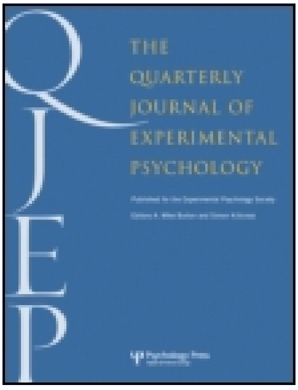

The Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/pqja20

Ninth Bartlett Memorial lecture. Thinking as a skill

P. N. Johnson-Laird [a]

[a] Centre for Research on Perception and Cognition, Laboratory of Experimental Psychology, University of Sussex, Brighton BN1 9QG, England

Version of record first published: 29 May 2007

To cite this article: P. N. Johnson-Laird (1982): Ninth Bartlett Memorial lecture. Thinking as a skill, The Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology, 34:1, 1-29

To link to this article: http://dx.doi.org/10.1080/14640748208400855

Quarterly Journal of Experimental Psychology (1982) 34A, 1–29

NINTH BARTLETT MEMORIAL LECTURE. THINKING AS A SKILL

P. N. JOHNSON-LAIRD

Centre for Research on Perception and Cognition, Laboratory of Experimental Psychology, University of Sussex, Brighton BN1 9QG, England

There are two conflicting views about the nature of thought: it is invariably rational or invariably irrational. Bartlett argued that thinking is a high level skill, and this idea suggests an obvious third possibility: thought is sometimes rational and sometimes irrational. This view is defended in the present paper, which argues that the doctrine of logical infallibility is either falsified by the results of some experiments on syllogistic reasoning or else empirically vacuous. There is no need to postulate a mental logic of the sort that Piaget and others have proposed. The rapid implicit inferences of daily life depend on the ability to interpret sentences by constructing mental models of the states of affairs that they describe. Deliberate deductions depend on the further ability to search for alternative models that violate putative conclusions. All that you need to know to assess validity is the fundamental semantic principle of deduction: an inference is valid if, and only if, its conclusion is true in every situation in which its premises are true and there is no way of interpreting the premises so as to render the conclusion false. This principle guides the construction of all logics though it is not explicitly stated in any of them. The paper concludes by examining the ways in which people differ in their ability to reason, the practical need to improve this ability, and some of the pedagogical implications of the present studies.

Introduction

Sir Frederic Bartlett argued that thinking is a skill that evolved from bodily behaviour, and he studied it from this standpoint for many years. In the book in which he reported his investigations, he suggested that thinking begins with environmental information which is incomplete or fragmentary, and develops in a series of interconnected steps (and sometimes intuitive leaps) which eventually reach some terminus. He wrote: "we should be content to regard thinking as an extension of evidence, in line with the evidence and in such a manner as to fill up gaps in the evidence" (Bartlett, 1958, p. 20). He distinguished three kinds of gap-filling processes, and he claimed that all thinking depends on one or more of them:

> In the first the gap is filled by interpolation, in the second by extrapolation, and the third requires that the evidence given should be looked at from a special, and often an unusual point of view, and that it should be recomposed and reinterpreted to achieve a desired issue (Bartlett, 1958, p. 22).

What could thinking be if not a skill? One's first reaction is that there is no other

0272–4987/82/010001+29 $03.00/0

possibility. But, perhaps thinking might be a set of mental processes that are inborn rather than learned, that apply universally to any cognitive content, and that are exercised wholly without error. This view seems so contrary to common sense that, as you might suspect, it is held by some philosophers (and some psychologists). Thus, Spinoza's metaphysical dread of illogicality led him to the view that what seemed like an invalid inference was merely a valid inference based on other premises. This doctrine has surfaced more recently in Mary Henle's (1962) explanation of apparent errors of reasoning in terms of the forgetting or reinterpretation of premises, the importation of irrelevant information, and other such processes that lead reasoners to argue from a different set of premises than those explicitly presented to them. From this position it is a short step to the view, which Henle espouses, that the underlying competence of ordinary individuals untutored in logic cannot be at fault, errors may occur in performance, but "in all such cases some malfunction of an information processing mechanism has to be inferred" (Cohen, 1981). And this doctrine suggests, in turn, that there is an innate mental logic and indeed that the mind is entirely furnished with innate concepts (Fodor, 1980). Such claims are tendentious with the best of intentions. Their authors are impressed by the fact that human beings invented logic, mathematics, and the very concept of rationality. It follows, they believe, that people must in principle think logically; the laws of thought are indeed the laws of logic.

Curiously, a number of recent investigators have advanced exactly the opposite argument. They have proposed theories of deductive reasoning that render people inherently irrational (e.g. Erickson, 1974; Revlis, 1975; Evans, 1977, 1980; Guyote and Sternberg, 1978). Even if an individual draws a conclusion that happens to be valid, the underlying thought process fails to be logical since these theories preclude a complete examination of the consequences of premises. If you contemplate the follies and fallacies of the human condition, then you might well conclude that these theories are essentially correct. Indeed, if you are of a Romantic cast of mind, you might revel in irrationality like Walt Whitman, who delighted in his capacity for self-contradiction, or you might fear rationality like Tolstoy, who wrote that if human life were ever controlled by reason, then all possibility of spontaneity would be annihilated.

Such views inflate the importance of logic. A valid deduction, of course, is one that follows necessarily from the premises: if the premises are true, then the conclusion must be true. But, it is entirely possible to argue validly from false premises and to make valid deductions that proceed from alternative and perhaps incompatible points of view—as Sydney Smith observed of two women shouting at each other from houses on opposite sides of the street: "they'll never agree, they're arguing from different premises". Moreover, any set of premises yields an infinite number of valid conclusions. Most of them are wholly trivial, such as a conjunction of all the premises, or a disjunction of all of them. Logic provides only a set of procedures for testing whether a given conclusion follows validly from a set of premises; it does not tell you which particular conclusion you should draw. Hence, when you draw a specific valid conclusion, you must be guided by more than logic: you must follow certain other principles that lead to a conclusion that is not wholly trivial.

There is, in short, a controversy about human reasoning. Some theorists, principally philosophers, argue that we are invariably rational; other theorists, principally psychologists, argue that we are invariably irrational. But these contrary points of view are not exhaustive. There remains a third possibility, which is implicit in Bartlett's writings. If thinking is a skill, then its exercise will vary in expertise: human beings may well be rational in some circumstances, but not in others. These three alternatives: invariable rationality, invariable irrationality, and variable rationality, exhaust the universe of possibilities; my task in this paper will be to try to decide amongst them.

Implicit and explicit inference

For most of us, the paragon of deductive ability is Sherlock Holmes, who was wont to make such remarks as: "Beyond the obvious facts that he has at some time done manual labour, that he takes snuff, that he is a Freemason, that he has been in China, and that he has done considerable amount of writing lately, I can deduce nothing else." Here is a rather longer extract from a Holmes story, "Charles Augustus Milverton," that I would like you to read. All that you need to know is that Milverton is the wickedest man in London, a blackmailer, and that Holmes and Watson are about to burgle his house to recover some compromising letters:

> With our black silk face-coverings, which turned us into two of the most truculent figures in London, 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, and we should make too much noise getting in. 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, and we had become felons in the eyes of the law. The thick warm air of the conservatory and the rich choking fragrance of the exotic plants 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.(!) Still holding my hand in one of his, he opened a door, and I was vaguely conscious that we had 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 realised 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 tiptoe, waited for me to follow, and then very gently closed the door. We were in Milverton's study, and a portiere at the farther side showed the entrance to his bedroom.
>
> It was a good fire, and the room was illuminated by it. Near the door I saw the gleam of an electric switch, but it was unnecessary, even if it had been safe, to turn it

on. At one side of the fireplace was a heavy curtain which covered the bay window that we had seen from outside. On the other side was the door which communicated with the veranda.

You may have noticed that Holmes does not make a deduction in this extract, and to compensate for this uncharacteristic lapse, I want you to attempt to make one. Imagine the plan of the house as a square with the veranda running along its lower side. The question is: which way did Holmes and Watson walk along the veranda —from right to left or from left to right?

In my experience very few people are able to make this inference unless they are asked the question before they read the passage, in which case the answer is relatively easy to work out. The crucial evidence occurs, as Bartlett would have said, in disguise. (For those readers who are still perplexed, the answer can be found at the end of this paper.) Yet, it would be a mistake to think, if you were unable to answer the question, that you made no inferences while you were reading the passage. I estimate that you drew about 40 inferences during the process of understanding it. Consider, for example, the way in which you understood the sentences:

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 . . .

You inferred that "the place" refers to the greenhouse, not the drawing-room, and that its door was locked. You inferred that Holmes removed a circle of glass from a pane in the door, that he put his hand through the resulting hole in the pane, and turned the key in the lock of the door to unlock it. And you inferred that Holmes opened the door, and that the two men entered the greenhouse. These inferences are so obvious that you may feel that they hardly merit the term. Yet they are inferences none the less, and they play a crucial role in enabling you to build up an integrated representation of the passage. They depend on general knowledge, and if like the Martian in Craig Raine's (1979) poem you lacked that general knowledge, then you would be unable to make the inferences, and the passage would be incomprehensible to you. It was only when workers in Artificial Intelligence attempted to program computers to understand texts in natural language that the ubiquity of these inferential steps was discovered. The computer is the best Martian of all.

What this passage from Conan Doyle has revealed to us is the existence of two sorts of thinking. There are explicit inferences that require a deliberate and conscious effort, and there are implicit inferences that proceed so smoothly and automatically that one is not normally aware of making them. Bartlett would have said that implicit inferences are intuitive leaps.

There is another important difference between implicit and explicit inferences apart from the conscious-unconscious split between them. Generally, when an implicit inference is made, it goes beyond the literal evidence. For example, when you read in the Sherlock Holmes story: "An instant afterwards he had closed the door behind us . . ." you inferred that Holmes and Watson had entered the house. But the story might have continued: "for at the last moment Holmes had turned us away from the house and from the only criminal act that I had ever known him contemplate." At the point at which these interpolations occur they can seldom be definitive: they are plausible inferences based on general knowledge rather than valid deductions. Some psychologists are accordingly tempted to suppose that a probabilistic mechanism underlies them. However, there is no reason to suppose that individuals compute probabilities in determining, say, the reference of a pronoun. The mechanism is more like one that yields a conclusion by default. The conclusion is justified provided that there is no (subsequent) evidence to overrule it. The inference lacks the mental imprimatur of an explicit deduction in which an attempt is made to test the validity of a conclusion.

The ability to make appropriate implicit inferences is a skill that children must acquire; without it, they will be unable to construct an integrated representation of discourse. This claim has been corroborated by a number of experimental studies. Til Wykes, a former student of mine, has shown that young children of about 4 to 5 years of age 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. 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 same entity as the subject of the previous clause (see Wykes, 1978). In a further joint study, we discovered that children are poor at making inferences based on general knowledge in order to work out the meaning of such sentences as: "The Smiths saw the Rocky mountains flying to California" (Wykes and Johnson-Laird, 1977). Similarly, children presented with a sentence such 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 the relevant inferences if they are explicitly asked to do so. The point is that they do not readily make implicit inferences as a normal part of understanding discourse. Adults, however, as my colleague Alan Garnham (1981) has shown, make such inferences even when they play no essential part in the construction of an integrated representation.

Skill in making implicit inferences is equally important in reading. My student, Jane Oakhill, has shown that an important distinction between excellent and average readers lies precisely in their ability to make such inferences. In one study, Oakhill (1982) 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 in their ability to understand what they read. In order to eliminate the effects of other possible differences in reading ability, the two groups of children were asked to listen to accounts of simple episodes. Each account consisted of three sentences, such as:

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 passages, their memory for them was tested. Those who made implicit inferences in order to build up an integrated representation of an episode should assume that the sentence:

The car was near the crossroads

occurred in the original passage. Given the nature of the episode, it is extremely plausible that the car was near the crossroads, since it crashed into the bus, and the bus was near the crossroads. The assertion:

The bus skidded on the ice

is much less plausibly inferred, since there is no reason to draw this conclusion in building a representation of the events in the passage. The results of the memory test using such sentences showed, as expected, that the good readers tended to make more recognition errors based on plausible inferences than did the average readers. The good readers, however, performed better than the average ones in recognising the original sentences from the passages, and in rejecting the implausible inferences. It seems that excellent readers are likely to make implicit inferences in order to build up an integrated representation of a story, whereas average readers are 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. However, in a series of additional studies, Oakhill has so far failed to isolate any other major distinction between her two groups of readers.

The doctrine of mental logic

Suppose that you are sitting on a tube train going to Uxbridge, and someone asks you, "Does this train go to Ickenham?" You look at the map and discover that every train that goes to Uxbridge goes to Ickenham, and so you reply, "Yes." At the heart of what you did is a simple valid deduction of the form:

This is an x

Every x is a y

Therefore, This is a y.

Of course, you had to know how to obtain the information about the train's route, which in turn probably required an inferential process. Similarly, you may have considered whether or not the train was likely to rush through the station at

Ickenham without stopping, an event which, though not strictly bearing on the accuracy of your answer, would be likely to aggrieve your questioner. Nevertheless, your conclusion hinges on a deductive inference that has the abstract form shown above.

Ever since Aristotle, logicians have been sensitive to the form of inferences, and they have proposed formal rules like the one above, in order to specify the set of valid inferences. In psychology, there is a tradition culminating in the work of Jean Piaget which holds that there is a mental logic. The mind uses formal rules of inference, which are not consciously accessible, to guide the process of valid deduction. In your inference about the tube train, you could have used such a rule as a kind of abstract template, and, by filling in the specific values of the variables, you could have made the deduction. In Piagetian lore, of course, a complete logic has been acquired once a child has attained the level of formal operations.

When I first began to study thinking, I too subscribed to the doctrine of mental logic, but over the years my doubts about it have gradually grown. They have been fostered by three main problems.

First, there are the problematic phenomena that occur in a well-known task invented by Peter Wason, which we jointly investigated (see Wason and Johnson-Laird, 1972). The task seems simple. The experimenter lays out four cards in front of a subject, displaying the following symbols:

E K 4 7

The subject already knows that each card has a number on one side and a letter on the other side. The experimenter presents the following generalisation:

If a card has a vowel on one side then it has an even number on the other side.

The subject's task is to select those cards that need to be turned over to find out whether the rule is true or false. The order of turning the cards over is not at issue: each card must be considered on its own merits alone.

Although the problem is easy to understand, it is hard to solve. Table I summarises the results of four such experiments, which for the sake of simplicity I have

TABLE I

The frequencies of initial selections of cards to test a rule of the form: if a card has a vowel on one side then it has an even number on the other side. The frequencies are pooled from four experiments (see Wason and Johnson-Laird, 1972)

E and 4	59
E	42
E, 4, and 7	9
E and 7	5
Others	13
Total	128

presented as though they all involved the generalisation above. Nearly all of the 128 subjects appreciated the need to turn over the card bearing the vowel: if it has an even number on its other side then the generalisation is unscathed, but if it has an odd number on its other side then the generalisation is plainly false. Likewise, most subjects realised that there was no need to select the card bearing the consonant, since the generalisation has no implications for such cards. Some subjects choose the card bearing the even number; other subjects do not. If this card is turned over to reveal a vowel, then the generalisation is unscathed; if it is turned over to reveal a consonant, then it would seem to be equally unscathed. Hence it is not really necessary to choose this card, but its selection is a venial sin of commission. The serious error arises with the card bearing the odd number. Very few subjects elect to turn it over, and this sin of omission is puzzling because if the card has a vowel on its other side then the generalisation is blatantly false. In other words, the reason for selecting the card with the odd number on it is precisely the same as the reason for selecting the card with the vowel on it: both might reveal the combination of a vowel with an odd number, and thus refute the generalisation.

Quite why subjects fail to make the correct selection remains something of a mystery. A number of factors appear to be at work, including an uncertainty about whether the generalisation implies its converse, a tendency to consider properly only those cards bearing values explicitly referred to in the generalisation, and a propensity towards seeking confirmation rather than disconfirmation. The task is sufficiently complicated to have engendered a sizeable body of studies with results that are not wholly consistent. What is clear, however, is that the introduction of a more realistic content can lead to a vastly improved performance. With a generalisation such as:

Every time I go to Manchester I travel by train

and a set of cards representing destinations and modes of transport:

Manchester Leeds Train Car

many more subjects see the need to select the card signifying the journey made by car, because if its destination is Manchester then the generalisation is false (Wason and Shapiro, 1971). But transfer from realistic conditions to abstract ones is minimal. Subjects in another experiment (see Johnson-Laird, Legrenzi and Legrenzi, 1972) were presented with four envelopes lying on a table and asked to select those that they wished to turn over in order to test whether or not they conformed to the rule:

If an envelope is sealed then it has a 50 lira stamp on it.

The subjects readily appreciated the need to turn over the envelope bearing only a 40 lira stamp on it, but the insight did not transfer to the standard version of the selection task. Each subject carried out two trials in both conditions, in an alternating order, and the results are presented in Table II. In fact, 22 out of the 24 subjects made more insightful selections in the realistic conditions, and the two remaining subjects showed equal insight in the two conditions.

Once more, it remains unclear quite what the critical difference is between the realistic and the unrealistic materials. But, the effects on performance certainly

TABLE II

The numbers of subjects (n=24) making a correct selection on both trials, one trial, and neither trial, in testing a conditional rule presented in a realistic and an abstract form (from Johnson-Laird, Legrenzi, and Legrenzi 1972)

	Realistic form	Abstract form
Both trials correct	17	0
One trial correct	5	7
Neither trial correct	2	17

cast doubt on the doctrine of a mental logic, which should be indifferent to content. Faced with a complex situation:

> . . . the subject will ask himself two kinds of questions: (a) whether fact x implies fact y . . . To verify it, he will look in this case to see whether or not there is a counter-example x and non-y. (b) He will also ask whether, on the contrary, it is y which implies x . . . (Piaget, in Beth and Piaget, 1966, p. 181).

The subjects in the card-turning task do, indeed, search for counterexamples, but their search is more likely to be complete with realistic materials, a phenomenon that is difficult to reconcile with the notion that it is directed by formal rules of inference, since, by definition they are neutral with respect to content.

The second source of my doubts about mental logic is the question of its acquisition. It would seem that in order to learn a logic you might have to be able to reason validly, but if you can reason validly then you might not need a logic, and certainly what would require an explanation would be the origin of your logical ability. In fact, several conjectures have been made in the literature about how the mind acquires its logic, but none of them is completely satisfactory.

Some theorists such as Falmagne (1980) have drawn a parallel between learning logic and learning language. Children encounter valid deductions in verbal guise and, it is said, they abstract rules of inference from them in the same way that they learn grammatical rules. Adults, alas, are not noted for sustained public demonstrations of logical thinking, and so this conjecture presumes that children can tell the difference between valid and invalid inferences, or else begs the question by assuming that adults have learned the difference. Although such teaching procedures may be useful in extending logical competence to new patterns of inference, they can hardly account for its original acquisition.

Piaget attempted to explain the development of logical thinking without relying on the conventional principles of learning theory. He argued that children construct logic by internalising their own actions and by reflecting upon them. The mastery of logical thinking ultimately grows from the mental operations created by this reflective process. Unfortunately, Piaget never described this theory in a form that is completely explicit and at a level of detail that would allow it to be modelled by a computer program. The vagueness of his account masked its inadequacies from Piaget himself and later proponents; the effort required to understand it is so great that to succeed exhausts one's critical faculties. Consider just one difficulty:

if thought is internalised action, then what is it that controls such action in the first place? The answer can hardly be simple reflex arcs, since there is no way in which internalising reflexes and their effects can give rise to the notion of truth or to the formal rules of the propositional calculus. Moreover, what is the underlying mechanism for this mysterious process of internalisation? An ability to internalise events might turn out to be nothing else than the ability to think. Hence, if one were to maintain, contrary to Piaget, that action is externalised thought, how in principle could the issue be decided? In making these highly critical comments about Piaget's theory, I do not wish to impugn his genius for asking the right questions or his inventiveness as an experimenter. Thinking may be a form of internalised action, but the nature of this claim remains to be clarified.

Faced with the difficulty of explaining the learning of logical competence, there is a natural temptation to suppose that it is inborn (Fodor, 1980), just as the principles of universal grammar are supposedly innate. The trouble with this view is that the best argument in its favour is the current failure of other approaches to the problem: if no one has succeeded in explaining how an ability could be learned, then perforce it must be inborn. Positive arguments that the ability to reason is innate are as hard to come by as positive arguments that it arises from divine intervention.

The third and severest problem for mental logic is that people make mistakes. They draw invalid inferences that should not occur if deduction is guided by rules of inference. The main response to this discrepancy is an heroic denial of the phenomenon. Mary Henle (1978) bravely declares:

> I have never found errors which could unambiguously be attributed to faulty reasoning. If they are found under clear conditions, I will be forced to a drastic revision of my view of the relation of logic to thinking.

As I have already mentioned, she suggests that mistakes arise because people misunderstand or forget premises, and because they import additional and unwarranted premises into their reasoning. They fail to stick to the original logical problem. Even with the most charitable interpretation of my own and others' errors in inferential performance, I believe that this defence is mistaken. However, there are no generally agreed criteria by which to make an independent assessment of whether an error violates logic. It is easy, all too easy, to explain errors away. My claim is accordingly that the notion that logical errors never occur is either false or else lacking in empirical content. In order to demonstrate errors in reasoning under clear conditions, let us consider the process of syllogistic inference.

Experiments with syllogisms

Imagine that there is a room full of people, including some scientists, parents, and drivers, and that you are told the following facts about them:

Some of the scientists are parents

All of the parents are drivers.

What conclusion would you draw? The reader may care to pause for a moment to consider what follows from the premises. The majority of subjects given this problem in the laboratory draw the valid conclusion:

Some of the scientists are drivers.

Only a few subjects draw the equally converse conclusion:

Some of the drivers are scientists.

I will call this asymmetry in the responses, the "figural bias," for reasons that will become clear presently. The point to be emphasised is that this syllogism is very easy: nearly every subject makes the correct response with a latency that is typically around 4 to 5 s. Here is a very much harder syllogism:

All the beekeepers are artists

None of the chemists are beekeepers.

Once again, the reader should imagine a roomful of people including beekeepers, artists, and chemists, and attempt to decide what, if anything, follows from the premises. One sample of highly intelligent university students produced the following responses to this problem:

None of the chemists are artists	60% of the subjects
None of the artists are chemists	10% of the subjects
"There's no valid conclusion"	20% of the subjects
Some of the chemists are not artists	10% of the subjects

Not a single subject drew the correct conclusion: "Some of the artists are not chemists" (see Johnson-Laird and Steedman, 1978). The result is entirely typical of performance with this sort of syllogism, and there are no grounds for supposing that the subjects forgot a premise or distorted a meaning, since they had both premises in front of them throughout their attempt to make an inference. Indeed, I see no possibility of explaining away the mistakes that does not also empty the doctrine of logical infallibility of all its empirical content. One final defence might be to claim that:

Syllogisms are highly artificial problems

Psychologists should not study artificial problems

Therefore, pyschologists should not study syllogisms.

In fact, as Kate Ehrlich has shown in an unpublished study, exactly the same phenomena occur if the formal dress of the syllogism is abandoned in favour of a more naturalistic presentation of premises with the same underlying logic. There appears to be no option but to conclude that individuals are capable of thinking logically in some cases but make genuine deductive errors in other cases. It should also be noted that there are marked differences from one individual to another in inferential skill. I therefore propose to abandon the doctrine of mental logic in favour of an alternative approach.

Reasoning with mental models

The theory of reasoning which I wish to propose is based on the concept of a mental model, a notion that was originally introduced by Bartlett's protegé, Kenneth Craik (1943). What is assumed here is that a knowledge of the language enables speakers to construct a model of the state of affairs corresponding to a description: connected discourse can be mentally represented, not only in a linguistic way, but also in a form that is similar to a model based on perceiving or imagining the events instead of merely reading or hearing about them (Johnson-Laird, 1970). The thesis of a "procedural semantics," which derives from work in artificial intelligence (Woods, 1967; Davies and Isard, 1972; Longuet-Higgins, 1972), has inspired theoretical work on the nature of the mental lexicon (Miller and Johnson-Laird, 1976); and it is natural to assume that such procedures are used in the construction, manipulation, and interrogation of mental models (Johnson-Laird, 1980, 1981). There is also experimental evidence to support the hypothesis that individuals can represent discourse both in a superficial linguistic format and in in the form of mental models (Mani and Johnson-Laird, 1982). The fundamental principle of the present theory is that reasoning consists in the construction of mental models on the basis of the premises, and the search for alternative models that might render putative conclusions false. The theory is based on six principal assumptions.

(1) Reasoners interpret premises by constructing an integrated mental model of them. This process, of course, is not peculiar to reasoning but can occur whenever individuals interpret coherent discourse. However, the assumption leads to an important but subtle claim: the logical properties of an expression are not directly represented in the mind (except perhaps in the case of logicians) but emerge naturally as a consequence of the use of the expression in the construction and search processes. This point should become clearer by considering what the theory has to say about the two example problems (the easy syllogism and the hard syllogism). In the case of the first premise of the easy syllogism:

Some of the scientists are parents

the reasoners' knowledge of the language enables them to imagine some arbitrary number of scientists:

scientist
scientist
scientist

and to mentally tag them in some way to indicate that some of them are identical to parents:

scientist=parent
scientist=parent
(scientist) (parent)

The parenthetical items indicate that there may be scientists who are not parents, and parents who are not scientists. The information from the second premise, "All the parents are drivers," can then be directly added to this model to yield an integrated representation of both premises:

scientist= parent=driver

scientist= parent=driver

(scientist) (parent=driver)

(driver)

where again there is an arbitrary number of parenthetical items representing the possibility of drivers who are not parents.

(2) The conclusion that is drawn from a mental model is based on novelty and parsimony: it interrelates those items that have not been explicitly related in the premises, i.e. that have not been directly used in setting up the model. A valid deduction adds nothing to the semantic content of the premises, but reasoners are not immediately aware of the consequences of that content. Hence, they draw conclusions that bring out these new relations but refrain from restating the premises, which can be taken for granted.

(3) The order in which information is expressed in a conclusion follows the principle that working memory operates on a "first-in first-out" basis. It is easier, for example, to recall the premises in the order in which they were presented than in the opposite order. Likewise, it is easier to formulate a conclusion in which the terms occur in the order in which they entered working memory.

The consequences of assumption (2) and (3) lead to an overwhelming bias for the following conclusion to the easy syllogism:

Some of the scientists are drivers

where the two terms that were not related in the premises are now interrelated in the order in which they entered the mental model.

An inference depends on forming an integrated model of the premises. The arrangement, or "figure", of the terms in a syllogism may readily permit this process of integration. With premises in the figure:

A – B

B – C

the two occurrences of the middle term, B, are temporally adjacent, and it is a straightforward matter to construct a mental model of the first premise and then to integrate within it the information from the second premise. The conclusion will then take the form:

A – C

according to the previous assumptions.

(4) If the premises are in a figure that does not permit an immediate integration, then additional operations have to be carried out to make it possible. With premises in the figure:

B – A

C – B

the occurrences of the middle term are separated in time, and cannot be immediately integrated. A natural way of proceeding is to construct a model of the second premise, C – B, renew the interpretation of the first premise, B – A, and then add the information it contains to the model of the second premise. Granted the "first-in first-out" principle of working memory, the integrated model will yield a conclusion, if any, of the form:

C – A.

The two remaining figures are still more complicated. There are two possible routes by which to integrate premises in the figure:

A – B
C – B.

Reasoners can construct a model of the first premise, and then switch round their interpretation of the second premise so the two occurrences of the middle term are temporally contiguous. Alternatively, they can construct a model of the second premise, renew their interpretation of the first premise, and then switch it round so as to make the integration possible. Switching round an interpretation must not be confused with the operation of converting a premise, though the two notions are related. The converse of "Some A are B" is "Some B are A" and the two assertions are equivalent in that when one is true the other is true; the converse of "All A are B" is "All B are A" but they are not equivalent. If reasoners formed the converse of a premise, they would often fall into error, and several theorists have proposed that such illicit conversions do occur (e.g. Chapman and Chapman, 1959; Revlis, 1975). However, switching round an interpretation does not affect logical accuracy since it concerns only the order of information in working memory. The interpretation of "All A are B" takes the form:

a=b
a=b
(b)

If this interpretation is switched round, it takes the form:

b=a
b=a
(b)

This revision is logically accurate: an illicit conversion would only occur if the parenthetical token representing the possibility of b's that are not a's was dropped in the process. The purpose of switching round an interpretation is to bring the two occurrences of the middle term into temporally adjacent positions.

The difficulty of the remaining figure:

B – A
B – C

is still greater. It calls both for re-ordering the premises and for switching round an interpretation. There are again two alternative procedures. Reasoners may switch round a model of the second premise, renew their interpretation of the first premise, and then add its information to the model. Alternatively, after they have initially interpreted both premises, they may renew their interpretation of the first premise, switch round their model of it, renew their interpretation of the second premise, and add its information to the model.

The reader familiar with Hunter's (1957) classic account of three-term series problems will recognise the similarity of assumption (4) to his account. The operations that are required to form a mental model from premises in each of the four figures are summarised in Table III, together with the resulting biases in the form

TABLE III

The operations required to form an integrated model of premises in the four figures, together with the predicted response bias

	Figure of premises					
Predicted operations	A–B B–C	B–A C–B	A–B C–B		B–A B–C	
Renew interpretation of of first premise	0	1	0	1	1	1
Switch round an interpretation	0	0	1	1	1	1
Renew interpretation of second premise	0	0	0	0	0	1
Predicted response bias	A–C	C–A	A–C	C–A	C–A	A–C

of the conclusions. There is an increase in the number of operations required to form a mental model over the four figures.

(5) The greater the load on working memory, the harder it will be to make an inference. One factor that should plainly increase the load on working memory is the need to carry out the additional operations required to form an integrated model in certain figures. Table IV presents the relevant results from three experiments. In Experiment I, 20 American students were given all 64 possible pairs of premises with a sensible everyday content and asked to state what followed from each pair of premises (see Johnson-Laird and Steedman, 1978). Experiments II and III are part of an unpublished investigation carried out by Bruno Bara and myself; Experiment II consisted of a replication of the American study carried out with 20 Italian students at the University of Milan; Experiment III was a further replication with another group of 20 subjects who were given only 10 s in which to make their response to each problem. In all three experiments, the figure of the problems produced the predicted trend in the difficulty of drawing a valid conclusion. The experiments also yielded a highly reliable bias in the form of the conclusions. Over 80% of the conclusions to problems in the figure:

A – B
B – C

were of the form:

A – C

whereas over 80% of the conclusions to problems in the figure:

B – A
C – B

were of the form:

C – A

The remaining two figures showed no reliable bias either way, which suggests that subjects made use of both the alternative strategies available to them in these cases.

TABLE IV

The percentages of valid conclusions as a function of the figure of the premises in three experiments

	Figure of the premises			
	A–B B–C	B–A C–B	A–B C–B	B–A B–C
Experiment I	60	50	53	49
Experiment II	51	48	35	22
Experiment III	40	27	33	16

(6) Ordinary individuals who have not been taught logic do not make use of rules of inference in order to make valid deductions. They have instead one essential piece of semantic information. They know that an inference is valid if the conclusion is true in every state of affairs in which the premises are true. In other words, a putative conclusion follows validly from a set of premises, if it is true when the premises are true and there is no way of interpreting the premises so as to render it false. Because there is no such way of re-interpreting the premises in the easy syllogism, the conclusion that was derived above is a valid one. But, consider now the second more difficult syllogism.

Its premises are:

All of the beekeepers are artists
None of the chemists are beekeepers.

Since the two occurrences of the middle term, "beekeepers", are not adjacent, the figure demands that the initial model is constructed on the basis of the second premise. This negative premise can be interpreted by forming a model in which the two classes are isolated from each other:

chemist
chemist
— — — — — — — — — —
beekeeper
beekeeper

The force of the broken line is to indicate that no chemist is, or can subsequently be represented as, identical to a beekeeper. A more explicit notation would indicate that no chemist is identical with any beekeeper, and vice versa. There would be a relation of the following sort:

chemist≠beekeeper

between each possible pairing of chemist and beekeeper. One way in which the information from the first premise can be added to the model is as follows:

chemist
chemist
— — — — — — — — — —
beekeeper=artist
beekeeper=artist
(artist)
(artist)

This model suggests the conclusion, "None of the chemists are artists", drawn by 60% of the subjects in the American experiment. Reasoners who attempt to search for an alternative interpretation of the premises may succeed in finding a second integrated model of the premises:

chemist
chemist = artist
— — — — — — — — — —
beekeeper=artist
beekeeper=artist
(artist)

This model shows that the previous conclusion is false, and suggests instead the conclusion, "Some of the chemists are not artists," which was drawn by 10% of the American subjects. There is, however, a third possible model of the premises:

chemist = artist
chemist = artist
— — — — — — — — — —
beekeeper=artist
beekeeper=artist

which suggests that even the last conclusion is invalid. At this point, it is tempting to respond that there is no valid conclusion interrelating the chemists and artists, a response that was made by 20% of the subjects. However, there is one relation that is common to all three models of the premises: "Some of the artists are not chemists." The difficulty of this response, as reflected in the fact that not a single subject made it, is a consequence of having to construct three different models of the premises and to evaluate each model in an order that violates the "first-in first-out" principle of working memory.

Some syllogistic premises yield only a single mental model, others yield two alternative models, and still others yield three models. On the assumption that a greater number of models will place a greater load on working memory, a trend in difficulty can obviously be predicted. The relevant results from Experiments I, II, and III, are summarised in Table V. In each experiment, there was a

TABLE V

The percentages of correct valid conclusions in three experiments on syllogistic reasoning. The percentages are shown as a function of the number of mental models that have to be constructed to yield the correct conclusion

	One model	Two models	Three models
Experiment I	92	46	28
Experiment II	80	20	9
Experiment III	62	20	3

highly reliable trend: the greater the number of models that have to be constructed, the poorer the performance. In fact, we have yet to test a subject who does not perform best on the one-model problems.

The explanation of the effects of the figural arrangement of terms has received further empirical support. In Experiment II, we recorded the latencies of the subjects' responses. Table VI presents the mean latencies of the correct valid conclusions for the one-model problems, the only sort that produced enough correct responses for the latencies to be analysed. Even with the one-model

TABLE VI

The mean latencies (s) to produce the correct valid conclusions to one model syllogisms in Experiment II. The results are shown as a function of figure

	Figure of the premises			
	A–B B–C	B–A C–B	A–B C–B	B–A B–C
Mean latencies	11·6	12·9	18·7	22·1

problems some subjects failed to produce any correct responses in some figures, but we were able to rank order the mean correct latencies for 14 subjects as a function of figure. The mean ranks for the four figures were: 1·7, 2·3, 2·6, and 3·4 (Page's $L=387$, $P<0{\cdot}0005$). This trend is, of course, exactly what is predicted by the

number of additional operations required for each figure (see Table III above).
As I noted earlier, Hunter's (1957) model of three-term series problems, such as:

Ann is taller than Betty

Betty is taller than Carol

Therefore, Ann is taller than Carol

makes use of very similar operations to those embodied in assumption (4). As a theory of such problems, however, it has been overshadowed in recent years by an emphasis on the effects of "end-anchoring" (Huttenlocher, 1968) and on linguistic effects such as the apparent negativity of some comparative adjectives like "shorter" (Clark, 1969). However, Bruno Bara, Patrizia Tabossi, and I, have obtained direct evidence for the operations of assumption (4) in the conclusions that subjects draw spontaneously to the premises of three-term series problems that concern kinship relations, e.g.:

Arthur is related to Bertrand

Charles is related to Bertrand

where the same term can be used as its own converse, thus eliminating any effects of potential negativity. We found that, in general, there was a bias towards drawing conclusions in which the end term in the first premise was also the first term in the conclusion:

Arthur is related to Charles.

These problems are so easy that presumably the subjects could always resort to the strategy that yielded such conclusions for the two symmetric figures:

A – B B – A
C – B B – C.

However, premises in the figure:

A – B
B – C

produced the greatest bias for A – C conclusions (77%), whereas the one figure that has no strategy that allows it:

B – A
C – B

eliminated the bias for this conclusion (47%).

Reasoning without logic

A formal logic is a calculus for proving the validity of an inference. It provides us with a systematic method for establishing in effect that there is no interpretation of the premises that is consistent with a denial of the conclusion. In fact, logicians have developed a variety of such methods. I will not discuss these methods, but turn directly to the claim that the manipulation of mental models enables valid inferences to be made without recourse to rules of logic. This claim, as I know from the reaction of audiences to whom it has been addressed, is both hard to understand and hard to believe, it is viewed as almost on a par with the Pelagian heresy in some quarters.

The crux of the matter is that a system of inference may perform in an entirely logical way even though it does not employ rules of inference, inferential schemata, meaning postulates, or any other sort of machinery conventionally employed in a logical calculus. The rest of the argument is simple once this point is grasped, and so I will labour it awhile. The theory applies to any sort of deductive inference though I have illustrated it here only with respect to syllogisms (see Johnson-Laird, in press, for a general account).

During the course of developing the theory of mental models, I have written several computer programs (in the high-level list-processing language, POP-10) that model the process of syllogistic inference. The most recent of these programs constructs a model of one premise, adds the information from the second premise, and then draws a conclusion interrelating the end items according to the assumptions described above. It then searches for an alternative model of the premises that would render the conclusion false. If the conclusion is affirmative, it examines the model for tokens representing the middle term that are not linked to both end terms. On finding such a token, it then breaks a link that is part of a chain that identifies two end items, and uses the free middle item in order to construct a new model that is true to the premises. For example, the following sort of model:

```
a=b=c
a=b=c
  (b)
  (b)
```

can be reconstructed as:

```
a=b=c
a=b
  b=c
  (b)
```

where both are models of the premises of the form:

All the A are B
All the C are B

The program uses a similar method for attempting to destroy negative premises. At each step, it produces whatever conclusion is currently warrantable, and it continues to try to reconstruct the model so as to falsify a conclusion until it has run out of any further possibilities, a point at which it always arrives since the models have only a finite number of entities in them. Nowhere in the program are there any rules of inference, or inferential schemata, that are employed to guide the search process. The program searches in a way that resembles Bartlett's (1958) account of the sectional map-reader: it is not random but guided by a goal. It embodies merely a knowledge of the truth conditions of syllogistic premises, and the principle of searching for models that refute conclusions, the principle that likewise underlies, without being directly represented in, any logical calculus or formal system of deduction.

Such is the scepticism of certain defenders of the doctrine of mental logic that they refuse to believe that the program does not employ rules of logic. What about the machinery for searching for end items, destroying identities, and testing truth conditions? They argue that it must surely embody logical rules; indeed, they say, a computer programming language has such rules built into it. If what they mean is that the program, or programming languages in general, make use of rules of inference, then the claim is false. If the program was governed by rules of inference, then it would not be able to model the errors that people make. It is easy to write a program that makes invalid deductions, and indeed the first stage in the output of my program is often to produce an invalid conclusion. The power of programs comes from the computational machinery of recursive functions that they realise, and this machinery suffices for developing algorithms for both logical and illogical inference.

The theory of mental models assumes that human reasoners construct models and search for alternatives, not necessarily in a random way, but certainly not in a wholly systematic way, either. Moreover, since even the most intelligent individuals have difficulty with certain syllogisms, and are aware of it, they have an obvious motivation to try to externalise and to systematise the search for alternative models of the premises. Hence, the theory suggests an obvious reason for the development of logic as a intellectual discipline.

When Aristotle invented logic, his method was to determine which pairs of syllogistic premises yield valid conclusions (see Kneale and Kneale, 1962). An inference of the form:

Every man is an animal

No stone is a man

Therefore, No stone is an animal

certainly yields a true conclusion. In order to determine whether inferences of this form were valid, Aristotle changed the content of the premises whilst holding their form constant, e.g.:

Every man is an animal

No horse is a man

Therefore, No horse is an animal.

The conclusion is manifestly false, but the inference is identical in form to the previous example. Since the form can lead to false conclusions from true premises, it must be invalid. Instead of searching for interpretations of premises that refute conclusions, Aristotle held form constant and searched for premises with a true content leading to a blatant falsehood and in this way he refuted all arguments of the same form. He used semantics to determine the set of valid syllogisms, and then he developed a logic, a set of principles for establishing validity.

Individual differ ences

My aim was to decide amongst three mutually exclusive hypotheses about reasoning: that it was invariably logical, that it was invariably illogical, that it was sometimes logical and sometimes illogical. Only the last of these alternatives appears to conform with the facts; it is the only one of them that truly treats thinking as a skill. Skills, of course, have to be acquired, and individuals differ in their mastery of them. Reasoning is no exception. In the studies of syllogistic reasons ing that my colleagues and I have carried out, the extent of individual difference-has been painfully apparent. The best subject that we have tested responded correctly to 85% of the premises that yield a valid conclusion interrelating the end items; the worst subject that we have tested responded correctly to only 15% of them. In fact, one subject might be thought to have done still worse: she abandoned the experiment on the grounds that she was not personally acquainted with the individuals referred to in the premises. However, this reaction reflects, not necessarily an inability to reason, but a reluctance to engage in the laboratory "game" of reasoning, a reaction common among people who live in a non-literate culture (see Scribner, 1977; Luria, 1977).

What causes individuals to differ in their ability to make inferences? No certain answer is yet known. In the past, psychologists have often been content to correlate performance in a particular reasoning task either with scores on intelligence tests or with factors postulated to underlie test performance (e.g. Guildford, 1959; Frandsen and Holder, 1969). They have been happy to treat "general intelligence", or some such notion derived from a factor analysis of test data, as a primitive and unanalysed commodity that gives rise to observed differences in ability. Whatever the general merit of investigating "individual differences" by way of mental tests, their use is of little value in the study of thinking. The data they yield are too gross to elucidate differences in mental processes from one individual to another.

The theory of mental models offers an explanatory framework that helps to make sense of differences in reasoning ability. It specifies the separate components underlying inferences and places several constraints on the possible differences amongst individuals. The theory assumes that syllogistic inference, for example, depends on three component skills: (1) an ability to form an integrated model of the premises; (2) an appreciation that an inference is only sound if there are no counterexamples to it, together with a capacity to put this principle into practice; (3) an ability to put into words the common characteristics of a set of mental models. Bruno Bara and I have begun to explore the differences in the detailed performance of subjects carrying out syllogistic reasoning, and I will describe some of our preliminary findings in the light of the three main abilities postulated by the theory.

The main difficulty in constructing an integrated model is that a representation of one premise must be held in working memory while information from the other premise is combined with it. Although all the subjects that we have tested performed best with one-model problems, we have tested two subjects who failed to perform reliably better than chance with these problems. (They were quite unable to cope with premises that required more than model.) The figural arrangement of terms had a striking effect on their performance: they could only form a model from premises in the two theoretically easier figures:

A – B	B – A
B – C	C – B

With premises in the other two figures, which require interpretations to be switched round, they either declared erroneously that there was no valid conclusion or forgot one of the end terms and mistakenly replaced it with the middle term so as to form a conclusion that was blatantly inconsistent with the premises. Their tendency to assert that there was no valid conclusion if the figures required interpretations to be switched round gave rise to a spuriously good performance with invalid syllogisms in these figures.

Only where a valid inference depends on constructing alternative models of the premises are genuine differences in inferential ability to be observed. A reasoner must appreciate the need to construct and to evaluate different models, and must be able to carry out this procedure within the processing limitations of working memory. Some subjects seem not to perceive the need to consider alternatives. The hallmark of their performance is a string of erroneous conclusions combined with a reluctance to respond that there is no valid conclusion interrelating the end items. Three of the subjects that we have tested failed disastrously with premises that required more than one model to be constructed and responded correctly to invalid problems on less than one in five occasions. Other subjects evidently perceive the need to consider alternative models, but are wholly incapable of assessing them correctly. The hallmark of their performance is a tendency to respond, "No valid conclusion," whenever there is more than one model of the premises. They, too, do spuriously well with the class of problems that have no valid conclusions, but they make this same response to the two- and three-model premises that have a valid conclusion. Any subject who performs better with invalid syllogisms than with valid syllogisms is showing signs of this syndrome.

Most of the subjects that we have tested are able to construct some alternative models, but from time to time they fall down in assessing their implications. They are particularly prone to error in those figures that require interpretations to be switched round, failing to detect either that a putative conclusion is violated by one alternative model or else that there is a conclusion common to all the alternatives. It is noteworthy that only one subject that we have tested showed any competence with the most difficult syllogisms of all, namely, those with three models where the conclusion runs counter to the figural bias.

There are a number of other differences in performance between the subjects, including their susceptibility to figural effects, which I will not discuss here. My aim has been to establish that the theory of mental models provides a framework suitable for describing individual differences, and suggests one important clue to their cause. Apart from a knowledge of the semantic principle of refutations, perhaps the single biggest factor in reasoning is the processing capacity of working memory, since the effects of both the number of models and the figure of the premises appear to arise as a consequence of its limitations. The same point has emerged from studies of other sorts of inference. In an experiment carried out by Johnson-Laird and Wason (1970), for example, the task was to check whether a description of the contents of an envelope was correct. The subjects selected diagrams one at a time from a set laid out in front of them. As they selected a diagram, the experimenter told them whether or not the same diagram was in the envelope. A sensible strategy in this task is to choose diagrams that do not fit the description on the envelope: if such a diagram is in the envelope, then plainly the description is false. Some subjects, however, choose diagrams that fit the description. This choice is uninformative once it is known that the envelope is not empty, because there is no reason why a diagram that fits the description cannot be outside the envelope. What was unexpected was that a complex disjunctive description had a striking effect on subjects' insight into the task. A subject would perform perfectly with one description, only to lose that insight on the very next trial when the disjunction occurred. The point to be stressed is that the content of the particular description has no bearing on the "logic" of the task, yet it had a considerable impact on performance. As Johnson-Laird and Wason commented: "it is possible that this complex disjunctive description occupies a greater amount of short-term memory than a single complex rule, and thus leaves a smaller amount of 'computing space' available for handling the selection of the diagram." More recently, Baddeley and his colleagues have made a comprehensive examination of the role of working memory in simple verbal inferences. They have found that when subjects are asked to hold in mind a string of digits, then their performance in reasoning tasks is adversely affected (see, for example, Baddeley and Hitch, 1974; Hitch and Baddeley, 1976).

Some practical implications

In a modern society, it is advantageous to be able to think logically. That might seem a surprising claim; let me give just one illustration to support it. We are all governed by rules and regulations that confer upon us certain rights and duties. The government and its agencies issue numerous leaflets to keep us informed. These leaflets are, however, notoriously difficult to understand. A decade ago, A. R. Jonckheere, Sheila Jones, Peter Wason, and others, introduced the technique of converting complicated rules into "logical trees". A typical passage from an official leaflet reads as follows:

> The earliest age at which a woman can draw a retirement pension is 60. On her own insurance she can get a pension when she reaches that age, if she has then retired from regular employment. Otherwise she has to wait until she retires or reaches age 65. At age 65 pension can be paid irrespective of retirement. On her husband's insurance, however, she cannot get a pension, even though she is over 60, until he has reached age 65 and retired from regular employment, or until he is 70 if he does not retire before reaching that age.

The same information is much more easily understood (see Wason, 1968; Jones, 1968) when it is presented as a logical tree, such as:

1. Are you under 60 years of age? If YES, you are not entitled to a pension.
If NO, read item 2.

2. If you are claiming on your own insurance, read question 3.
If you are claiming on your husband's insurance, read question 5.

3. Are you under 65 years of age? If YES, read question 4.
If NO, you are entitled to a pension.

4. Are you working? If YES, you are not entitled to a pension.
If NO, you are entitled to a pension.

5. If your husband's age is:
(a) less than 65, you are not entitled to a pension.
(b) between 65 and 69, read question 6.
(c) 70 or more, you are entitled to a pension.

6. Has your husband retired? If YES, you are entitled to a pension.
If NO, you are not entitled to a pension.

One of the leaflets that Wason and Jones successfully rendered comprehensible concerned the eligibility of widows for death grant. I recently read the new leaflet (NI.49 issued in 1979) on this topic to see what lessons Whitehall had learned. The crucial extract runs as follows:

> Death grant is payable where either of the following conditions is satisfied by the person on whose contributions the grant is claimed:
>
> The contributor must have paid or been credited with at least 25 contributions of any class at any time between 5 July 1948 or the date of entry into insurance, if later, and 5 April 1975, or the date on which he reached 65 (60 for a woman), or died under that age, whichever is the earliest; or
>
> Since 6 April 1975 the contributor must have actually paid contributions in any one tax year (6 April to the following 5 April) before the relevant year, on earnings of at least 25 times the lower earnings limit for that year. The relevant year is usually the income tax year in which the death occurred, but if immediately before the date of death, the person on whose contributions the grant is claimed was himself dead or over 65 (60 for a woman), it is either the year in which he reached that age, or the year in which he died, whichever is earlier.

I hope that that clears up any confusions the reader might have had about the matter! On a flying visit to University College, London, the late Richard Crossman told Jonckheere, Jones, and me, that there were dangers in clarity: people would get what they were entitled to and that would cost the government a lot of money. I used to think that he said it as a joke.

What is needed, evidently, are ways of improving reasoning ability, and educationalists have developed a variety of methods that are supposed to do so. They include the pedagogical use of stories illustrating logical principles (Lipman and Sharp, 1978), practice with special reasoning problems (Feuerstein, Hoffman and Miller, 1980), and courses on thinking and problem solving (e.g. Whimbey and Lochhead, 1980). Psychologists have become increasingly involved in such matters, especially since the start of the project to raise the intelligence of the entire population of Venezuela. (The International newsletter, Human Intelligence, has published several reports on this project which includes work carried out by researchers at Harvard University, Bolt Beranek and Newman, Inc., and many other research organisations.)

My own work suggests that the most common cause of difficulty in reasoning for individuals living in a literate society is the limited processing capabilities of working memory. Its effects have been apparent in every subject that we have tested. However, it must be emphasised that there appears to be a spontaneous improvement in reasoning ability simply as a consequence of practice (with no knowledge of results). The subjects in Experiment I (see Table V) were tested again 1 week later. They were given no forewarning that they would be re-tested, but their overall performance increased by 10%, and 19 out of the 20 subjects returned an improved score. One striking differential effect of practice occurred with the valid conclusions drawn in the most difficult figure, B – A, B – C, which requires an interpretation to be switched round. Here, there was an overall improvement of 20%, and half of it was due to a decline in erroneous responses that there was no valid conclusion. The effect of practice must in part be to increase the efficiency of the encoding operations of working memory. Experience with the task may also produce a growing awareness of the logical properties of the problems. Some subjects may begin to notice, for example, that two negative premises never yield an interesting valid conclusion.

Several people have suggested to me that diagrams of mental models might serve a useful pedagogical function in teaching the principles of deduction. Although the prospect is appealing, it may be dangerous. Whenever I have presented a reasoning problem informally, I have noticed the difficulties that people get themselves into if they use Euler circles. The problem is that there is no simple algorithm for using them that one can learn like one learns, say, the algorithm for long multiplication. Merely drawing circles does not guarantee that all their possible combinations will be considered exhaustively. The same problem applies to the notation that I have invented for depicting the structure of mental models; if there were a simple algorithm, then doubtless most of us would have mastered it when we first learned to reason. Educators are probably better advised to ensure that their students understand the fundamental principle of deductive inference and get plenty of opportunities to put it into practice.

Conclusions

Bartlett believed that thinking is a high-level skill. He considered only a restricted variety of thought processes, and I have restricted myself to a still narrower aspect of what he would have called, "closed system thinking." I have argued that such thinking consists of three essential skills:

(1) *The ability to construct mental models of the situations described by sentences.* This is part of the process of verbal comprehension, and implicit inferences are nothing more than the use of pieces of general knowledge to aid this process of constructing a single mental model. They are rapid and automatic because no attempt is made to test their validity.

(2) *The ability to search for different models of the same premises in order to check whether an inference is valid.*

Most people appear to be aware of the principle at stake, here, but there are marked differences in individual skill at putting it into practice. A major cause of the difficulty of making explicit deductions is the need to form integrated models, and to search for alternatives, within the processing capacity of working memory. The sequence and timing of the operations integrating the premises are critical: they show up in the effects of figure on the form of conclusions, on the time that it takes to make an inference, and on the chances of drawing the correct conclusion.

(3) *The ability to put into words the common characteristic of a set of mental models.*

Here, again, people differ in skill. It is a rare individual indeed who can cope with the assessment of a set of models that have to be evaluated in violation of the "first-in first-out" principle of working memory.

There is no need to postulate rules of logic in the mind. There is no need to suppose that human beings are intrinsically irrational. Logical thinking is a skill that is exercised with varying degrees of success. Logic itself is a consequence of our happy ability to search for refutations, not the cause of that ability.

The solution to the Sherlock Holmes riddle is that Watson and he must have gone along the veranda from right to left (given that it runs along the lower side of the plan). The reason is because they entered the house from one end of the veranda, passed from room to room, and then turned right into Milverton's study with its door opening onto the veranda. By a nice coincidence, Sheila Jones lives in the house in Hampstead that is generally believed to be Conan Doyle's model for the story.

This research was supported by a grant from the Social Science Research Council. Many individuals have helped me, and I am particularly grateful to Bruno Bara, Kate Ehrlich, Alan Garnham, Dave Haw, Janellen Huttenlocher, Maria Legrenzi, Paolo Legrenzi, Kannan Mani, George A. Miller, Jane Oakhill, Mark Steedman, Patrizia Tabossi, Peter Wason, and Til Wykes, for some fruitful collaborations over the years. Alan Garnham, Leslie Henderson, Jane Oakhill, and Keith Oatley, very kindly read an earlier version of this paper and made many helpful criticisms of it. I also thank my colleagues Steve Isard, Christopher Longuet-Higgins, and Stuart Sutherland, for ideas, encouragement, and advice. Finally, I am grateful to the Experimental Psychology Society and to its officers for the invitation to give this lecture and for helping to make its occasion a pleasure (for me) rather than a duty.

References

BADDELEY, A. D. and HITCH, G. (1974). Working memory. In BOWER, G. H. (Ed.), *The Psychology of Learning and Motivation*, Volume 8. New York: Academic Press.

BARTLETT, F. (1958). *Thinking: An Experimental and Social Study*. London: Allen and Unwin.

BETH, E. W. and PIAGET, J. (1966). *Mathematical Epistemology and Psychology*. Dordrecht: Reidel.

CHAPMAN, I. J. and CHAPMAN, J. P. (1959). Atmosphere effect re-examined. *Journal of Experimental Psychology*, **58**, 220–6.

CLARK, H. H. (1969). Linguistic processes in deductive reasoning. *Psychological Review*, **76**, 387–404.

COHEN, L. J. (1981). Can human irrationality be experimentally demonstrated? *Behavioral and Brain Sciences*, **4**, 317–70.
CRAIK, K. J. W. (1943). *The Nature of Explanation*. Cambridge: Cambridge University Press.
DAVIES, D. J. M. and ISARD, S. D. (1972). Utterances as programs. In MICHIE, D. (Ed.) *Machine Intelligence*, Volume 7. Edinburgh: Edinburgh University Press.
ERICKSON, J. R. (1974). A set analysis of behavior in formal syllogistic reasoning tasks. In SOLSO, R. (Ed.), *Theories in Cognitive Psychology: The Loyola Symposium*. Potomac, MD.: Erlbaum.
EVANS, J. ST B. T. (1977). Towards a statistical theory of reasoning. *Quarterly Journal of Experimental Psychology*, **29**, 621–35.
EVANS, J. ST B. T. (1980). Current issues in the psychology of reasoning. *British Journal of Psychology*, **71**, 227–39.
FALMAGNE, R. J. (1980). The development of logical competence: a psycholinguistic perspective. In KLUWE, R. H. and SPADA, H. (Eds), *Developmental Models of Thinking*. New York: Academic Press.
FEUERSTEIN, R., HOFFMAN, M. B. and MILLER, R. (1980). *Instrumental Enrichment*. Baltimore: University Park Press.
FODOR, J. A. (1980). Fixation of belief and concept acquisition. In PIATELLI-PALMARINI, M. (Ed.), *Language and Learning: The Debate between Jean Piaget and Noam Chomsky*. Cambridge, Mass.: Harvard University Press.
FRANDSEN, A. N. and HOLDER, J. R. (1969). Spatial visualization in solving complex verbal problems. *Journal of Psychology*, **73**, 229–33.
GARNHAM, A. (1981). Are inferences need driven? Unpublished paper, Laboratory of Experimental Psychology, University of Sussex.
GUILFORD, J. P. (1959). Three faces of intellect. *American Psychologist*, **14**, 469–79.
GUYOTE, M. J. and STERNBERG, R. J. (1978). A transitive-chain theory of syllogistic reasoning. Technical report No. 5, Department of Psychology, Yale University.
HENLE, M. (1962). On the relation between logic and thinking. *Psychological Review*, **69**, 366–78.
HENLE, M. (1978). Foreword to REVLIN, R. and MAYER, R. E. (Eds), *Human Reasoning*. Washington, D.C.: Winston.
HITCH, G. J. and BADDELEY, A. D. (1976). Verbal reasoning and working memory. *Quarterly Journal of Experimental Psychology*, **28**, 603–21.
HUNTER, I. M. L. (1957). The solving of three-term series problems. *British Journal of Psychology*, **48**, 286–98.
HUTTENLOCHER, J. (1968). Constructing spatial images: a strategy in reasoning. *Psychological Review*, **75**, 550–60.
JOHNSON-LAIRD, P. N. (1970). The perception and memory of sentences. In LYONS, J. (Ed.), *New Horizons in Linguistics*. Harmondsworth, Middx: Penguin.
JOHNSON-LAIRD, P. N. (1980). Mental models in cognitive science. *Cognitive Science*, **4**, 71–115.
JOHNSON-LAIRD, P. N. (1981). Comprehension as the contruction of mental models. *Proceedings of the Royal Society of London B*, **295**, 353–74.
JOHNSON-LAIRD, P. N. (In press). *Mental Models*. Cambridge: Cambridge University Press. Cambridge, Mass.: Harvard University Press.
JOHNSON-LAIRD, P. N., LEGRENZI, P. and LEGRENZI, M. S. (1972). Reasoning and a sense of reality. *British Journal of Psychology*, **63**, 395–400.
JOHNSON-LAIRD, P. N. and STEEDMAN, M. J. (1978). The psychology of syllogisms. *Cognitive Psychology*, **10**, 64–99.
JOHNSON-LAIRD, P. N. and WASON, P. C. (1970). Insight into a logical relation. *Quarterly Journal of Experimental Psychology*, **22**, 49–61.
JONES, S. (1968). *Design of Instruction*. London: H.M.S.O.
KNEALE, W. and KNEALE, M. (1962). *The Development of Logic*. Oxford: Clarendon Press.

LIPMAN, M. and SHARP, A. M. (1978). *Growing Up With Philosophy.* Philadelphia: Temple University Press.

LONGUET-HIGGINS, H. C. (1972). The algorithmic description of natural language. *Proceedings of the Royal Society B*, **182**, 255–76.

LURIA, A. R. (1977). *The Social History of Cognition.* Cambridge, Mass.: Harvard University Press.

MANI, K. and JOHNSON-LAIRD, P. N. (1982). The mental representation of spatial descriptions. *Memory and Cognition*, In press.

MILLER, G. A. and JOHNSON-LAIRD, P. N. (1976). *Language and Perception.* Cambridge: Cambridge University Press. Cambridge, Mass.: Harvard University Press.

OAKHILL, J. (1982). Constructive processes in skilled and less-skilled comprehenders' memory for sentences. *British Journal of Psychology*, In press.

RAINE, C. (1979). *A Martian Sends a Postcard Home.* Oxford: Oxford University Press.

REVLIS, R. (1975). Two models of syllogistic reasoning: feature selection and conversion. *Journal of Verbal Learning and Verbal Behavior*, **14**, 180–95.

SCRIBNER, S. (1977). Modes of thinking and ways of speaking: culture and logic reconsidered. In JOHNSON-LAIRD, P. N. and WASON, P. C. (Eds), *Thinking: Readings in Cognitive Science.* Cambridge: Cambridge University Press.

WASON, P. C. (1968). The drafting of rules. *New Law Journal*, **118**, 548–9.

WASON, P. C. and JOHNSON-LAIRD, P. N. (1972). *Psychology of Reasoning: Structure and Content.* London: Batsford. Cambridge, Mass.: Harvard University Press.

WASON, P. C. and SHAPIRO, D. (1971). Natural and contrived experience in a reasoning problem. *Quarterly Journal of Experimental Psychology*, **23**, 63–71.

WHIMBEY, A. and LOCHHEAD, J. (1980). *Problem Solving and Comprehension: A Short Course in Analytical Reasoning*, 2nd Edition. Philadelphia: The Franklin Institute Press.

WOODS, W. A. (1967). Semantics for a question-answering system. Mathematical Linguistics and Automatic Translation Report NSF-19, Harvard Computational Laboratory.

WYKES, T. (1978). Inference and children's comprehension of prose. Unpublished D.Phil. thesis, University of Sussex.

WYKES, T. and JOHNSON-LAIRD, P. N. (1977). How do children learn the meanings of verbs? *Nature*, **268**, 326–7.

Revised version received October 1981