13 Logical Thinking: Does it Occur in Daily Life? Can it Be Taught?

P. N. Johnson-Laird
MRC Applied Psychology Unit

Overview

There is an important distinction between two sorts of inference that occur in daily life. On the one hand, *implicit* inferences are rapid, effortless, and usually outside conscious awareness; they play a crucial role in the comprehension of discourse. This chapter reports some experimental results showing that children are often poor at making such inferences, and that one difference between good readers and bad readers is precisely the ability to make such inferences. On the other hand, *explicit* inferences are made in answering questions and solving problems. They can be genuinely deductive, unlike implicit inferences that tend to be plausible conjectures that may subsequently be discounted.

Psychologists committed to the view that conscious thought is rational have argued that there is a mental logic underlying deduction. This chapter reviews some alternative systems of logic, including programs based on the "resolution" principle, the PLANNER programming language, and the so-called method of "natural deduction." However, it goes on to argue, from a consideration of how children might acquire logic, that valid inferences can be made without the use of any rules of inference.

A summary of some experimental studies suggests that there are at least two levels at which discourse is normally represented: a relatively superficial representation close to the linguistic form of sentences, and a more profound representation that takes the form of a *mental model* of the state of affairs described by the discourse. This concept is used to explain how valid deductions are made

without logic. Human reasoners follow the fundamental semantic principle governing all deduction: An inference is valid if there is no way of interpreting the premises that is consistent with the denial of the conclusion. Implicit inferences may well occur in the construction of mental models. What distinguishes explicit deductions, however, is an attempt to construct and to evaluate alternative mental models of the same premises. A number of experimental findings are reported that support this hypothesis. Fallacious inferences occur as a consequence of failures to carry out the search for counterexamples in a systematic and comprehensive way. These failures seem to be the result of inherent limitations in the capacity of working memory, because the main differences in logical ability from one individual to another concern those deductions that demand the construction of more than one mental model. Some pedagogical implications of the theory are briefly sketched.

INTRODUCTION

At the end of a lecture on ethics, Epictetus, the Stoic philosopher, recommended the study of logic to his audience because it was useful. One of his listeners was unconvinced and asked: "Sir, would you *demonstrate* the usefulness of the study of logic?" Epictetus smiled and replied: "That is my point. How could you, without the study of logic, test whether my demonstration would be valid or not?"

The late Yehoshua Bar-Hillel, who recounts this story, comments that if the audience had signed up to take the course in logic that Epictetus announced, then they must surely have been very disappointed. Epictetus had shown merely that there was a need for a theory that would make it possible to test the validity of arguments in ordinary language, not that he possessed such a theory. Indeed, as Bar-Hillel (1970) emphasizes, logic until very recently was insufficiently powerful to cope with natural language.

To a psychologist, the story of Epictetus suggests a different moral. No matter how much we understand the logic of an inference, we are not thereby vouchsafed any insight into its underlying mental processes. In particular, logic does not determine which inference should be drawn from a given set of premises; there are always a potentially infinite number of valid conclusions that follow from any premises. Likewise, the fact that an inference is valid does not guarantee that the process of thought that yielded it invariably delivers valid conclusions. The impotence of logical knowledge in the face of psychological phenomena is increased still further by the fact that many inferences that people make are invalid. But, what exactly is an inference?

The answer to this question ought to take the form of a theory, but it is nevertheless useful to draw a rough line around the domain to which the theory is intended to apply. This delineation is provided by a working definition. An

inference is a mental process by which new information is obtained from old, either by transforming the old information or by combining two separate pieces of old information. This broad specification encompasses much, but it has the advantage of excluding nothing that might reasonably be taken as an inference. There are many sorts of inference, but I argue that the major distinction to be recognized by any psychological theory is between what I call *explicit* and *implicit* inferences.

Suppose you were to read in the paper: "There was a fault in the signaling circuit. The crash led to the deaths of 10 passengers." You are likely to infer that the passengers were killed in the crash. The text does not make this assertion, and it might even continue: "They were arrested when the airplane crashed and subsequently shot as spies." Plainly, you have jumped to a conclusion based partly on the content of the passage and partly on your general knowledge. You make such inferences automatically, rapidly, almost involuntarily, and often without being consciously aware of what you are doing. Because a valid inference is one for which, if the premises are true, the conclusion must be true, the most important feature of the inference that you drew, to a logician, is its invalidity. To a psychologist, however, the striking feature of your inference is the tacit and automatic way in which it was drawn. It is an example of an *implicit* inference.

Suppose, on the other hand, you read the following two assertions.

Arthur is not related to Bill

Bill is not related to Charles

and then you are asked: Is Arthur related to Charles? You are now faced with a task that requires a moment's thought before you can respond correctly. You need to figure out that Arthur could be related to Charles but need not be. Perhaps surprisingly, not everyone is able to do so: When Bruno Bara and I tested some university students in an experiment that included this inference, one or two of them concluded that Arthur was not related to Charles. Happily, the majority of subjects gave the correct answer. Once again, however, the important psychological feature of this deduction is perhaps not its validity, but the fact that it requires a conscious and cold-blooded effort. You must make a voluntary decision to try to answer the question, and you are aware of trying to make a deduction. It is an example of an *explicit* inference. Let us examine each sort of inference in more detail.

Implicit Inference

Ever since Helmholtz, there have been psychologists who argue that there are unconscious inferences underlying perception, and other psychologists who vehemently deny this claim. It is not my intention to enter into this controversy, but

what seems certain is that implicit inferences, which are unconscious, occur ubiquitously in understanding discourse. Indeed, without an ability to make such inferences written and spoken discourse would not function in its customary way. In order to understand the following discourse it is necessary to make a variety of inferences:

> The pilot put the plane into a spin just before landing on the strip. He just got it out of it in time. Wasn't he lucky?

Every word in the first sentence is ambiguous, and the appropriate meanings can only be recovered by making inferences from linguistic context and general knowledge. To make sense of the second sentence, a number of inferences have to be made in order to determine the referents of the pronouns: The first "it" refers to the plane, and the second "it" refers to the spin. The third sentence is not to be taken as a question, though it is interrogatory in form. An inference from the context establishes that it has the force of an assertion. At the point at which most of these inferences are made, they can seldom be securely established. They are plausible conjectures rather than valid deductions. Many psychologists are accordingly tempted to suppose that a probabilistic inferential mechanism underlies them. However, there is no need to suppose that individuals compute probabilities in determining, say, that a pronoun refers to one entity rather than to another. The mechanism is much more like one that yields a conclusion by default—the conclusion is justified provided that there is no (subsequent) evidence to overrule it: It lacks the guarantee, the mental imprimatur, associated with all explicit deduction.

If the present thesis about the role of implicit inferences is correct, then children must acquire the ability to make such inferences in order to understand discourse. This hypothesis has been borne out by a number of experimental studies. Til Wykes, a former student of mine, showed that young children (about 4½ years old) have considerable difficulty in correctly acting out with glove puppets such pairs of sentences as:

Jane needed Susan's pencil.
She gave it to her.

The task is much easier for them if gender can be used as a cue:

Susan needed John's pencil.
He gave it to her.

In general, the greater the number of pronouns in a sentence, the harder it is for young children to understand it properly. They appear to adopt a syntactically based procedure for assigning referents to pronouns rather than an inferential one. They assume that a pronoun refers to the subject of the previous clause (see Wykes, 1978). In a further study, we discovered that children are poor at making commonsense inferences to work out the meaning of such sentences as:

"The Smiths saw the Rocky Mountains flying to California" (Wykes & Johnson-Laird, 1977). Similarly, children presented with such sentences as, "The man stirred his cup of tea," tend not to infer spontaneously that the man used a *spoon* to stir his tea. In all these cases, it was clear from control studies that the children are able to make relevant inferences. The point is that they do not usually do so as a normal part of understanding discourse.

The ability to make implicit inferences is equally important, of course, for reading. Jane Oakhill, a former student of mine, has shown that an important distinction between good readers and average ones lies precisely in their inferential ability. In one study, Oakhill gave a sample of 168 children (aged 7 to 8 years) a variety of vocabulary and reading tests. She was then able to select two groups matched on vocabulary and phonic skills, but differing considerably in their general reading ability. The two groups of children took part in an experiment that investigated the extent to which they made inferences when *listening* to very simple stories. Each story consisted of three sentences:

The car crashed into the bus.
The bus was near the crossroads.
The car skidded on the ice.

After the children had heard eight such stories, their memory for them was tested. A child who has built up an integrated mental representation of the events in the story might well assume that the sentence, "The car was near the crossroads," had originally occurred in the story. Given the nature of the original events, this inference is extremely plausible. A sentence such as, "The bus skidded on the ice," is much less plausibly inferred, because there is no reason to make this inference in building a representation of the events in the story. The results of the memory test using such sentences showed, as expected, that good readers tended to make more errors based on plausible inferences than did average readers. Good readers, however, performed better than average ones in recognizing the original sentences from the stories and in rejecting implausible inferences. It seems that a really good reader is likely to make implicit inferences in order to build up an integrated representation of a story, whereas an average reader is less likely to do so. Obviously, this study tells us nothing about causal direction. Good readers may be good because they spontaneously make inferences, or they may make such inferences because they are good readers as a result of other factors. But, in a series of additional experiments, Oakhill has so far failed to find any other major distinction in the abilities of her two groups of readers. Their memory spans for digits, words and short sentences, and the size of their vocabularies do not differ significantly.

One addendum to this work is worth noting. The procedure is based on one devised by Paris and his colleagues (e.g., Paris & Carter, 1973), though these investigators were not concerned with differences in reading ability. These studies have been criticized on the grounds that the sentences used in the memory tests

allowed children to detect the new sentences, which had not occurred in the original stories, solely because they contained words not in the original stories. The materials used by Oakhill, as the preceding example shows, were carefully selected so as to obviate this criticism.

If understanding discourse depends on the ability to make implicit inferences, then an important pedagogical task is to inculcate this ability in those who lack it. Unfortunately, we do not know whether the difference in reading ability reflects an inability to make inferences, or merely—as with the children in Wykes's studies—the failure to mobilize them spontaneously. What we do know is that implicit inferences are so automatic that most people are unaware of making them. Like many skills, children must somehow pick up the ability in a wholly tacit way. This characteristic suggests that we must be especially careful that we do not unwittingly interfere with the normal acquisition of this skill if we try to enhance it. The educational task is more akin to trying to promote the development of a child's native tongue than to giving explicit instruction in reading.

Explicit Inference

Logical thinking in daily life is most likely to occur in those explicit inferences that are consciously made in trying to answer questions or to solve problems. The following dialogue illustrates such rational sequences of thought:

1. Does this train go to Ickenham?

   Yes—it's going to Uxbridge, and all trains that go to Uxbridge go to Ickenham.

2. Play a let.

   Why?

   You served out of turn, and the rules of badminton state that any point on which a player serves out of turn is a let.

3. Could the suspect have committed the murder?

   No. The victim was stabbed to death in a cinema during the afternoon showing of *Bambi*. The suspect was traveling to Edinburgh on a train throughout that afternoon.

The last example is certainly a valid inference, because there is no way in which the premises could be true and the conclusion false. Yet there is no existing logic that directly establishes its validity, which depends, of course, on many obvious but unstated premises, such as:

For one person to stab another it is necessary for them to be at the same place at the same time.

There are no cinemas on trains traveling to Edinburgh.

Trains to Edinburgh do not pass through cinemas on the way there.

The inference also depends on the meaning of the terms *during* and *throughout*. It is crucial that the suspect was on the train throughout the afternoon—if he was able to leave it, he might have done the murder. How then does the legal mind—or what the rest of us have to serve the same function—make this particular deduction?

The majority of psychologists have argued that there must be a mental logic that underlies the ability to make valid deductions. The attraction of this hypothesis is that it explains how it is possible for logic to have been invented in the first place. It also leads naturally to a quest for the correct specification of mental logic. The fact that people often perpetrate fallacies is mildly embarrassing to the doctrine, but it is easy—all too easy—to explain invalid inferences away: "I have never found errors which could be unambiguously attributed to faulty reasoning," Mary Henle (1978) remarks characteristically. The trouble is that there appear to be no independent "ground rules" for deciding whether a mistaken inference is the result of a logical error or some other deficiency. Nevertheless, the doctrine of mental logic has led to much research into reasoning, and, not surprisingly, to attempts to enhance children's logical ability by teaching them patterns of valid inference (see Falmagne, 1980). Before one can readily assess the pedagogical implications of such studies, it is important to try to establish the nature of mental logic and whether or not it really exists.

SYSTEMS OF LOGIC AND RULES OF INFERENCE

If there is a mental logic, then its most essential component must consist of a set of rules of inference, or some such schemata enabling conclusions to be drawn from premises. In an orthodox formulation of a logic, a rule of inference is essentially syntactic in nature. It governs uninterpreted expressions, specifying what can be derived from them wholly in terms of their form, not their meaning. The semantic content of the deduction is irrelevant. Indeed, its irrelevance is the essential foundation on which all formal logic rests: The principle of validity can be captured in a wholly general way.

One group of workers in artificial intelligence has capitalized on the content independent aspect of logic and developed programs that operate proof procedures for the predicate calculus. One of the major developments in formal logic during this century is Church's proof that there can be no mechanical decision procedure for this calculus: There can be procedures that will determine sooner or later that an inference is valid, but there can be no procedure that is guaranteed to discover that an inference is invalid. Hence, the quest in mechanical theorem

proving is to cut down the time it takes to discover that an inference is valid (if indeed it is), because as the program grinds away there is no way of knowing whether it will ultimately reveal that the inference is valid or go on computing forever. Programs have been devised that use just a single rule of inference, the so-called "resolution" rule (see e.g., Robinson, 1965):

$$\begin{array}{r} A \text{ or } B \\ \text{not } -A \text{ or } C \\ \hline \therefore B \text{ or } C \end{array}$$

Thus, whenever an assertion and its negation occur in disjunctive premises, they can be deleted to have a disjunction of whatever is left. The uniform proof procedure has the overall pattern of a *reductio ad absurdum*, but in order to apply the resolution rule it is necessary to translate the premises into a special notation in which the quantifiers are eliminated and all the connectives are transformed into disjunctions.

Only if a deduction is valid will the uniform proof procedure ultimately yield a proof. It is accordingly important to cut down the time to prove valid arguments, and the problem is to find the best way of eliminating assertions and their negations; a variety of heuristic procedures have been devised to help to speed up the search (see Robinson, 1979). A uniform theorem prover is undoubtedly intelligent, but it is highly artificial. Indeed, its critics within artificial intelligence have pointed out that it is both inflexible and remote from ordinary reasoning (see e.g., Winograd, 1972). One might add that, because there are doubts about whether natural language can be accommodated within the orthodox predicate calculus, there must also be doubts about such a formalization.

A very different approach to rules of inference is represented by the development of PLANNER and its cognate languages (Hewitt, 1972). Programs written in PLANNER-like languages have a data base that consists of a set of assertions, specified in a predicate-argument format, such as:

(SCIENTIST FRED)
(DRIVER BILL)

The assertion, "Fred is a scientist," is accordingly true with respect to this data base, and PLANNER enables the programmer to implement procedures (1) that will evaluate a sentence with respect to the assertions in the data base and return its truth value, and (2) that will take a sentence and add the corresponding assertion to the data base. However, if the assertion is of the form, "All scientists are drivers," then rather than tamper with the data base—going through it and adding an assertion about each individual who is a scientist to the effect that he or she is a driver, PLANNER allows a form of representation in which such a data base is merely *described* rather than actually established. For example, a procedure known as a consequent theorem can be set up:

```
(CONSEQUENT(x)(DRIVER x)
  GOAL(SCIENTIST x))
```

and in effect added to the data base. What this procedure says is that *x is a driver* is true for any x provided that the goal of showing that *x is a scientist* is achieved—the consequent is true provided that the goal is satisfied. It is just one of the ways in which any general assertion can be represented by a procedure in a PLANNER system.

If the program's goal is to show that *Fred is a driver*, then the preceding procedure can be called if there is no simple assertion to that effect in the data base. If the consequent theorem satisfies its goal, i.e., discovers that there *is* an assertion that *Fred is a scientist*, the desired conclusion follows at once. By representing general assertions in the form of rules of inference, PLANNER allows the programmer to take into account information specific to their content, for example, hints or heuristics about how to achieve a particular inferential goal. Hence, PLANNER-like systems seem more plausible psychologically than uniform theorem provers. However, people do possess certain general inferential abilities, and no studies in artificial intelligence have yet illuminated them.

If human beings possess a mental logic, it is likely to have more than one rule of inference unlike a "resolution" system, and rather fewer rules of inference than a PLANNER system. This consideration renders a system based on the so-called method of "natural deduction" quite plausible, and a number of psychologists have proposed such a theory (see Braine, 1978; Johnson–Laird, 1975; Osherson, 1975). A natural deduction system puts no premium on parsimony and contains inference schemata for each operator and connective in the logic, e.g.:

$$\frac{A \text{ and } B}{\therefore A}$$

$$\frac{A \text{ or } B,\ \text{not } (A)}{\therefore B}$$

There are a number of technical difficulties with any psychological theory based on natural deduction, but they are not insuperable. The major problem, however, is to explain how children could acquire such a system of inferential schemata.

THE ACQUISITION OF MENTAL LOGIC

Any version of the doctrine of mental logic needs to explain how logic gets into the individual human mind. Prior to its acquisition, that mind will not be capable of valid reasoning, and so obviously the event is momentous. In fact, however, it runs the risk of paradox, for how could logic be acquired by someone who

was not already able to reason soundly? Three main answers have been given to this question, but none of them will do.

First, it is said that logic is acquired according to the ordinary processes of learning as delineated by conventional theory. Children are positively reinforced for making valid deductions and not reinforced (or even perhaps punished) for making invalid deductions. Unfortunately, even casual observation shows that there can be few children who are brought up under such a regimen; training in logic is the prerogative of philosophers, not parents. Moreover, the hypothesis begs the question in assuming that parents have themselves somehow acquired the distinction between valid and invalid inferences. A related question begging conjecture is that children are able to infer rules of inference by abstraction from the valid inferences that they encounter in daily life. Hence, children are obliged to be already able to discriminate between validity and invalidity as a precursor to the learning process. But, if they already have the ability to make this distinction, why should they need to learn rules of inference? The trouble with both proposals is that they assume the prior existence of logic in order to account for the acquisition of logic.

Second, it is said that mental logic is inborn. It is part of our innate intellectual equipment, just as there are supposedly genetically endowed constraints on the possible forms of human grammar (see Chomsky, 1965). Although this supposition may well be correct, a disinterested psychologist might suspect that it is a rather convenient way of passing on the problem to biology.

Third, Piagetians argue that logic is neither learned nor innate but constructed. It is the result of actions on the world, the internalization of those actions, and reflection on their organization, in a hierarchical sequence of stages that gradually liberates the individual from direct dependence on the evidence of the senses. Unfortunately, neither Piaget nor his colleagues have ever spelled out the nature of the mechanism guiding this developmental sequence in a sufficiently clear and explicit way to allow it to be either modeled or effectively evaluated. Accounts of the theory use vagueness to mask its potential inadequacies from its proponents (and others).

Perhaps, however, there is no logic in the mind, and perhaps humanity is intrinsically and irredeemably irrational. The follies and horrors of the human condition certainly lend credence to this view. Yet human beings cannot be wholly irrational. Logic could not have been invented by a species incapable of logical thought. Indeed, it was originally invented to help people to think more precisely. What could be more rational than the desire to make valid inferences, an appreciation that the unaided mind was not invariably able to do so, and the invention of a technology for reasoning? Psychology has in the past too often assumed that there is a dichotomy: People either have a mental logic and are rational or else they lack such a logic and are irrational. What has hitherto been unquestioned is that these two alternatives are exhaustive. In fact, there is a third possibility. Human beings do not possess a mental logic, but they are capable

of rational thought. Before I spell out the case for this point of view, I want to introduce the notion of a *mental model*.

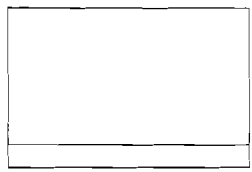
Mental Models

Let us suppose that you are reading the famous story, *Charles Augustus Milverton*, by Conan Doyle (1905). In this story, Sherlock Holmes and Dr. Watson set out to burgle the house of the eponymous Milverton, a blackmailer and "the wickedest man in London." The following sequence of events then occurs:

> We stole up to the silent, gloomy house. A sort of tiled veranda extended along one side of it, lined by several windows and two doors.
>
> "That's his bedroom," Holmes whispered. "This door opens straight into the study. It would suit us best, but it is bolted as well as locked. Come round here. There's a greenhouse which opens into the drawing room."
>
> The place was locked, but Holmes removed a circle of glass and turned the key from the inside. An instant afterwards he had closed the door behind us. The thick warm air of the conservatory took us by the throat. He seized my hand in the darkness and led me swiftly past banks of shrubs which brushed against our faces. Holmes had remarkable powers, carefully cultivated, of seeing in the dark. [!] He opened a door, and we entered a large room in which a cigar had been smoked not long before. He felt his way among the furniture, opened another door, and closed it behind us. Putting out my hand I felt several coats hanging from the wall, and I understood that I was in a passage. We passed along it, and Holmes very gently opened a door upon the right-hand side. Something rushed out at us, and my heart sprang into my mouth, but I could have laughed when I realized that it was the cat. A fire was burning in this new room, and again the air was heavy with tobacco smoke. Holmes entered on tip-toe, and waited for me to follow. We were in Milverton's study, and a doorway at the farther side showed the entrance to his bedroom.
>
> It was a good fire and the room was illuminated by it. At one side of the fireplace was a heavy curtain, which covered the bay window we had seen from outside. On the other side was the door which communicated with the veranda. A desk stood in the centre, with a turning chair of shining red leather. In the corner between a bookcase and the wall, there stood a tall green safe, the firelight flashing back from the polished brass knobs upon its face.

Doubtless, in reading this passage, you were more than usually aware of the role of implicit inferences in comprehension—the windows and door referred to in the second sentence were those of the house, for example, though this fact is not stated explicitly. You probably also noticed that Holmes does *not* make one of his celebrated deductions. In fact, that omission is deliberate on my part

because I want you to try to make a deduction. Here is a simple plan of the house with the veranda running down one side of it:



Which way did Holmes and Watson make their way along it—from left to right or from right to left?

In my experience, about one in 100 people can spontaneously give the right answer for the right reason. However, if you read the passage again with the aim of solving this riddle, it is relatively simple. (The solution, for those who are still perplexed, can be found at the end of this chapter.)

In order to make this inference you must build a mental model of the spatial layout. The fact that you are unlikely to be able to draw the correct conclusion unless you are forewarned suggests that there are at least two sorts of representation for discourse: a relatively superficial representation close to the linguistic form of the discourse and a mental model that is much closer to being a representation of a state of affairs—in this case the plan of a house—than to a set of sentences.

My colleagues and I have investigated this hypothesis about levels of representation in a series of experiments (see Johnson-Laird, 1983). Kannan Mani and I took the idea that readers can construct a mental model of a spatial layout, and examined it in a properly controlled study (Mani & Johnson-Laird, 1982). The subjects heard a series of spatial descriptions such as the following one:

The spoon is to the left of the knife.
The plate is to the right of the knife.
The fork is in front of the spoon.
The cup is in front of the knife.

They then judged whether a diagram such as:

spoon knife plate
fork cup

was consistent or inconsistent with the description. If you think of the diagram as depicting the arrangement of the objects on a table top, then obviously it is consistent with the description. Half the descriptions were determinate as in the example, but the other half were grossly indeterminate. The indeterminacy was introduced merely by changing the last word in the second sentence:

The spoon is to the left of the knife.
The plate is to the right of the spoon.
The fork is in front of the spoon.
The cup is in front of the knife.

This description is indeterminate in that it is consistent with at least the two radically different arrangements shown here:

spoon	knife	plate	spoon	plate	knife
fork	cup		fork		cup

After the subjects had evaluated a whole series of pairs of descriptions and diagrams, which were presented in a random order and each with a different lexical content, they were given an unexpected test of their memory for the descriptions. On each trial they had to rank four alternative descriptions in terms of their resemblance to the actual description that they had been given. The alternatives consisted of the original description, a description that was inferrable from a model of the original description, and two confusion items that described completely different arrangements. The inferrable description for the previous example included the sentence:

The fork is to the left of the cup.

which could be inferred from the layout corresponding to the description (either the determinate or the indeterminate one).

The subjects remembered the gist of the determinate descriptions very much better than that of the indeterminate descriptions. The percentage of trials on which they ranked the original and the inferrable description as closer to the original than the confusion items was 88% for the determinate descriptions, but it was only 58% for the indeterminate descriptions. All 20 subjects conformed to this trend, and there was no effect of whether or not a diagram had been consistent with a description. However, the percentage of trials on which the original description was ranked higher than the inferrable description was 68% for the determinate descriptions, but it was 88% for the indeterminate descriptions. This difference was also highly reliable.

A plausible explanation for the pattern of results is that subjects construct a mental model for the determinate descriptions but abandon such a representation in favor of a superficial linguistic one as soon as they encounter an indeterminacy. Mental models are relatively easy to remember but encode little or nothing of the original sentences on which they are based, and subjects accordingly confuse inferrable descriptions with them. Linguistic representations are relatively hard to remember, but they do encode the linguistic form of the sentences in a description.

Working Memory and Inference

A crucial factor in the construction of a mental model, or indeed of any sort of integrated representation, is the capacity of working memory. The representation must be held there while at the same time the relevant information from the current sentence is extracted and added to it. This problem is not obviated by merely allowing subjects to have the written premises in front of them throughout the task; the integration of premises has to occur in working memory, unless the subjects are allowed to use paper and pencil as an external substitute for it.

Kate Ehrlich and I have demonstrated the importance of working memory in another series of experiments on spatial reasoning (Ehrlich & Johnson-Laird, 1982). In one such experiment, the subjects listened to a verbal description of a spatial layout, and then they attempted to draw a diagram of the corresponding arrangement. The main variable that we manipulated was the continuity of the descriptions. Each description occurred in three different versions (though with different lexical contents for an individual subject). A continuous description such as:

The sugar is on the left of the fork.

The mug is in front of the fork.

The ashtray is on the right of the mug.

allows a mental model of the layout to be built up in a continuous sequence. The first premise corresponds to the arrangement:

1. sugar fork

The second premise allows the mug to be added:

2. sugar fork
   mug

The third premise allows the ashtray to be added so as to complete the layout:

3. sugar fork
   mug ashtray

and the subject can proceed to translate this mental model into an actual diagram.

A discontinuous version of the same description is created by reordering the premises:

The ashtray is on the right of the mug.

The sugar is on the left of the fork.

The mug is in front of the fork.

The first premise yields the arrangement:

1. mug ashtray

But the second premise makes no reference to either of these objects and a subject is obliged to create a separate representation of it:

2a. mug ashtray 2b. sugar fork

It may even be that subjects at this point abandon the attempt to construct a mental model and rely instead on a superficial linguistic representation. Only when the third premise is presented can the representations of the first two premises be integrated:

3. sugar fork
mug ashtray

As we expected, the task was very much harder with the discontinuous descriptions. The continuous descriptions yielded 69% correct diagrams, whereas the discontinuous descriptions yielded only 42% correct diagrams. The effect is not simply a consequence of there being two consecutive sentences that have no referent in common. In a third "semicontinuous" condition, the sentences were presented in the order:

The mug is in front of the fork.

The ashtray is on the right of the mug.

The sugar is on the left of the fork.

Here the third sentence makes no reference to any item in the second sentence, but the task is not reliably harder (60% correct diagrams) than in the continuous case—presumably because the third sentence does refer to an entity already present in the model of the previous premises.

The limitations of working memory play a crucial part in the form of the conclusions that reasoners draw in their own words. This effect was originally apparent in a study of syllogistic reasoning (see Johnson-Laird & Steedman, 1978). For example, with such premises as:

Some of the parents are drivers.

All of the drivers are scientists.

there is an overwhelming bias to draw the conclusion:

Some of the parents are scientists

rather than its equally valid converse:

Some of the scientists are parents.

In general, any syllogism in which the terms are arranged:

A – B

B – C

creates a bias toward subjects drawing a conclusion:

A – C

and any syllogisms in which the terms are arranged:

B – A
C – B

creates a bias toward subjects drawing a conclusion:

C – A

This so-called "figural" effect applies both to valid and invalid conclusions and is singularly reliable. Whenever I have lectured on it, I have invariably illustrated the phenomenon, and audiences at universities as far afield as Milano, Padova, Nijmegen, Amsterdam, Edinburgh, Cambridge, Chicago, San Diego, and New York have universally conformed to it.

For several years I took the view that the figural effect was the result of the way in which syllogistic premises were mentally represented—a matter that is taken up again later—but it now seems much more likely to be a consequence of how information is put together in working memory.

Bruno Bara and I have recently discovered that there is a figural effect in very simple three-term series problems, such as:

Ann is taller than Beryl.

Beryl is shorter than Carol.

Who is tallest?

Studies of such problems have invariably either asked a specific question, as in the example, or else presented a specific conclusion for evaluation. Bara and I, however, presented the premises to subjects and asked them to draw conclusions in their own words (Johnson-Laird & Bara, 1984). Likewise, in order to obviate any differences as a result of the linguistic contrast between such pairs of antonyms as "taller" and "shorter," we chose to use only a single relational term. We found that with premises of the form:

Ann is related to Beryl.

Beryl is related to Carol.

nine of ten subjects drew a conclusion of the form:

Ann is related to Carol.

With premises of the form:

Beryl is related to Ann.

Carol is related to Beryl.

there was no reliable bias either way. In the case of symmetrical problems such as:

Ann is related to Beryl

Carol is related to Beryl

or:

Beryl is related to Ann.

Beryl is related to Carol.

there is a slight bias toward drawing a conclusion in which the end individual in the first premise (Ann) is the subject.

The results of this experiment suggest that when reasoners combine the information presented in premises, they try to form a mental model of the first premise to which they add the information in the second premise. Thus, the premise, "Ann is related to Beryl," yields the following sort of model:

$a \rightarrow b$

They then interpret the second premise and substitute the relation to Carol in place of the middle term.

$a \rightarrow c$

On the plausible assumption that working memory operates according to a "first-in first-out" principle, the resulting model is then translated into the conclusion:

Ann is related to Carol.

The same procedure with a problem of the form:

Beryl is related to Ann.

Carol is related to Beryl.

yields a conclusion of the appropriate bias. The fact that the phenomenon is considerably reduced in this case suggests that the "figure" of the premises is less conducive to such a substitution. The first premise yields a model of the form:

$b \rightarrow a$

where the middle term is first into working memory. The second premise introduces the middle term last. There may be accordingly be a tendency to scan the first model in the opposite direction, and then to make the required substitution.

Constructing a superficial linguistic representation is effortless and automatic for a native speaker of the language. Constructing a mental model requires effort and places a load on working memory. A mental model has a structure that corresponds to a state of affairs rather than to a set of sentences, and this structure

has to be constructed by the reasoner. Mental models may take the form of images in certain cases, but that is not essential: There are grounds for supposing that everyone can construct such models, but many people claim to be bereft of imagery.

How to Reason Validly Without the Use of Logic

If you are told as a matter of fact:

Consultant surgeons in Brighton earn £20,000 per annum

and you subsequently learn that:

Arthur is a consultant surgeon in Brighton,

then you will have little difficulty in inferring that

Arthur earns £20,000 per annum.

The inference is so simple that its underlying mechanism eludes introspection. It may depend on a rule of inference as adherents of mental logic suppose, and I have described a number of systems that have been implemented as computer programs that are also capable of the inference. What I want to outline now is a very different way of making the same inference that relies, not on any rules of inference or inferential schemata, but on mental models.

Let us first consider an overt, though somewhat impractical, way of making the inference. Suppose you were able to gather together in one room all the consultants in Brighton. The first premise asserts that all the *surgeons* among them earn £20,000, and so you hand out to each of them a placard, which they are to carry, bearing the legend, "I earn £20,000 per annum." The second premise asserts that Arthur is a consultant surgeon in Brighton. You now search through the room until you come upon Arthur: He will be carrying a placard that says he earns £20,000 per annum. Thus, you may readily draw the appropriate conclusion. All that is necessary to convert this outlandish procedure into a psychological theory is to suppose that you carry out the entire procedure in your mind. You construct a *mental model* that satisfies the premises and derive the conclusion from an inspection of its contents.

At first sight, this way of making inference may seem too simple to be true. You may feel that a sort of deception has occurred. In order to dispel this feeling, let us consider another example and deal with it in a more abstract way. We suppose that the premises are of the form:

All the A are B.

All the B are C.

Instead of employing a roomful of people, we suppose that the reasoner merely imagines arbitrary numbers of individuals or entities of the appropriate sorts. A mental model of the first premise accordingly contains some arbitrary number of members of the class, A, which we designate thus:

a

a

a

Because the premise asserts that all of them are also members of the class, B, each *a* must be identical to a member of that class:

a = b

a = b

a = b

The premise is entirely consistent with the possibility that there are *b*'s that are not *a*'s—there may be, or there may not be—and this possibility must also be represented in the mental model. We use the notational convention that anything within parentheses denotes a possible individual. Hence, a complete mental model for the premise, "All the A are B," has the form:

a = b

a = b

a = b

(b)

In order to draw an inference, your task is to form an integrated representation of both premises. Such an integration is only possible of course because the same set of individuals is referred to in both of them. A representation of the second premise, "All the B are C," could, by itself, take the form:

b = c

b = c

b = c

(c)

But, obviously, because B refers to the same class in both premises, it should be represented by the same number of individuals, and hence the representations of the two premises can be combined as follows:

a = b = c

a = b = c

a = b = c

(b) = c

(c)

A more plausible maneuver, however, is to operate directly on the model of the first premise and to substitute *c*'s for each *b* within it, and add an optional *c*, as sanctioned by the second premise:

a = c

a = c

a = c

(c)

Alternatively, *a*'s can be substituted for *b*'s in the model of the second premise, yielding the same ultimate result. In any case, the integrated model yields the conclusion "All A are C."

When an inference is made in this way, the reasoner imagines a state of affairs that satisfies the description provided by the premises and then draws a conclusion that is consistent with that state of affairs omitting any reference to those entities referred to in both premises, i.e., the so-called "middle-term" that makes the inference possible.

The example, of course, was chosen with benevolence aforethought. Here is a case where there is more than one way of combining the information in the premises. Suppose they are of the form:

Some A are B

No B are C

The first premise is satisfied by the state of affairs:

a = b

(a) (b)

Likewise, the second premise is satisfied by:

b ≠ c

b ≠ c

though strictly speaking, this notation is a slight oversimplification because there should be a nonidentity between every *b* and *c*. In seeking to form an integrated

model that satisfies both premises, a prudent reasoner ought to consider all the different possibilities. It is easier to discern them from a composite representation formed by sticking the two models directly together rather than by substituting the information in one into the representation of the other. Such a model would have the following form:

a = b ≠ c

(a) (b) ≠ c

Because there is no specified relation between (*a*) and (*b*), it can obviously be either an identity or a nonidentity. Hence there are two possible integrations:

a ≠ c	a ≠ c
(a) ≠ c	(a) = (c)

These two models are consistent with the conclusion:

Some A are not C.

And they are also consistent with its converse:

Some C are not A.

However, the number of possible *a*'s that are not *b*'s is arbitrary, and accordingly the following integrated model is also possible:

a ≠ c
//
(a)

(a) = c

This model plainly refutes the putative conclusion, "Some C are not A," because all the *c*'s are *a*'s. But there is no way in which to make all the *a*'s identical to *c*'s,and it is therefore valid to conclude:

Some A are not C.

This more complicated deduction was drawn entirely without relying on any mental logic, rules of inference, or inferential schemata. There is one fundamental principle that guides it: An inference is valid if there are no counterexamples to it. What reasoners must do in order to be rational is to ensure that there is no way of interpreting the premises that is consistent with a denial of the conclusion. They must try to consider all the different ways in which the information in the premises can be combined in an integrated model, and this task is obviously one in which the meanings of the premises must not be violated. Logic, and particularly the method of natural deduction, can be conceived of as a device for

making this search systematically; human beings, however, very often fail to examine all the possibilities—they have no logic to help them. Moreover, logic alone can never guide the reasoner to a particular conclusion—it always permits an infinite number of alternative valid conclusions from any premises whatsoever. The overwhelming majority of such conclusions are entirely trivial, consisting of such assertions as the mere conjunction or disjunction of the premises, and so human reasoners must obviously be guided by principles entirely outside logic to the particular conclusions that they draw. The heuristic principle that people appear to follow is to build a model that establishes a connection between those items that are referred to only in separate premises.

There are many other different sorts of valid deduction, and I cannot deal with all of them here. What should be reasonably evident, however, is that the same general principle of constructing mental models with a view to finding counterexamples can apply to any sort of deduction. The reader who wishes to see the application of this thesis to other forms of inference is referred to Johnson–Laird (1978, 1980a, 1980b). The present account of reasoning without logic has taken its truth very much for granted. The first example of an inference presented earlier was one in which the premises yield only a single integrated model; the second example was one in which the premises yield three different models. In general, valid conclusions deriving from these sorts of premises require one, two, or three different mental models. We can predict that the greater the number of models to be constructed, the harder the task will be. The results of many experiments overwhelmingly confirm this prediction (see Johnson–Laird, 1983).

Individual Differences in Reasoning Ability

The inferences in the previous section are known to logicians as "syllogisms." People differ considerably in their ability to make such inferences, which have long been used in tests of intelligence. The present theory of inference, which is described in full in Johnson–Laird and Bara (1984), throws considerable light on the source of these individual differences. The first component of importance is whether or not a person is prepared to play the game of making inferences in a laboratory setting. Sylvia Scribner (1977) and her colleagues have shown that people in nonliterate cultures are often not prepared to play this game. The following dialogue illustrates the performance of such a nonparticipant. The subject was given the following problem:

All Kpelle men are rice farmers.

Mr. Smith is not a rice farmer.

Is he a Kpelle man?

The following dialogue then ensued:

S: I don't know the man in person. I have not laid eyes on the man himself.
E: Just think about the statement.
S: If I knew him in person, I can answer that question, but since I do not know him in person, I cannot answer that question.
E: Try and answer from your Kpelle sense.
S: If you know a person, if a question comes up about him you are able to answer. But if you do not know the person, if a question comes up about him it's hard for you to answer.

This dialogue illustrates that the Kpelle subject is not prepared to make inferences about people that he does not know. Yet, at the same time, it also shows that he is quite capable of just such inferences. The claim that underlies his behavior can be put in the following form:

If I do not know an individual, then I cannot draw any conclusions about that individual.

I do not know Mr. Smith.

Therefore I cannot draw any conclusions about Mr. Smith.

Luria (1977) reports very similar findings in a study with nonliterate Uzbekistanian women. As Scribner argues, it seems likely that literacy or schooling, rather than other cultural differences, is the critical variable. In our experiments, we have encountered only one adult subject—a student in an Italian university—who was not prepared to play the game of deductive inference.

In order to form an integrated mental model of premises, a subject must clearly be able to understand them and to know what would count as a state of affairs that would satisfy them. The subject must be able to hold a representation of both premises in working memory so as to combine the information that they contain. This problem is not merely one of remembering the premises, because it is not obviated by having the premises in front of the subject throughout the whole task. It is necessary that both should be mentally encoded simultaneously, or at least that sufficient information from both should be present in working memory to permit the integrated model to be formed. However, these accomplishments are merely the normal ability to understand one's native language, and to form a mental model of discourse. Although people do differ in their verbal competence, particularly in the speed with which they can understand or produce complicated discourse, this source of variation tends to be smaller than other components of deductive inference; there is very little difference in subjects' skill with those syllogisms that require only a single model to be constructed.

Where more than one model is required, then the subjects must appreciate the need to construct them, carry out this process without error, and be able to remember all of them so as to determine what, if anything, they have in common. With those premises that yield a valid conclusion interrelating the items in the separate premises, the single biggest source of individual differences is a subject's

capacity to cope with two or three alternative models. With premises that do not permit a valid inference to be drawn interrelating the end terms, the situation is more complicated. There are some individuals who are prone to responding "No valid conclusion," whenever the going gets tough, i.e., whenever it is possible to form more than one model. They are right for the wrong reasons with these problems; they are wrong, of course, with premises that yield valid conclusions.

Conclusions

I have argued that there is an important psychological distinction between implicit and explicit inferences. Implicit inferences occur as an automatic part of the comprehension of discourse, whereas explicit inferences are consciously made in attempts to answer questions or solve problems. Implicit inferences tend to occur as an aid to constructing a representation of discourse, and their conclusions are plausible rather than valid. Explicit inferences include those deductions that are intended to be valid. Contrary to the tradition that they depend on a mental logic, the theory presented here argues that they depend on two basic skills: (1) the ability to construct mental models of situations described in sentences, which is a process that occurs in much of the ordinary comprehension of discourse; and (2) the ability to construct and to evaluate alternative mental models of the same premises in order to determine whether or not there are any counterexamples to a putative conclusion. A major cause of difficulty in deduction is indeed the need to consider alternative models within working memory.

Logical thinking does occur in daily life, but the errors that occur in the laboratory suggest that ordinary individuals do not possess a mental logic. If my thesis that errors arise largely as a consequence of the limitations of working memory, then there is perhaps little that can be done pedagogically to enhance logical skill. Yet one should not be too pessimistic. The simple experience of inferential tasks without feedback on the correctness or incorrectness of performance can lead to a significant improvement in performance (see Johnson-Laird & Steedman, 1978). The teaching of logic may likewise effect an improvement in performance or at least suggest the use of overt techniques to relieve the load on working memory. In general, however, the techniques of logic are too complicated to have an immediate practical application, and the standard logical calculi are remote from ordinary language. There is one as yet untested device that may prove to be effective. It would be a simple matter to train people to use paper and pencil in building overt models of premises and to teach them to try to search more exhaustively for counterexamples to putative conclusions. Sherlock Holmes, you may recall, asserted that: "When you have eliminated the impossible, whatever remains, however improbable, must be the truth." My aim has been to argue that logic is a consequence, not a cause, of our unsystematic

but happy ability to search for counterexamples. The more overt we can make this task, the more likely we are to succeed in it.

The Solution to the Sherlock Holmes Riddle

The solution to the Sherlock Holmes riddle is that Watson and he must have gone along the veranda from right to left, as it is shown in the plan. Having entered the house from near one end of the veranda and passed from room to room, they turned *right* from the passageway into Milverton's study with its door that opened directly onto the veranda.

ACKNOWLEDGMENTS

Many individuals helped in this research, and I am particularly indebted to Bruno Bara, Kate Ehrlich, Alan Garnham, Dave Haw, Jane Oakhill, Patrizia Tabossi, for some fruitful collaborations, and to Steve Isard, Christopher Longuet–Higgins, and Stuart Sutherland, for much encouragement and advice.

REFERENCES

Bar-Hillel, Y. (1970). Argumentation in pragmatic languages. In Y. Bar-Hillel, *Aspects of language*. Amsterdam: North–Holland.

Braine, M. D. S. (1978). On the relation between the natural logic of reasoning and standard logic. *Psychological Review*, *85*, 1–21.

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.

Conan Doyle, Sir Arthur. (1905). *The return of Sherlock Holmes*. London: Murray.

Ehrlich, K., & Johnson-Laird, P. N. (1982). Spatial descriptions and referential continuity. *Journal of Verbal Learning and Verbal Behavior, 21*, 296–306.

Falmagne, R. J. (1980). The development of logical competence: A psycholinguistic perspective. In R. H. Kluwe & M. Spada (Eds.), *Developmental models of thinking*. New York: Academic Press.

Henle, M. (1978). Foreword for R. Revlin & R. E. Mayer (Eds.), *Human reasoning*. Washington, DC: Winston.

Hewitt, C. (1972). *Description and theoretical analysis of PLANNER* (MIT AI Laboratory Rep. MIT-AI-258). Cambridge, MA: MIT, Artificial Intelligence Laboratory.

Johnson-Laird, P. N. (1975). Models of deduction. In R. J. Falmagne (Ed.), *Reasoning: Representation and process in children and adults*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Johnson-Laird, P. N. (1978). The meaning of modality. *Cognitive Science*, 2, 17–26.

Johnson-Laird, P. N. (1980a). Mental models in cognitive science. *Cognitive Science, 4*, 71–115.

Johnson–Laird, P. N. (1980b, June). *Propositional representation. Procedural semantics → mental model*. Paper presented at the Royaumont Conference on Cognitive Psychology, Paris. Reprinted in J. Mehler, E. C. T. Walker, & M. Garrett (Eds.), *Perspectives on mental representation*. Hillsdale, NJ: Lawrence Erlbaum Associates, 1982.

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

Johnson–Laird, P.N., & Bara, B. (1984). Syllogistic inference. *Cognition, 16*, 1–61.

Johnson-Laird, P. N., & Steedman, M. J. (1978). The psychology of syllogisms. *Cognitive Psychology, 10*, 64–99.

Luria, A. R. (1977). *The social history of cognition*. Cambridge, MA: Harvard University Press.

Mani, K., & Johnson-Laird, P. N. (1982). The mental representation of spatial descriptions. *Memory & Cognition, 10*, 181–187.

Osherson, D. (1975). Logic and models of logical thinking. In R. J. Falmagne (Ed.), *Reasoning: Representation and process in children and adults*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Paris, S. G., & Carter, A. Y. (1973). Semantics and constructive aspects of sentence memory in children. *Developmental Psychology, 9*, 109–113.

Robinson, J. A. (1965). A machine-oriented logic based on the resolution principle. *Journal of the Association for Computing Machinery, 12*, 23–41.

Robinson, J. A. (1979). *Logic, form and function: Mechanization of deductive reasoning*. Edinburgh: Edinburgh University Press.

Scribner, S. (1977). Modes of thinking and ways of speaking: Culture and logic reconsidered. In P. N. Johnson–Laird & P. C. Wason, *Thinking: Readings in cognitive science*. Cambridge: Cambridge University Press.

Winograd, T. (1972). *Understanding natural language*. New York: Academic Press.

Wykes, T. (1978). *Inference and children's comprehension of prose*. Unpublished doctoral thesis, University of Sussex.

Wykes, T., & Johnson–Laird, P. N. (1977). How do children learn the meanings of verbs? *Nature, 268*, 326–327.