the problem space for an act of creation. The creator possesses only the criteria of the domain, and although they place constraints on what can be created within that domain, they still allow an indefinite number of possibilities. If the criteria allowed only one possible continuation at each step in grappling with the task, the process would be trivial. There would be no choice about what to do next. Since creators almost always have a choice of continuations, it follows that the mechanisims of creation must be treated as nondeterministic.

This account is the beginning of a theory of creativity at the computational level. It tells us what a creative process has to do, namely, it must start with a set of criteria and make nondeterministic choices among the options they offer. Given these foundations, a striking conclusion can be derived: Only three general classes of procedure are capable of creativity. These classes have nothing to do with the details of creative mechanisms, though they place constraints on them, but rather concern the overall architecture of the creative process. They are as follows:

1. *Neo-Darwinian procedures*. They make arbitrary combinations of or changes in existing elements so as to generate a vast number of putative products. They then exploit the criteria of the domain so as to filter out those products that are not viable. Some theorists have argued that such procedures are the only mechanism for creativity (e.g., Skinner, 1953).
2. *Neo-Lamarckian procedures*. They form initial combinations of or changes in existing elements under the immediate guidance of the criteria of the domain. If a choice between equally viable alternatives arises, it is made arbitrarily. The choice has to be arbitrary, since by definition all the available criteria of the domain are used in the initial generation of ideas.
3. *Multistage procedures*. They make use of some criteria in the initial generative stage; they then use other criteria as filters. This procedure is perhaps the one that Perkins has in mind when he suggests that creativity is a pro-

cess of search and selection. Choices between equally viable alternatives may arise. Once again, certain of these choices must be made arbitrarily, since the complete set of criteria does not pinpoint a single unique product.

The mental criteria that a creator exploits can obviously differ in terms of their completeness. For certain domains, the criteria are complete. That is to say, the creator has sufficient criteria to guarantee that the result of the creative process will be at least viable. Completeness is thus a desirable property for all creation that occurs extempore, such as the making of artifacts in media that allow no second chances and the improvization of music, dance, poetry, and other forms of art. In such cases, the creator's mental operations define a problem space in which all routes lead to at least a satisfactory outcome. We can therefore think of the criteria as defining just the set of feasible routes, and it is natural to suppose that the creative procedure in this case will be neo-Lamarckian: a nondeterministic walk through a problem space that leads only to viable outcomes. Given the limitations of human processing capacity, the computational power of the procedure is likely to be weak, that is, to call on the minimal possible memory for the results of intermediate computations. I have examined jazz improvization as a test case of this sort of creativity and shown that it can be modeled in computer programs based on such procedures. Musicians improvizing in a particular genre have a tacit grasp of the criteria of the genre, which they can use to generate music spontaneously. If their grasp of the criteria is inadequate, they will produce unacceptable music and fail to find gainful employment as improvizers.

The creation of a poem, painting, or symphony is usually carried out within the conventions of an existing genre. Likewise, the creation of science normally occurs within the constraints of an existing paradigm (Kuhn, 1970). These sorts of creativity nearly always depend on a multistage procedure. There is no complete set of criteria that leads only to viable outcomes, but the initial generative stage can be partially constrained by some criteria. The result, however, almost always calls for further revision or elaboration, and this process may be governed by criteria that the creator is unable to exercise in the generative stage – we are all better critics than creators. This division of labor is exemplified by the cases that Perkins discusses. The problem space contains many routes that fail to terminate in an acceptable goal, but the creator is not striving to achieve a single unique goal; there are many acceptable goals.

In fact, the notion of problem spaces containing goals a priori is often merely a convenient fiction for creation. Unlike conventional problem solving, the process may not be goal driven in any realistic sense. Creators can start off with no very clear goal. They make a sequence of choices on the basis of often tacit criteria. They may not recognize their goal until after they have achieved it, or they may fail to achieve any worthwhile result.

However, a multistage procedure can call for considerable computational power, that is, memory for the intermediate results of computations. Thus, the creation of tonal chord sequences of the sort used in most Western music evidently calls for a considerable use of such memory (see, e.g., Johnson-Laird, in press; Steedman, 1982). Writing or a notation of some sort, of course, relieves the creator of the actual burden of remembering intermediate results, and in certain forms of art, such as sculpture and painting, the work itself provides such a record.

In the case of a major innovation in art or science, there are grounds for doubting whether there could ever exist criteria that always guarantee a successful outcome. On the one hand, there are too few instances of revolutions within a particular domain. On the other hand, it is hard to see what different innovations could possibly have in common. What criteria are common, for example, to both the invention of perspective and the invention of Cubism? What criteria are common to both the transition in physics to Newtonian mechanics and the transition to the special theory of relativity? By criteria here, I have in mind knowledge that would be effective in reducing the processes required to generate *all* successful revolutions within a particular domain. Indeed, it hardly has to be said that even the best of innovators may try out many bad ideas before discovering a good one.

Let us consider the invention of a profound analogy as a special case of this sort of creativity. By definition, the analogy does not depend on preexisting rules that establish mappings between the source and target domains. The innovation depends on the invention of such mappings. Establishing a mapping is a process that resembles the construction of a complex proposition that links an element in one domain with an element in the other. We can think of all domains of knowledge as constituting a vast epistemic space, which embraces knowledge of the solar system, of atoms, of waves, of clocks, of clouds, and so on. The task of creating a profound analogy consists initially in constructing a mapping from one domain to another. The more distant the two domains are from one another (before the construction of the analogy) the larger is the number of domains that might serve as the source, and the longer is the chain of links that will have to be established to form the mapping. Granted that at each point in the construction of a chain there are several possible continuations, the mapping is like the construction of a novel sentence – a sentence that captures the content of the mapping. Plainly, the number of possible sentences increases exponentially with the length of the sentence; and it soon ceases to be feasible to explore all possible mappings.

There are computer programs that have produced novel proofs of theorems, interesting mathematical conjectures, rediscoveries of scientific laws, and works of art (see Chapter 11). Their success depends on their operating in highly constrained domains, using a neo-Lamarckian procedure or assis-

tance from the user (or both). Even a neo-Darwinian procedure will work if, like nature, one is prepared to use it over and over again in a cumulative way in billions of experiments every year for a period extending over millions of years and to countenance a high proportion of failures. The point of my argument is that there can be no feasible program that is *guaranteed* to make profound discoveries routinely by using analogies or any other procedure.

How, then, do those few exceptional individuals, whom we recognize as geniuses, succeed in making innovations? Is there perhaps some mental commodity, or "potency" to use Perkins's term, that leads to success – a higher degree of intelligence, a larger working memory, a more rapidly functioning brain, a larger number of associative connections, a higher degree of motivation, or an infinite capacity for taking pains? I suspect not. What evidence there is suggests that creativity is not merely a matter of some such property being enhanced; there are many highly intelligent and dedicated individuals (by any measure) who lack the spark of originality. My conjecture is that geniuses have mastered more constraints, but they have their knowledge in a form that can directly govern the generative stage of creation. Knowledge is the key – in this case knowledge of the specific domain, since, as I have argued, there are not likely to be any general criteria for innovation. But knowledge alone is not enough. To return to the inference about the murder in the cinema, everyone recognizes the ingenuity of the solution that the suspect used a posthypnotic suggestion that the victim stab himself. Yet very few people succeed in thinking of this solution for themselves. Conscious critical knowledge, which is relatively easy to acquire (and for educators to test), is impotent when it comes to the unconscious generation of ideas.

How knowledge comes to work in the generative stage of creativity is perhaps the most important mystery confronting students of thinking. One conjecture is that it does so only as a result of an individual's attempts to create. The only way to learn to be creative is by trying to create. If there is any truth in this conjecture, the pedagogical moral is that the best method of fostering creativity may be to encourage children to attempt to create within a particular domain as soon as they have acquired the rudiments of technique.

Free will, self-reflection, and metacognition

I have now discussed several types of thought. Is there any other sort? There is indeed one very important additional mechanism. A salient element of our conscious experience is self-reflection. We have the capacity to reflect upon what we are doing – our own process of thought becomes itself an object of thought at a higher level – and as a result of this self-reflection we may modify our performance. For example, if you are having some success in solving problems of a particular class but then you are stumped by a

certain problem, you can ask yourself: "What was I doing when I succeeded with the earlier problems?" Or, to take an example from Chapter 5, if a problem reminds you of some other domain, you may say to yourself: "I should try to draw an analogy here." Such thoughts are based on your ability to scrutinize your own performance, that is, to raise yourself up one level to become a spectator of your own thoughts and behavior. This procedure may help you to reformulate how you should proceed at the lower level of actual performance.

You cannot inspect your own thought processes in complete detail. If you could, there would hardly be any need for books on the psychology of thinking. What you have access to is something like a *model* of your own abilities – an incomplete and perhaps partially erroneous representation of their major features (see Johnson-Laird, 1983, chap. 16). This ability of the mind to inspect models of its own performance and then in turn to use these models in thinking is the basis of all the so-called metacognitive skills that you possess. This account is one way in which Sternberg's idea of "metacomponents" can be explicated (see Chapter 10). Hence you can think about how you remember things and take remedial steps to improve your memory (see Chapter 13 for some observations of the development of this ability in childhood). You can think about how you get on with people and work out a strategy for coping with difficult social situations. But self-reflection does not stop here. It, too, can be the object of itself: you can think about your own metacognitive thoughts. When you start to think about how you ordinarily deal with problems of a certain sort, you may realize what you are doing and think, "This is one of those problems that I can tackle by thinking about the way I have solved similar problems in the past, but whenever I use this ability, I tend to concentrate too much on previous successes." There does not appear to be any barrier that in principle prevents you from reflecting about such thoughts at a still higher level.

The ability to reflect at ever higher levels is essential to freedom of choice. When you follow a plan, you sometimes carry out a "cast iron" sequence of actions, that is, a deterministic sequence like that which underlies calculation. But often you observe the outcomes of your actions and, as a result, may modify the plan or even on occasion abandon it altogether. You usually have the freedom to choose among several options at various points in its execution, particularly if you are engaged in the creative exercise of your imagination.

The concept of freedom that I here invoke refers to freedom of will – the propensity that thinkers from Descartes (1637/1911–12) to Dostoyevsky (1864/1972) invariably cite in order to cast doubt on the feasibility of a science of the mind. Scientists often retort that free will is an illusion (e.g., Skinner, 1971); yet its existence is entirely compatible with the capacity for self-reflective thinking.

Suppose, for instance, that you are confronted with a choice between putting milk or lemon in your tea. Sometimes, you decide what you want almost automatically and without thinking about it. (If you are Richard Feynmann, you may even choose both milk and lemon!) On other occasions you may be unable to make up your mind. Sooner or later in this case, you will say to yourself, "This is ridiculous; I'll have to choose one of them." And you may then, as a result of this higher-order reflection, make an abitrary decision. You may even ensure that it is arbitrary by recourse to external means. You may spin a coin or, like the hero of Luke Rhinehart's novel *The Diceman*, toss dice.

What gives you free will is the self-reflective ability to think about *how* you will make a decision and thus to choose at a metalevel a method of choice. At the lowest level, you can make a choice without thinking about it at all. You just pour milk into your tea or put a slice of lemon into it:

Level 0: Pour milk into your tea.

At the metalevel, you think about what to do and make a decision based, say, on a simple preference (see Chapter 6):

Level 1: By assessing preferences, you choose from:
Level 0: Pouring milk into your tea.
Putting a slice of lemon into your tea.

How did you arrive at this method of choice? You did not think about it consciously. It was a tacitly selected method that came to mind as the right way to proceed. Perhaps most choices are made this way. But the metalevel method need not be tacitly chosen. You can confront the issue consciously (at the meta-metalevel). And indeed if you do reflect about the matter, you may assess different methods of choice and try to choose rationally from among them:

Level 2: Making a rational assessment, you choose from:
Level 1: Assessing preferences
Taking your spouse's advice } to choose from:
Spinning a coin
Level 0: Pouring milk into your tea
Putting a slice of lemon into your tea.

The method of decision at the highest level is, of course, always tacitly selected – it just comes to mind. If it were chosen consciously, there would be a still higher level at which that decision was made. In theory, there need be no end to the hierarchy of decisions about decisions about decisions, but the business of life demands that you do something rather than get lost in speculation about how to decide what to do. The buck must stop somewhere.

We have free will, not because we are ignorant of the roots of many of our decisions, which we certainly are, but because our models of ourselves enable us to choose how to choose, and among the range of options are

those arbitrary methods that free us from the constraints of an ecological niche or any rational calculation of self-interest.

Intentionality and self-reflection

Once you have decided what to do and how to do it, you can act intentionally to try to achieve your goal. There are computer programs that are goal driven, that is, that try to achieve a stated goal. Some cognitive scientists have argued that these programs have intentions. However, it seems more accurate to say that they act *as though* they had intentions. What is missing from them is self-knowledge. At the lowest level (like the computer programs), human beings can

Level 0: Construct a model of a possible future state of affairs.
Compute what to do to try to bring about that state of affairs.
Carry out this plan.

Unlike a computer program, human beings have access to a model of these abilities, and moreover they can use it by:

Level 1: Determining what to do by consulting a model of:
Level 0: Constructing a model of a possible future state of affairs.
Computing what to do to try to bring about that state of affairs.
Carrying out this plan.

In other words, people know that they can act to try to achieve some goal, and they can use this knowledge in determining what to do.

Once again, as the theory allows, people know that they can take into account their self-knowledge in making decisions. They can

Level 2: Determine what to do by consulting a model of:
Level 1: Determine what to do by consulting a model of:
Level 0: Construct a model of a possible future state of affairs.
Compute what to do to try to bring about that state of affairs.
Carry out this plan.

In other words, people know that they know that they can act to try to achieve some goal, and they can use this knowledge in determining what to do. Even this level is not necessarily the top of the hierarchy.

Of course, most of us recognize that the road to hell is paved with good intentions. We know that our having a particular intention, such as to give up smoking, is not necessarily sufficient to produce the appropriate actions. In the light of this knowledge, we sometimes take special steps to try to ensure an intended outcome.

When you are thinking about something, you can be so deeply engrossed in it that you forget all about your own condition. But you can perceive yourself as thinking about a problem – perhaps as a precursor to a metacognitive step. This state of self-awareness is phenomenologically distinct

from ordinary perception and is perhaps the central riddle of human consciousness. What gives rise to self-awareness according to the present theory is the self-reflective mode of processing. Normal perception yields a model of the world; self-awareness depends on the mind constructing a model of itself constructing the model of the world. You perceive yourself perceiving the world or cogitating about it. Once again, the model representing perception is a radically incomplete one, but it is sufficient to create the subjective experience of self-awareness.

Conclusions

I have described a variety of types of basic thinking and above them all a higher-order type: self-reflection. Since we can carry out a calculation in the midst of a daydream, or daydream in the midst of a calculation, their names are merely convenient labels that reflect combinations of underlying distinctions. The taxonomy founded on these distinctions can be summarized in terms of the following questions:

Does a process of thought have a goal? If not, it is of the family of associative thinking, which includes the genera of dreams and daydreams. If it has a goal, it falls into the major family of thinking, which psychologists call problem solving. There are many genera here, and their classification continues:

Is the thought process deterministic? If it is, obviously it leads to a single precise goal and constitutes the genus of calculation. If it is not deterministic, then again there are many genera, and the classification continues:

Is there an explicit starting point? If not, the process is in the family of creative processes, of which there are three main species (neo-Darwinian, neo-Lamarckian, and multistage). If there is an explicit starting point, it is in the family of reasoning processes, and the classification continues:

Does the reasoning process increase semantic information? If so, it is a species of induction. If not, it is a species of deduction.

Figure 15.3 presents the outlines of this taxonomy, which can obviously be refined into many subspecies. The taxonomy omits self-reflection (metacognition), which depends on having access to a model of a thought process. All the genera of problem solving appear to be potential candidates for self-reflection. When thinking lacks a goal, however, matters are less clear. If you are daydreaming and start to reflect on the process, you can indeed influence its nature. Often, however, your metacognitive thoughts lead you to abandon the daydream and to enter into deliberations about some problem that emerges from it. If you are having a real dream and start to reflect on the process, the dream becomes what is sometimes known as "lucid": You are aware that you are dreaming. Most people find it difficult to influence the content of a lucid dream, but they can usually at least decide to wake

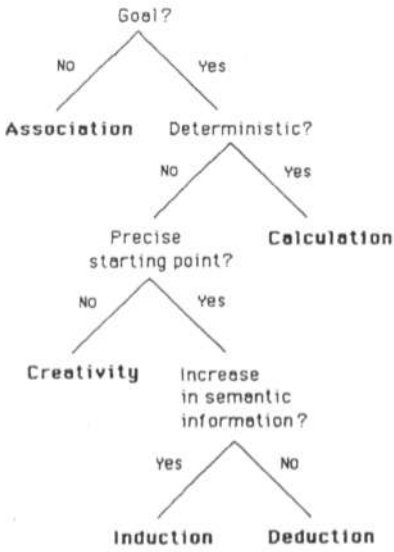

Figure 15.3. A summary of a taxonomy of thought (excluding self-reflection).

up. The essential point is that metacognition often changes the character of purely associative thinking. It either introduces a goal or brings it to a halt. It is hard to be an innocent witness of one's own thought processes.

Is the taxonomy complete? Perhaps. The underlying distinctions seem to be both fundamental and exhaustive, and they embrace all the varieties of thought that have been discussed in this book. A species of thinking that they failed to encompass would be a major discovery. It is clear, however, that thinking in practice may call for combinations of different genera.

The taxonomy derives primarily from an analysis at the computational level – an analysis in terms of what a thought process is computing. The reader may wonder whether it implies anything about the way thinking is carried out at the algorithmic level. In fact, one theme that has emerged in this chapter is that all the species of thought can be explained on the assumption that each is guided by knowledge and depends on representations of the world. There is only one domain – deductive reasoning – in which accounts based on the formal manipulation of uninterpreted symbols are still pursued by some theorists.

References

Bacon, F. (1889). *Novum organum* (T. Fowler, Ed.). New York: Oxford University Press. (Original work published 1620).

Baddeley, A. D. (1966). The capacity for generating information. *Quarterly Journal of Experimental Psychology*, *18*, 119–29.

Bar-Hillel, Y., & Carnap, R. (1952). An outline of a theory of semantic information. In Y. Bar-Hillel, *Language and information*. Reading, MA: Addison-Wesley, 1964.

Bowerman, M. (1977). The acquisition of word meaning: An investigation of some current concepts. In P. N. Johnson-Laird & P. C. Wason (Eds.), *Thinking: Readings in cognitive science* (pp. 239–53). Cambridge University Press.

Brook, S. (1983). *The Oxford book of dreams*. New York: Oxford University Press.

Chomsky, N. (1980). *Rules and representations*. New York: Columbia University Press.

Church, A. (1936). A note on the Entscheidungsproblem. *Journal of Symbolic Logic*, *1*, 40–1, 101–2. Reprinted in M. Davis (Ed.), *The Undecidable*. Hewlett, NY: Raven Press, 1965.

Descartes, R. (1911–12). *Discours de la méthode*. In *The philosophical works of Descartes* (E. T. S. Haldane & G. T. R. Ross, Trans.). Cambridge University Press. (Original work published 1637.)

Dostoyevsky, F. (1972). *Notes from the underground*. Harmondsworth: Penguin Books. (Original work published 1864.)

Fodor, J. A. (1983). *The modularity of mind: An essay on faculty psychology*. Cambridge, MA.: Bradford/MIT Press.

Galton, F. (1883). *Inquiries into human faculty and its development*. London: Macmillan.

Gentner, D. L. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Psychology*, *7*, 155–70.

Goodman, N. (1955). *Fact, fiction, and forecast* (2nd ed.). Cambridge, MA: Harvard University Press.

Hacking, I. (1975). *The emergence of probability*. Cambridge University Press.

Hesse, M. (1966). *Models and analogies in science*. Notre Dame, IN: Notre Dame University Press.

Hopcroft, J. E., & Ullman, J. D. (1979). *Introduction to automata theory, languages, and computation*. Reading, MA: Addison-Wesley.

Inhelder, B., & Piaget, J. (1958). *The growth of logical thinking from childhood to adolescence*. London: Routledge & Kegan Paul.

James, W. (1890). *The principles of psychology*. New York: Holt.

Johnson-Laird, P. N. (1983). *Mental Models: Towards a cognitive science of language, inference, and consciousness*. Cambridge University Press; Cambridge, MA: Harvard University Press.

(1986). *Semantic information: A framework for induction* (Mimeo). Cambridge, Eng.: MRC Applied Psychology Unit.

(in press). Freedom and constraint in creativity. In R. J. Sternberg (Ed.), *The Nature of Creativity*.

Johnson-Laird, P. N., Herrmann, D. J., & Chaffin, R. (1984). Only connections: A critique of semantic networks. *Psychological Review*, *96*, 292–315.

Jung, C. G. (1919). *Studies in word association*. New York: Moffat Yard.

Keil, F. C. (1979). *Semantic and conceptual development: An ontological perspective*. Cambridge, MA: Harvard University Press.

Kuhn, T. S. (1970). *The structure of scientific revolutions* (2nd ed.). University of Chicago Press.

Lashley, K. S. (1951). The problem of serial order in behavior. In L. A. Jeffress (Ed.), *Cerebral mechanisms in behavior*. New York: Wiley.

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. New York: Freeman.

McCloskey, M. (1983). Naive theories of motion. In D. Gentner & A. L. Stevens (Eds.), *Mental Models*. Hillsdale, NJ: Erlbaum.

Mill, J. S. (1847). *A system of logic Book* 3. London: Macmillan.

Miller, G. A., Galanter, E., & Pribram, K. (1960). *Plans and the structure of behavior*. New York: Holt, Rinehart & Winston.

Miller, G. A., & Johnson-Laird, P. N. (1976). *Language and perception*. Cambridge University Press; Cambridge, MA: Harvard University Press.

Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice-Hall.

Peirce, C. S. (1931–58). *Collected papers* (8 vols.; C. Hartshorne, P. Weiss, & A. Burks, Eds.). Cambridge, MA: Harvard University Press.

Popper, K. R. (1972). Conjectural knowledge: My solution to the problem of induction. In *Objective knowledge: An evolutionary approach*. New York: Oxford University Press (Clarendon Press).

Rumelhart, D. E., Smolensky, P., & McClelland, J. L. (1986). PDP models of schemata and sequential thought processes. In J. L. McClelland, D. E. Rumelhart, & the PDP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition: Vol. 2: Psychological and biological models*. Cambridge, MA: Bradford/MIT Press.

Saussure, F. de (1960). *Course in general linguistics*. London: Owen.

Skinner, B. F. (1953). *Science and human behavior*. New York: Macmillan.

(1971). *Beyond freedom and dignity*. New York: Knopf.

Smoke, K. L. (1932). An objective study of concept formation. *Psychological Monographs, 42* (Whole No. 191).

Steedman, M. J. (1982). A generative grammar for jazz chord sequences. *Music Perception*, 2, 52–77.

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology, 4*, 207–32. Reprinted in D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgement under uncertainty: Heuristics and biases*. Cambridge University Press, 1982.

Wason, P. C. (1977). 'On the failure to eliminate hypotheses . . .' – a second look. In P. N. Johnson-Laird & P. C. Wason (Eds.), *Thinking: Readings in cognitive science*. Cambridge University Press.