# 11

# Strategies and Tactics in Reasoning

Philip N. Johnson-Laird
Fabien Savary
Monica Bucciarelli

This chapter reports three experiments that investigated the strategies that individuals developed for themselves in various sorts of reasoning. The results suggest that there are four levels of thinking: at the topmost level, there is metacognitive thinking, which can yield novel strategies for reasoning; at the second level are the strategies themselves; At the third level are the components of strategies, which are a variety of 'tactics', such as representing a premise in a diagram; and at the fourth level are the largely unconscious processes that underlie the tactics. Any feasible theory of these various levels must, at present, contain many nondeterministic components, and the best way to express such a theory is in the form of a grammar.

Not long ago, a visiting speaker, a distinguished cognitive scientist, came to Princeton. He visited our laboratory, and one of us, by way of entertaining him gave him a tricky inferential problem, one that almost everyone gets wrong. He got it wrong, too. Later, he explained:

> *I said to myself, "This is one of Phil's silly inference problems, and so I've got to be careful." But I still got it wrong!*

Despite his insight—a metacognitive one, our visitor failed to come up with an appropriate strategy for dealing with the inference—as indeed had we when we first attempted the inference in checking the output of a computer program.

Current theories of deductive reasoning have largely neglected the topic of inferential strategies, their variety, and the constraints that determine the particular strategy that a reasoner adopts (see Evans, Newstead, & Byrne, 1993, for a review of current theories). The aim of the present chapter is to make good this neglect. But, what exactly *is* a strategy in reasoning? The best way to answer this question is to consider an example, although not the tricky problem that we gave our distinguished visitor. Consider instead the following simple problem about marbles in a box:

> *There is a red marble in the box if and only if there is a brown marble in the box.*
> *Either there is a brown marble in the box or else there is a gray marble in the box, but not both.*
> *There is a gray marble in the box if and only if there is a black marble in the box.*
> *Does it follow that:*
> *If there is not a red marble in the box then there is a black marble in the box?*

You, the reader, are invited to solve this problem, and then to try to characterize how you went about it.

The correct answer to the problem is "yes": If there is not a red marble in the box, then there is a black marble in the box. But, how did you solve the problem? One strategy is to use a supposition, that is, an assumption for the sake of argument. Thus, you might have said to yourself:

> *Suppose that there is not a red marble in the box. It follows from the first premise that there is not a brown marble in the box, either. And, in this case, it follows from the second premise, that there is a gray marble in the box. The third premise now implies that there is a black marble in the box. So, if there is not a red marble in the box, then it follows that there is a black marble. Hence, the conclusion follows.*

This protocol, which in fact is typical of what some reasoners say, reveals one strategy that can be used to solve the problem: the suppositional strategy. But, as we will see, it is not the only strategy that people use to solve this problem.

A working definition of a *strategy* is that it is the sequence of steps that an individual follows in solving, or attempting to solve, a problem. Thus, the first step in the foregoing strategy is to make a supposition (corresponding to the antecedent clause of the conditional conclusion). The second step is to combine this supposition with the first premise in order to draw an intermediate conclusion. And the next step is to use this conclusion to make another inference, and so on, until one arrives at the consequent proposition in the conclusion.

Each step in a strategy is what we refer to as a *tactic*. The mental processes underlying a tactic are seldom, if ever, available to consciousness. Thus, noone knows for sure how people make an inference of the form known as Modus Ponens, the tactic that occurs in the earlier second step. People do not report how they carry out tactical steps in their "think aloud" protocols. Their nature is therefore highly controversial. Some psychologists argue that Modus Ponens depends on the use of a formal rule of inference (see Braine & O'Brien, 1991; Rips, 1994). Other psychologists, including the present authors, argue instead that the minds of logically untrained reasoners do not contain tacit rules of inference, and that an inference such as Modus Ponens depends on constructing mental models of the premises (see Johnson-Laird & Byrne, 1991). In contrast to tactics, the overall strategy that reasoners use is potentially available to introspection, and can be revealed by reasoners' verbal reports, especially if they have to think aloud as they tackle a problem (see also Evans, chap. 1, this volume). In our view, reasoning tactics probably depend on the manipulation of mental models, but reasoning strategies are, as yet, almost wholly unknown. They are a matter for empirical investigation.

The plan of the present chapter is simple. It begins with a brief account of how psychologists have thought about strategies in the past. It presents an experimental investigation of the strategies underlying reasoning with sentential connectives (such as *"if"* and *"or"*). The results call for a distinction between strategies and tactics, that is, the components of strategies, which in turn depend on unconscious processes. The chapter then formulates a new way to frame theories that are nondeterministic, and illustrates this method in an account of one particular strategy for sentential reasoning. The chapter then turns to reasoning with quantifiers (such as *"all"* and *"some"*). It reports a study in which the participants had to construct models of premises that refuted putative conclusions, and it then describes a study of syllogistic reasoning. The results of these two studies suggest that all current theories of syllogistic reasoning need to be revised. The chapter concludes with an appraisal of strategic and tactical thinking in reasoning.

## STRATEGIES IN REASONING: A BRIEF REVIEW

In the past, psychologists have defended two main views about reasoning strategies. On the one hand, some have argued that reasoners rely on a single deterministic strategy. This view is explicit in Rips' (1989) account of the suppositional strategy that he claims reasoners use to solve so-called "knight-and-knave" problems, such as:

> *There are only two sorts of people: knights, who always tell the truth, and knaves, who always lie.*
> *Arthur says 'Lancelot is a knight and Gawain is a knave'.*
> *Lancelot says 'Arthur is a knave'.*
> *Gawain says 'Arthur is a knave'.*
> *What are Arthur, Lancelot, and Gawain?*

Likewise, Rips' (1994) more recent PSYCOP computer program for reasoning in general follows a single deterministic strategy. A similar view is defended by Martin Braine and his colleagues (see Braine & O'Brien, 1991). On the other hand, Johnson-Laird and Byrne (1990) have argued that naive reasoners use a variety of different strategies for knight-and-knave problems. Consider the previous problem, for example. Many people report that they solved it when they noticed that Lancelot and Gawain are making the same assertion and so they must both be either knights or else knaves. Hence, Arthur's assertion cannot be true, because he assigns Lancelot and Gawain to different categories. Johnson-Laird and Byrne modeled five distinct strategies for knight-and-knave problems, including both a suppositional strategy and the one sketched for the previous problem, which is outside Rips' account. Subsequently, Byrne and Handley (1997) obtained good evidence for the use of several strategies for these problems.

There are other embarrassments to the thesis that there is just a single deterministic strategy. Girotto, Mazzocco, and Tasso (1997) report robust effects of the order of the premises on simple conditional inferences. Such effects seem inexplicable in terms of a single strategy using formal rules of inference. Likewise, reasoners use suppositions relatively rarely in reasoning (see Wason & Johnson-Laird, 1972), which is surprising if the single strategy is based on them. In the light of these results, we began to suspect that the main reason for postulating a single strategy is that previous experimental studies of reasoning had tended to use pairs of premises at most, which leave little room for alternative strategies. They had also failed to gather pertinent evidence about strategies. In order to remedy these defects, we carried out a study of reasoners' strategies in coping with a set of deductions that hinged on sentential connectives.

## A STUDY OF STRATEGIES AND TACTICS IN SENTENTIAL REASONING

In our first experiment, participants were given a set of different inferences about marbles in a box. The problem in the introduction is a typical example of such an inference. Here it is again in an abbreviated form:

*Red if and only if Brown.*
*Either Brown or else Gray, but not both.*
*Gray if and only if Black.*
*Does it follow that:*
*If not Red then Black?*

The inferences had premises that were mainly exclusive disjunctions or biconditionals, as the example illustrates, and their conclusions were either conditionals or disjunctions. For half the problems the conclusions were valid, and for the other half they were invalid. The important feature of the problems, however, is that most of them supported only two alternative possibilities. Each mental model represents a possibility, and so each problem called for two mental models. The premises of the preceding problem, for example, yield the following two models, shown here on separate lines:

*Red Brown*
*Gray Black*

It is evident that the conclusion follows from the premises.

The participants were encouraged to use paper and pencil to evaluate the inferences. They were told to "think aloud" as they tackled each inference, and we video-recorded what they had to say and what they wrote down or drew. The camera was above them and focused on the paper on which they wrote, and they rapidly adapted to the conditions of the experiment.

Problem:
Red iff brown
Brown ore gray
Gray iff black
If not red then black?

| Verbal protocol: | Diagram: |
|---|---|
| Red iff brown | |
| If brown then red | **Br → R** |
| Brown ore gray | |
| If not brown or red then gray | **G** |
| | Br → R |
| Gray iff black | |
| If gray then black: | **Bl →** G |
| | Br → R |
| Yes | |
| Brown ore gray | |
| Red if brown | |
| If black then gray | |
| Not possible (brown and gray) | |
| If not red then necessarily gray | |
| Gray only if black | |
| Yes | |

FIG. 11.1. A typical protocol of a participant solving a reasoning problem transcribed using the interface devised by the second author. The protocol begins with the problem, stated in abbreviated form in which "iff" denotes "if, and only if," and "ore" denotes an exclusive disjunction. The left-hand side of the protocol summarizes what the participant said in thinking aloud, and the right-hand side shows what the participant drew. The participant, in fact, drew a single diagram, and the bold components show what was drawn contemporaneously with the verbal assertions on the left.

We tested eight Princeton students, who had no training in logic. The problems were easy, and none of the participants made any errors in evaluating them, although they were not always right for the right reasons. We devised a computer program that allowed us to transcribe each protocol, including any diagrams drawn by the participant, into a format that was readable in the high-level programming language, LISP.

Figure 11.1 presents a typical protocol of a participant (no. 5) tackling the problem above. The participant drew a single diagram, adding components to it incrementally. The finished diagram was:

*Bl* → *G*
*Br* → *R*

where "*Bl*" stands for "*Black*," "*G*" for "*Gray*," "*Br*" for "*Brown*," and "*R*" for "*Red*." Each row represents a possibility, that is, a mental model.

We believe that our protocols are typical of intelligent individuals thinking aloud as they make inferences. The protocols showed most of the major steps, and they allowed us to identify the participants' strategies and their component tactics. What they did not reveal, however, are the insightful processes in developing or creating the strategies, or the processes underlying the tactical steps, which are largely unconscious.

Our participants used four principal strategies in evaluating the inferences, and we were able to categorize every single protocol, though some protocols showed that the participant had changed mid-problem from one strategy to another. We describe each of the four principal strategies and present examples of them.

*The Suppositional Strategy.* In this strategy, reasoners begin by making an assumption, which corresponds either to the antecedent or to the consequent of the conditional conclusion. They use this supposition to derive a conclusion from a premise. They then use this intermediate conclusion to derive another conclusion from another premise, and so on, until they derive the other proposition (or its negation) in the conditional conclusion. Where individuals base their supposition on the antecedent clause (p) of a conditional of the form: "*if p then q*," then the strategy is closely related, if not identical, to the suppositional strategies postulated in formal rule theories (e.g., Braine & O'Brien, 1991; Rips, 1994) and to the suppositional strategy postulated by Johnson-Laird and Byrne (1991).

In some cases, however, reasoners invalidly based their supposition on the consequent clause (q) of the conditional conclusion and discharged the supposition when the derivation lead them to its antecedent clause (p). This strategy is not valid. If a conclusion of the form "*if A then C*" is interpreted as a "one-way" conditional, that is, *A* implies *C*, but *C* doesn't necessarily imply *A*, then the supposition must correspond to the antecedent, *A*, of the conditional. And if the conclusion is interpreted as a biconditional, that is, "*if A then C and if C then A*," then the suppositional strategy needs to be used twice, for example, once to show that the supposition of *A* yields the consequent, *B*, and once to show that the supposition of *C* yields the antecedent, *A*. None of our participants ever made such double suppositions.

There are two diagnostic signs of a suppositional strategy: Reasoners start by stating a supposition (or assumption), and they then derive a series of simple categorical conclusions, beginning by combining their supposition with a premise. Figure 11.2 is a typical protocol of a participant using a suppositional strategy.

Problem:
Pink iff black
Black ore gray
Gray iff blue
If not pink then blue?

| Verbal Protocol: | Drawing: |
|---|---|
| Pink iff black | |
| Black ore gray | |
| Gray iff blue | |
| If not pink then blue | |
| Assuming we have no pink | |
| There is no pink | Crosses out pink in printed premise |
| So there is no black | Crosses out black in both premises |
| There is gray | Circles gray |
| There is blue. | |
| Yes | |
| Not pink and blue | |
| Yes | |

FIG. 11.2. The suppositional strategy: a typical protocol. Figure 11.1 explains the abbreviations. As the protocol shows, the participant starts by reading the premises aloud.

The key phrase indicating the participant is making a supposition is, "*Assuming we have no pink*," which corresponds to the antecedent of the conditional conclusion. The participant then uses this assumption to draw a conclusion from the first premise. This step, in turn, leads to a further inference, and so on, in a series culminating in a conclusion corresponding to the consequent of the conclusion.

*The Compound Strategy.* Reasoners draw a conclusion from a pair of premises, or from one premise and a diagram of another premise, or from two diagrams representing separate premises. The conclusion is expressed either verbally or in the form of another diagram, or both. By combining such pair-wise inferences, reasoners derive the answer to the questioned conclusion. Figure 11.3 shows a typical protocol in which the reasoner combines the first two premises to yield an intermediate conclusion, and then combines this conclusion with the third premise to draw the final conclusion. One feature of the strategy is that even though neither the premises nor the conclusion to be evaluated made use of any modal terms, such as "*possibly*," the participants often drew a modal conclusion.

Problem:
White iff blue
If blue then pink
Pink ore brown
White or brown?

| Verbal Protocol: | Drawing: |
|---|---|
| ... | [draws a diagram of the first premise] |
| White if blue | Points to: blue → white |
| If blue then pink | |
| If blue then pink | Writes down premise |
| If pink then white [an intermediate conclusion] | Draws: pink → white |
| Pink ore brown | |
| Pink and white | Points to previous diagram |
| If brown then not white [conclusion] | Writes: brown, ~~white~~ |
| White ore brown | |
| Yes | |

FIG. 11.3. The compound strategy: a typical protocol. "If" in the premises refers to a "one-way" conditional, and "or" refers to an inclusive disjunction. The participant started by drawing a diagram for the first premise; we have omitted this stage from the protocol below.

When the participant draws the conjunctive conclusion, "*Pink and white*," in Fig. 11.3, he is really referring to a possibility rather than drawing a categorical conclusion. In other cases, however, the participants were quite explicit about the modal nature of their conclusions, for example:

*Red or else blue*
*Blue or else gray*
∴ *Possibly grey and red*

*The Chain Strategy.* This strategy is not one that we had encountered before, and we can find no mention of it in either the psychological or logical literature. The reasoner constructs a chain of conditionals leading from one constituent of a conditional conclusion to its other constituent. There is a resemblance to the suppositional strategy, but two crucial distinctions. First, reasoners do not announce that they are making an assumption. Indeed, they are not making an assumption, because they do not draw any intermediate conclusions. Second, they convert any premise that is not a conditional into a conditional, either verbally or in the form of a diagram representing a conditional. These conversions include cases where a biconditional, such as:

*Gray if and only if red*

yields the inference, either from the verbal premise or from a diagram representing it:

*If not gray, then not red*

---

Problem:
Gray iff red
Red or else white
White iff blue
If not gray then blue?

| Verbal Protocol: | Drawing: |
|---|---|
| . . . | [Draws separate diagrams for each premise] |
| | Points to diagrams: |
| If not gray then not red | r → g |
| If not red then white | r X w |
| White comes from blue | b → w |
| Yes | |

Problem:
Pink iff green
Green or else red
Red iff white
If not pink then white?

| Verbal Protocol: | Drawing: |
|---|---|
| . . . | [Draws separate diagrams for each premise]<br>Crosses out color terms in diagrams: |
| if not pink then not green | ~~pink~~ = ~~green~~ |
| if not g then r | ~~green~~ or red |
| if red then white | red = white |
| yes | |

---

FIG. 11.4. The chain strategy: two typical protocols. We have omitted the initial verbal protocol when the participants draw separate diagrams for each premise.

The aim is to make a chain in which the consequent of one conditional matches the antecedent of the next conditional. Figure 11.4 presents two typical examples of the chain strategy. The chain strategy is valid provided that reasoners construct a chain leading from the antecedent of the conditional conclusion to its consequent. However, reasoners often worked invalidly in the converse direction.

*The Model Strategy.* From our standpoint, the most interesting strategy was one in which the reasoners explicitly represented the possibilities compatible with the premises. As Fig. 11.5 shows, they drew a single integrated diagram that represented the possibilities. Some participants drew a vertical line down the page and wrote down the colors in the two possibilities on either side of it. Others arranged them horizontally. One participant, as Fig. 11.5 shows, merely drew circles around the terms in the premises themselves to pick out one of the two possibilities. Figure 11.1 gives a complete protocol of this strategy. A tell-tale sign of the model strategy is that the participants using it work through the premises in the order in which they are stated, and they include in the diagrams information from premises that are irrelevant to evaluating the conclusion.

---

Problem:

Blue or else brown
Brown or else white
White iff red
If blue then red?

Final diagram:

blue white red
brown

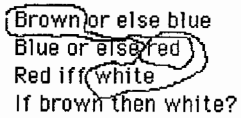
Problem:

Black or else pink
Pink or else grey
Grey iff white
If black then white?

Final diagram:

black | pink
white grey |

Problem and drawing:


Brown or else blue
Blue or else red
Red iff white
If brown then white?

---

FIG. 11.5. The model strategy: three typical protocols. We have presented only the premises and the final diagram that the participants drew (see Fig. 11.1 for a complete example of this strategy).

There were lines in the protocols that we could not understand, and there were also false starts and derivations that petered out. But, every participant correctly evaluated every problem, and we were able to categorize all their protocols into cases of one or more of the four strategies outlined. On the basis of the data, we worked out the relative proportions of the four sorts of strategies, that is, we calculated the total number of times each strategy occurred in the protocols, and then expressed them as percentages of the sum of the totals. The results were as follows:

*Suppositional strategy: 21% of overall use of strategies*
*Compound strategy: 19% of overall use of strategies*
*Chain strategy: 25% of overall use of strategies*
*Model strategy: 34% of overall use of strategies*
*Unknown strategies: 0% of overall use of strategies*

The most salient feature of the protocols was that the participants mix strategies, and switch from one strategy (compound) to another (chain) in ways that seem wholly unpredictable. Sometimes a switch occurs in the middle of a problem; sometimes from one problem to the next. Reasoners sometimes revert to a strategy that they used earlier in the experiment. They are plainly not following a single deterministic strategy of the sort postulated in current formal rule theories. Likewise, although the problems are all within the scope of sentential reasoning, the participants quite often draw intermediate conclusions that go beyond the scope of current formal rule theories. Sometimes, these conclusions are about possibilities. On other occasions, however, reasoners take a step that is difficult for formal rule theories to explain, for example:

*If red then not white*
*brown for white [where the participant points to a diagram of the form: brown → white]*
*∴ not (if red then brown)*

The most striking strategy, the model strategy, is entirely beyond the scope of rule theories. Yet, it was used at least one or more times by half the participants.

In our view, the four strategies are all entirely compatible with the use of models at the tactical level, and so we examine each strategy from this point of view. People reason from suppositions in many circumstances, and the strategy is compatible with the use of models. In an unpublished study carried out in collaboration with Victoria Bell, we asked the participants to draw possible conclusions or necessary conclusions from such suppositions as:

*Suppose everyone spoke the same language.*

As the model theory predicts, they were reliably faster to draw possible conclusions than to draw necessary conclusions. A key feature of our present experiment is that the participants did not make embedded suppositions; that is, having made one supposition, they did not make another before they had discharged the first. This lack of embedded suppositions may be because our problems could be solved without them, but we wonder whether logically naive individuals spontaneously embed one supposition within the domain of another (*pace* Rips, 1994). The compound strategy is also compatible with the use of models. In fact, some compound inferences can be explained at present only by the use of models, for example, those inferences yielding modal conclusions, which are beyond the scope of current formal rule theories (Braine & O'Brien, 1998; Rips, 1994). The chain strategy, likewise, can be accommodated within the model theory: The immediate inferences that convert disjunctions into conditionals, for example, could be based on models. Indeed, Richardson and Ormerod (1997; see also Ormerod, chap. 7, this volume) have studied how such conversions occur and argued that a version of the model theory gives a good account of them. Finally, the model strategy is isomorphic to the cumulative construction of a single set of models based on all the information in the premises.

## NONDETERMINISM AND THEORIES OF STRATEGIES AND TACTICS

A deterministic process is one in which each step depends solely on the current state of the process and whatever input, if any, it happens to have. Thus, a deterministic strategy unwinds in a fixed way like clockwork. The mind may be deterministic, but, as theorists, we have no option but to treat it nondeterministically, that is, our theories have to allow for different possible actions in the same theoretical state. This move is forced on us because we cannot predict precisely what will happen next in a piece of reasoning. We can put some constraints on the process from our observations of common patterns in inferential behavior, but the details are beyond the predictive power of our theories. Nondeterminism could merely reflect our ignorance: If we had a better understanding of the mind, then we would discern its deterministic nature. For example, the mind could be *chaotic* in the technical sense that its behavior is deterministic but soon becomes unpredictable like, say, the dripping of a tap. Another possibility is that the mind is genuinely nondeterministic, either because it can make arbitrary decisions or because its behavior is governed in part by quantum events. We must leave to future researchers the task of deciding amongst these competing interpretations.

One question we can answer, albeit speculatively, is why the mind appears to be nondeterministic from top to bottom. The answer is that there is a value in variation. Just as variation among individuals is a precursor to the evolution of species, so there is an advantage to variation in mental life. It yields novel and unpredictable thoughts, which are a prerequisite for learning and creativity. Lack of variation is equivalent to intellectual catatonia.

Granted the need for a nondeterministic theory, we need a precise and convenient way to express it. We offer a new way in which to couch such a theory, which depends on the following steps. First, the different theoretical possibilities are captured in a grammar. In the case of reasoning, we need a grammar of strategies in which each step calls on tactics of various sorts. Second, in implementing the theory in a computer program, these tactics must be modeled in explicit mechanisms for carrying out the appropriate inferential steps. Third, the computer program includes a parser that uses the grammar to parse the protocols. In our case, the grammar should allow each "think aloud" protocol to be parsed, and control each step in the corresponding tactics, such as drawing a diagram, or making an immediate inference. Hence, as the grammar is used to parse a protocol, the program will carry out the same inferential processes that the theory attributes to reasoners following that particular strategy. A good theory of strategies should account for all humanly possible strategies and, of course, for all the strategies observed in an experiment. A grammar is merely a parsimonious way in which to capture all the strategies and the unfolding of a particular sequence of tactical steps, from many possibilities, as a specific strategy is applied to a specific problem. Just as a grammar of a language embodies a theory of all the possible syntactic structures in the language, so a grammar of strategies should embody all the different tactical structures in all the possible strategies. Hence, a good theory of strategies will be one that parses *all* the protocols and carries out all their required tactical processes.

Two computational desiderata must be met by the resulting theory. First, the power of the grammar, its position in the Chomsky hierarchy, should be compatible with plausible assumptions about the operations of working memory. Second, the processes embodied in the theory must be comparable in tractability to those carried out by the mind. Theorists sometimes worry that a theory of mental processing postulates processes that are intractable; that is, as the input increases in size, so the process demands an increasing amount of time or memory, or both (see Oaksford & Chater, 1991). We know, however, that some apparently simple classes of inference are almost certainly intractable; for example, determining whether an inference about a possibility is valid takes a nondeterministic device an amount of time that is some polynomial of the length of the premises (it is, technically speaking, "NP" hard). And we also know that as such inferences increase in complexity, for example in the number of premises on which they are based, so the human inferential system collapses under the weight of the problem. Hence, a good theory of such inferences should also postulate mental processes that are intractable.

We illustrate the construction of a grammar for strategies by considering the way in which we have modeled the chain strategy. The major tactical steps in the strategy are as follows, with options shown in parentheses:

*1. Read each premise and grasp its meaning.*
*2. Draw a diagram based on the meaning of a premise.*
*(3. Check the diagram.)*

4. *Select a constituent proposition (antecedent or consequent) of the conditional conclusion.*
5. *Find a premise (or diagram) containing the constituent proposition.*
(6. *If the premise or diagram is not a conditional, make an immediate inference to a conditional with an antecedent that matches the constituent proposition.)*
7. *If the consequent matches the other constituent in the conclusion, or its negation, then the chain is complete. Otherwise, focus on the constituent proposition expressed by the consequent of the conditional, and continue from Step 5 above.*
8. *Evaluate the chain: if it reaches the other constituent of the conclusion, respond: Yes, the conclusion follows; if it reaches the negation of the other constituent of the conclusion, respond: No, the conclusion does not follow; otherwise, abandon the strategy.*

The full implementational details would overwhelm the reader, but here is a sketch of how the program works. It examines each item in a protocol to determine its tactical status, which includes the following cases, for example:

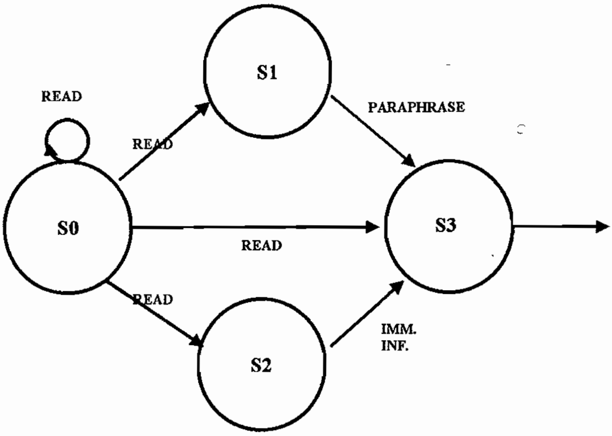
*READ the next premise and grasp its meaning*
*Make an IMMEDIATE-INFERENCE from a premise*
*Draw a DIAGRAM.*

This procedure yields an annotated version of the protocol, and it carries out all the required tactical steps, drawing diagrams, making inferences, and so on, as it proceeds through the protocol. Indeed, its ability to carry out these steps provides a check on the accuracy of its tactical assignments to each step in the protocol. We have assumed a so-called "regular" grammar of strategies. Such a grammar corresponds to a finite-state automaton, which is the simplest hypothesis about how strategies are generated. Finite-state automata do not require any working memory for intermediate results, and so in the Chomsky hierarchy they are the least powerful computational device capable of generating infinitely many sequences. (Of course, the program as a whole makes use of working memory as do human reasoners: Our assumption of a regular grammar concerns only the identification of tactical steps in a protocol.) In the grammar representing the chain strategy, each rule corresponds to a tactical step in the strategy. It specifies the state of the system by an arbitrary numerical label such as S0, the next tactical step, and the resulting state of the system after this step is taken, for example:

*S0 → read-premise S1*
*S1 → read-premise S2*
*S2 → immediate-inference S3*

Grammars are useful ways to control computer programs, but they are hard for readers to digest. Figure 11.6 shows a finite-state automaton and the equivalent grammar for drawing a diagram. As the figure shows, the system starts in state S0 and then has a choice of different routes.

Finite-state device:



Grammar:

READ S0
S0 → READ S1
READ S2
READ S3
S1 → PARAPHRASE S3
S2 → IMMEDIATE-INFERENCE S3
S3 → DRAW-DIAGRAM

FIG. 11.6. A nondeterministic finite-state device, and its corresponding grammar for reading a premise and drawing a diagram.

It reads a premise and may stay in the same state (S0)—so that it can read the premise repeatedly, or it jumps to a state (S1) where its next action is to paraphrase the premise, or to a state (S2) where its next action is to make an immediate inference from the premise, or to a state (S3) where its next action is to draw a diagram of the premise. Nothing in the automaton or grammar determines which of these routes is taken, and that is why the process is nondeterministic.

The representation of diagrams in our program is simple. One sort corresponds to individual premises, for example:

| | |
|---|---|
| *(draw a → b)* | *stands for drawing of one-directional arrow* |
| *(draw a ↔ b)* | *stands for two-directional arrow or '='* |

| | |
|---|---|
| *(draw a / b)* | *stands for some sort of separation between a and b, such as the drawing of separating circles round them* |
| *(draw a b)* | *stands for a grouping of a and b together* |
| *-a* | *stands for crossing out an a* |
| *+a* | *stands for ticking or affirming an a in some way.* |

---

<u>Problem</u>:
(a ore b)
(b ore c)
(c iff d))
(if a then d) Conclusion

| <u>Verbal Protocol</u>: | <u>Diagrams</u>: | <u>Tactical steps</u>: |
|---|---|---|
| Drawing initial diagrams: | | |
| (a ore b) | | Read-premise |
| (a or b) | | Immediate-inference |
| | (draw a / b) | Draw |
| (b ore c) | | Read-premise |
| | (draw b / c) | Draw |
| (iff d then c) | | Immediate-inference |
| (if d then c ) | | Immediate-inference |
| | (draw d → c) | Draw |
| (if d then c) | | Immediate-inference |
| | (show d → c) | Show-diagram |
| Checking previous steps, showing diagrams: | | |
| (a or b) | (show a / b) | Show-diagram |
| (b or c) | (show b / c) | Show-diagram |
| (b or c) | (show b / c) | Show-diagram |
| The chain strategy from d in conclusion: | | |
| (if d then c) | (show d → c) | Show-diagram |
| (if c then not b) | | Immediate-inference |
| | (draw c → - b) | Draw |
| (if c then not b) | (show c → - b) | Check previous two steps |
| (if not b then a) | | Immediate-inference |
| | (draw - b → a) | Draw |
| (if not b then a) | (show - b → a) | Check previous two steps |
| (if a then d) | | Read-evaluate-conclusion |
| (yes) | | Assert conclusion |

FIG. 11.7. An annotated protocol of the chain strategy in the format used by the computer program modeling the strategy. For simplicity, paraphrases of premises have been subsumed under the more general tactic of making an immediate inference.

```
Problem:
(a ore b)
(b ore c)
(c iff d))
(if a then d) Conclusion

The program's output:
Drawing initial diagrams:

(Read-premise (a ore b))
(Immediate-inference (a or b) from (a ore b))
(Diagram (draw a / b) from (a ore b))
(Read-premises (b ore c))
(Diagram (draw b/c) from (b ore c))
(Immediate-inference (iff d then c) from (c iff d))
(Immediate-inference (if d then c) from (c iff d))
(Diagram (draw d → c) from (c iff d))

. . . Checking previous steps, showing diagrams

The chain strategy from d in conclusion:
(Diagram (show d →_c) from (c iff d))
(Immediate-inference (if c then not b) from (b ore c))
(Diagram (draw c → - b) from (b ore c))
. . . checks two previous steps
(Immediate-inference (if not b then a) from (a ore b))
(Diagram (draw - b → a) from (a ore b))
. . . checks two previous steps                         Chain is complete.
(Read-conclusion (if a then d))
(Asserts-conclusion (Yes))

The parse was successful.
```

FIG. 11.8. A computer parse of the protocol in Fig. 11.7. We have omitted the program's output for the repetitions of certain steps. Everything within parentheses is an output of the program; our comments are on separate lines without parentheses.

Another sort of diagram is built up from the premises cumulatively. For example, a diagram that uses two horizontal lines to represent two possibilities (see Fig. 11.5), such as:

```
a        c          d
b
```

is represented by showing the cumulative steps in its construction:

```
(draw 1 a / b)
(draw 1 a c / b)
(draw 1 a c d / b)
```

where "1" is a number identifying the diagram. In addition to drawing diagrams, our participants often point to an existing diagram, particularly when they are checking a previous step. The program likewise can "show" a diagram, for example (show 1 a c d / b).

Figure 11.7 shows a complete "think aloud" protocol of a chain strategy in which we have substituted "*a*," "*b*," "*c*," and "*d*" for the propositions referring to the different colours. We have also added comments that label the tactical steps, and shown the diagrams drawn by the participant using the earlier notation. Our program can parse chain strategies and carry out the corresponding inferential processes. Figure 11.8 shows its output as it parses the "think aloud" protocol in Fig. 11.7. Our general goal, yet to be achieved, is for the program to deal with all four strategies.

## STRATEGIES AND TACTICS IN REFUTATIONS

We now turn to reasoning based on quantifiers, and to our attempts to identify the strategies and tactics of syllogistic reasoning. Syllogisms are logically simple inferences based on two premises, which each can be in one of four "moods":

*All the A are B* -- *abbreviated as the 'A' mood*
*Some of the A are B* -- *abbreviated as the 'I' mood*
*None of the A is a B* -- *abbreviated as the 'E' mood*
*Some of the A are not B* -- *abbreviated as the 'O' mood*

Here, are the premises of a typical syllogistic problem:

*Some of the chefs are musicians.*
*None of the musicians is a painter.*
*What conclusion, if any, follows?*

There are two controversies about syllogistic reasoning. The first is whether logically untrained individuals can reason at all syllogistically, or are merely selecting a conclusion that matches the mood of a premise (Chater & Oaksford, 1999; Martin Levine, personal communication, May 1994; Wetherick & Gilhooly, 1990). Granted that individuals can reason with quantifiers, the second controversy is whether they rely on formal rules of inference (Rips, 1994; Yang, Braine, & O'Brien, 1998); or on some sort of models, which may be Euler circles (Cardaci, Gangemi, Pendolino, & Di Nuovo, 1996; Fisher, 1981) or mental models (Johnson-Laird & Bara, 1984; Polk & Newell, 1995) or on both formal rules and Euler circles (Ford, 1995). The variety of theories confirms that even though syllogisms are logically simple, they are psychologically complex. What we aim to show is that none of these theories is quite right.

When we gave reasoners a paper and pencil and asked them to think aloud as they tackled syllogisms, the results were not too revealing, and so we had to devise a new procedure to help us to identify the participants' strategies (see Bucciarelli & Johnson-Laird, 1998). We gave them cut-out shapes to represent the premises, and we video-recorded them as they manipulated these shapes.

Our first study examined the competence of logically untrained individuals to search for counterexamples to putative conclusions. Part of our motivation was to check a computer program implementing the model theory (for an account of the program, see Bara, Bucciarelli, & Johnson-Laird, 1995). But, if people are unable to refute conclusions in this way, then Polk and Newell (1995) are correct in arguing that refutations play little or no role in syllogistic reasoning. The experiment was therefore designed to externalize the process of searching for counterexamples. The participants were given complete syllogisms, such as:

*Some of the chefs are musicians.*
*None of the musicians are painters.*
*∴ None of the chefs are painters.*

All the syllogisms referred to chefs, musicians, and painters, which were represented in cut-out shapes by chefs' hats, guitars, and palettes, respectively. The task was to add these shapes to six "stick" figures in order to construct an external model of the premises that refuted the putative conclusions. There were 20 syllogisms, four with valid conclusions, the remaining 16 with conclusions that could be correctly refuted by appropriate models of the premises.

The participants were able to construct counterexamples. The overall percentage of correct responses in cases where a conclusion could be refuted was 59%; and the overall percentage of correct responses in cases where a conclusion could not be refuted was 71%: Both results are much better than chance performance. Each participant was able to refute conclusions, and the range in performance was from 95% correct responses by the best participant to 25% correct responses by the poorest participant. This range in ability is quite characteristic of syllogistic reasoning. A major cause of error was that the participants often did not grasp what constituted a refutation of certain moods of conclusion. With conclusions in the O mood ("*Some of the A are not C*"), they often constructed a model in which some of the *A* were *C*, and they sometimes constructed a model in which none of the *A* were *C*. The correct counterexample calls for a model in which "*all the A are C*." Likewise, the participants occasionally thought that they had refuted a conclusion in the I mood "*Some of the A are C*," when they had merely constructed a model in which some of the *A* were not *C*. Such problems should be alleviated by expressing conclusions in the O mood using the logically equivalent form:

*Not all of the A are C*

because the participants appeared to refute an O conclusion by constructing a model of an assertion that omitted the term "not."

The participants varied in how they interpreted the four different moods of premises. Their preferred interpretation for first premises of the form, "*All the A are B*," was the co-extensive one in which each *A* is a *B*, and each *B* is an *A*. The result is an external model of the following form:

*a b*
*a b*

which represents two individuals who are both *A*'s and *B*'s. This interpretation was the preferred one for most sorts of syllogisms, but when the second premise was in the O mood, the participants were more inclined to build models of the *A* premise in which the *A*'s were properly included within the *B*'s:

*a b*
*a b*
*b*

Evidently, a second premise in the mood, "*Some of the B are not C*," helped reasoners to grasp that there could be *B*'s that are not *A*'s. Analogous patterns of influence occurred with other premises.

The most striking aspect of the results was the variety in the participants' strategies. The main strategies in a nutshell are as follows: The reasoners sometimes began by constructing a model of the first premise to which they added the information from the second premise; they sometimes proceeded in the opposite order. Sometimes, their initial model satisfied the conclusion, and so they modified the model in order to refute the conclusion; sometimes, they constructed an initial model of the premises that immediately refuted the conclusion. Here, to illustrate the variety of strategies, we summarize performance with the syllogism of the form:

*Some of the A are B.*
*None of the B are C.*
∴ *None of the A are C.*

Of the 20 participants, 16 correctly refuted the conclusion by constructing a model of the premises in which the conclusion was false. Five of these participants began by constructing a model of the premises that was consistent with the conclusion:

| | | |
|---|---|---|
| *a* | | |
| *a* | *b* | |
| | | *c* |

where we have ignored the actual numbers of tokens of each type that the participants constructed. Two of the five participants then refuted the conclusion by adding a $C$ to an $A$ (an operation that our computer program modeling the theory also carries out):

| | | | | | | |
|---|---|---|---|---|---|---|
| *a* | | | | *a* | | *c* |
| *a* | *b* | | *becomes* | *a* | *b* | |
| | | *c* | | | | *c* |

Another two of the five participants added an $A$ to a $C$ (which the program can also do):

| | | | | | | |
|---|---|---|---|---|---|---|
| *b* | *a* | | | *b* | *a* | |
| *b* | | | *becomes* | *b* | | |
| | | *c* | | | *a* | *c* |

The remaining participant of the five introduced a new $B$ and an $A$, and added a $C$ to the $A$:

| | | | | | | |
|---|---|---|---|---|---|---|
| *a* | *b* | | | *a* | *b* | |
| *a* | | | | *a* | *b* | |
| | | *c* | *becomes* | | *b* | |
| | | | | *a* | | *c* |
| | | | | | | *c* |

In contrast, 11 of the 16 participants who were correct refuted the conclusion in their initial model of the premises. Six of them did so with the following model:

| | | |
|---|---|---|
| *a* | | *c* |
| *a* | *b* | |

and three of them did so with the model:

| | | |
|---|---|---|
| *a* | *b* | |
| *a* | | *c* |
| | *b* | |

The other two out of the 11 built slight variants of the first of these models. Thus, the 16 participants reached a correct counterexample using at least five distinct procedures.

To what extent are these differences *strategic* as opposed to *tactical*? The distinction is harder to draw than in the case of sentential reasoning, but since we need a theory that encompasses both strategies and tactics, it may not be too important. In our view, the tactical steps are as follows:

*Read a premise and grasp its meaning.*
*Build a model (internal or external) based on a premise's meaning.*
*Add information based on a premise's meaning to an existing model.*
*Formulate a conclusion based on one or more models.*
*Evaluate a conclusion with respect to one or more models.*
*Construct an alternative model of the premises (in which the conclusion is false).*

Hence, some of the variation in the protocols for the previous example are different ways of carrying out a tactical step. For example, in seeking an alternative model of the premises, where a genuine alternative differs not merely in the numbers of tokens of different types, there are three main processes. Our computer program, as it happens, implements each of them. They are to add a token to one that is already in the model, to split an individual in the model into two separate tokens, and to join two separate individuals in the model into one. In principle, there is at least one other operation that reasoners could use: completely removing an individual from the model. Our participants only carried out this operation in special circumstances: They would occasionally remove an individual, only to restore the same individual immediately. The principle strategic variation in the task accordingly depends on two choices: what to model first, the first premise, the second premise, or the conclusion; and whether to model initially the conclusion or its refutation. In fact, the latter choice may only affect the external model. The internal model, we suspect, is likely to satisfy the conclusion at first, and then individuals can sometimes refute the conclusion as they an external model.

There are three major discrepancies between our program and the participants' performance. First, the program follows a deterministic strategy. Given a particular pair of premises, it always proceeds in the same way. Our participants, however, varied considerably in what they did, and they seemed likely to vary if they were to encounter the same problem twice (for evidence on this point, see Johnson-Laird & Steedman, 1978). Second, the program uses a fixed interpretation of the premises, whereas given a premise in a particular mood, our participants sometimes created one sort of model and sometimes another, a phenomenon that is much more in line with Polk and Newell's (1995) theory. Third, the program departs from human performance in its explicit representation of negation. Our participants, perhaps because they lacked any external symbols for negation, appeared to represent negatives only as "mental footnotes" on their external models.

Logically untrained individuals are able to construct external models of syllogistic premises, which are counterexamples to putative conclusions. This ability is beyond the explanatory scope of all current formal rule theories (Braine & O'Brien, 1998; Rips, 1994), which refute conclusions merely by failing to find formal derivations of them. Recursively speaking, our result is in turn a counterexample to formal rule theories. A critical issue, however, is whether individuals spontaneously use the strategy of searching for counterexamples when they have to draw syllogistic conclusions for themselves. In order to examine this issue, we carried out a further experiment.

## STRATEGIES IN SYLLOGISTIC REASONING

Our final experiment was designed to observe the external models that the participants built in drawing their own conclusions from syllogistic premises. For purposes of comparison, each participant also carried out the inferential task without being allowed to construct external models. The 20 participants drew their own conclusions from a set of 48 syllogisms (all the syllogisms in three of the four "figures" of the syllogism) in two conditions one week apart. Half the participants carried out the task first using external models and then without using them; and half the participants carried out the two conditions in the opposite order.

There was no reliable effect on accuracy of whether the participants constructed external models (51% correct overall) or not (55% correct overall). Likewise, there was no reliable difference between the first session of 48 syllogisms (51% correct) and the second session of 48 syllogisms (55% correct). However, the participants drew a slightly more diverse set of conclusions (a mean of 4.3 different conclusions to each problem) when they constructed external models than when they did not (a mean of 3.6 different conclusions).

Table 11.1.
The percentages of correct responses in the experiment on syllogistic reasoning in which the participants worked both with and without external models

| | *No external Model* | *External Model* |
|---|---|---|
| One-model with valid conclusion | 93 | 89 |
| Multiple model with valid conclusion | 29 | 21 |
| Multiple model with no valid conclusion | 60 | 57 |

Table 11.1 presents the percentages of correct responses to three sorts of syllogisms: those that have only one model and hence a valid conclusion (one-model syllogisms), those that call for multiple models to reach a valid conclusion (multiple-model syllogisms with a valid conclusion); and those that have multiple models without a conclusion in common (multiple-model syllogisms with no valid conclusion). In both experimental conditions, as the model theory predicts, the participants drew a greater percentage of correct conclusions to one-model problems than to multiple-model problems with (or without) valid conclusions.

In the external model condition, the participants constructed multiple models on 39% of trials, and all 20 participants built them, ranging from two participants who built such sequences on 75% of problems, down to one participant who built them on only 8% of the problems. The results corroborated a crucial prediction of the model theory: The participants were more likely to construct two or more models for the multiple-model problems with valid conclusions (37% of such problems) and with no valid conclusions (48% of such problems) than for one-model problems (11%). All 20 participants were in accord with this prediction. These percentages probably underestimate the construction of multiple models: When the participants constructed just a single model of multiple-model problems, they often made correct responses that were not consistent with that model, which suggests that they had considered an additional model of the premises in their mind's eye.

For the one-model problems, as we have remarked, the majority of conclusions were based on a single model. It is interesting to compare the models postulated in Polk and Newell's (1995) program with those constructed by our participants. As an example, consider the one-model problem based on the premises:

*Some B are A.*
*All B are C.*

Under one interpretation of the premises, Polk and Newell's program constructs the following model:

*b*        *c*
*b*   *a*   *c*

and then, as a result of reencoding the first premise, it constructs the model:

*b*        *c*
     *a*
*b*   *a*   *c*

from which it generates the valid conclusion: "*Some A are C.*" The most frequent response (9 participants) in our experiment was to construct just the first of these models (ignoring the number of tokens). Of the five participants who constructed multiple models, two constructed the foregoing sequence, although the first token constructed by all these participants was of the form: *b a c*.

The participants drew a less varied set of conclusions when they reasoned without the benefit of external models. The difference arises, we believe, because the reasoners constructed external models without being able to encode negative information explicitly. They then drew their conclusion based on the model without considering the presence or absence of negative tokens. Hence, they were more likely to draw a negative conclusion from affirmative premises, or an affirmative conclusion from negative premises, than when they reasoned without the benefit of external models. Yet, the participants' performance was of comparable accuracy whether or not they used external models, and the numbers of models had comparable effects in both experimental conditions (see Table 11.1).

Evidently, naive reasoners construct sequences of multiple models in drawing conclusions, especially from problems that support multiple models. This result was predicted by the model theory. Yet, the construction of multiple models is not necessarily equivalent to a search for counterexamples (see also Handley, Dennis, Evans, & Capon, chap. 12, this volume). Reasoners who construct more than one model may be just augmenting an initial model, but presumably they are not augmenting a model when, as often happened, they construct a model, modify it, revert to the original, modify it again, and so on. Likewise, participants must in general have been envisaging an alternative model when they constructed a single model from which they drew a correct response that was inconsistent with that model. But, again, we cannot be certain that refutation was the underlying motivation. Hence, a sequence of models is suggestive evidence, but no more, for the claim that reasoners are searching for counterexamples. It is bolstered, however, by the nature of the errors. Our analysis showed that there are three common causes of error. Some errors occur because reasoners overlook alternative models of the premises; some errors occur because reasoners construct the right set of models but assume that the models have nothing in common; and some errors occur because reasoners construct the right set of models but describe only one of them. To reach the right response to multiple-model syllogisms for the right reason, it is necessary to consider not just the initial models, but to search for alternatives, to grasp what, if anything, is common to all of them, and to describe it correctly.

Despite its successful predictions, the model theory and its computer implementation once again fail badly in accounting for the data. There were the same discrepancies as we observed in the previous experiment: The program uses a single deterministic strategy, it makes a single interpretation for each mood of the premises, and it represents negation explicitly. In constructing external models, our participants violated each of these principles.

## REPRESENTATIONS AND TACTICS IN SYLLOGISTIC REASONING

Our results count against all current theories of syllogistic reasoning. We showed in the previous section that they are incompatible with the present theory of mental models. We turn now to other accounts. It is impossible to prove that the processes postulated by a theory play no role in reasoning, and indeed many of these processes may occur, but what we can show is that no theory by itself can explain our results. The results of our experiments demonstrate that logically untrained individuals are able to reason from syllogistic premises. They are not merely generating conclusions in accordance with the "atmosphere" of the premises (*pace* Wetherick & Gilhooly, 1990) or selecting a conclusion that matches the form of the least informative premise (*pace* Chater & Oaksford, 1999). Granted that logically untrained individuals do reason, the principal controversy is whether they rely on formal rules of inference, or some form of mental model, or both.

Could it be that some reasoners do rely on formal rules? This claim is defended by Ford (1995). She argued that some of the participants in her study relied on a verbal substitution strategy. She classified participants as using this strategy if they spoke of replacing one term in a syllogism with another, or crossed out one term and replaced it with another. But, she also classified participants as using the strategy if they rewrote a syllogism as an equation or drew arrows between its terms (see Ford, 1995, fn 2, p. 18). This evidence may be consistent with a verbal strategy, but it is hardly decisive. Consider how the strategy is supposed to work: "*the subjects . . . take one premise as having a term that needs to be substituted with another term and the other premise as providing a value for that substitution*" (Ford, 1995, p. 21). Ford proposes a set of formal rules governing these substitutions. Apart from notational differences, Braine and Rumain (1983) have proposed the same rules. But, is the substitution procedure a purely verbal one dependent on formal rules of inference? And does the model theory, as Ford implies, group "*all people together as though they basically reason in the same fashion*" (Ford, 1995, p. 3)? Johnson-Laird and Bara (1984, p. 50) wrote: "*There are undoubtedly differences from one individual to another in the way in which they make syllogistic inferences. Our alternative implementations of the theory suggest a way in which some of these differences might be explained.*" Ironically, one of these alternatives was a substitution procedure based on models rather than verbal premises, that is, one token is substituted for another in a model of the premises (see also Johnson-Laird, 1983, p. 106). In fact, no evidence shows that the substitution strategy is purely verbal as opposed to based on mental models.

Current theories based on formal rules (e.g., Braine & O'Brien, 1998; Rips, 1994) postulate such rules for syllogisms as:

*All X are Y.*
*All Y are Z.*

∴ *All X are Z.*

The drawback of these rule systems is that models play no part in them, and so they are unable to explain the ability of our participants to construct external models, or to establish models that are counterexamples to putative conclusions. However, Rips's system has the power of a Universal Turing machine, and so it can be used as a programming language in which to implement any theory, including the mental model theory. His theory in this general sense is thus almost irrefutable, that is, no empirical results could ever show it to be false unless they demonstrated that mental processes are not computable (Johnson-Laird, 1997). But, formal rules in a narrower sense cannot explain how people are able to construct external models or to refute conclusions by constructing counterexamples to them. So let us turn to a competing model-based theory.

Euler circles represent each set referred to in a premise by a circle, and they represent the relation between the two sets by a simple topological relation between the circles. Hence, a premise of the form "*All the A are B*" calls for two separate representations: In one, the circle representing *A* lies wholly within the circle representing *B*, that is, the set *A* is properly included with set *B*; and in the other, the two circles coincide, that is, the two sets are co-extensive. Analogous topological relations between the two circles represent premises in the other moods. The traditional use of Euler circles calls for the construction of all the different diagrams for each premise, and all the different combinations for the pair of premises, a demand that leads to a combinatorial explosion (see Erickson, 1974). Stenning and his colleagues have devised a novel way to use Euler circles that obviates this explosion (see Stenning & Yule, 1997). Ford (1995) postulates a similar procedure: Reasoners assume that areas enclosed by circles can be empty, and they use the verbal premises as reminders of which areas cannot be empty. This procedure, as Ford allows, is equivalent to the use of optional elements in models. Hence, the main burden of Stenning's analysis and Ford's results is that reasoners do not use the traditional method of Euler circles. These authors, however, give no account of the sequences of models that reasoners construct or of the operations that they use to generate such sequences.

There is no doubt that some people rely on Euler circles. But, do individuals who have never seen circles used to represent sets spontaneously use Euler circles? As far as we know, no logician prior to Leibniz used circles to represent sets. The idea was a major innovation, and it was later popularized by Euler's letters to a Swedish princess. If naive individuals spontaneously use the method, why wasn't it invented earlier and why did it have to be popularized? The major disadvantage of Euler circles, however, is that they do not generalize to relational inferences, such as the following example (see Russell, 1946):

*All horses are animals.*

∴ *All horses' heads are animals' heads.*

In contrast, mental models represent finite sets of entities by finite sets of mental tokens, and they readily accommodate relations among entities (see Johnson-Laird, 1983). Ford (1995) appears to take for granted that because some of her participants drew Euler circles, it follows that these individuals were not using mental models. She writes: "*Thus, the spatial subjects used a type of representation specifically dismissed by Johnson-Laird and his colleagues, where the class itself and not the finite members of the class is represented*" (p. 41). Readers should note the equivocation in this claim. Ford is referring to the external representations drawn by her participants; Johnson-Laird and his colleagues are referring to internal mental representations. Moreover, contrary to Ford, some of her participants whom she classified as verbal reasoners did refer to individual entities, as the following extracts from four protocols show:

*i.* *. . . if there are any historians like suppose there's two historians right that means there are two weavers who are also historians so we can say some of the weavers are historians . . . (Eric)*
*ii.* *. . . could have a weaver that is not a historian and is a TC member (Catherine)*
*iii.* *. . . all of the historians are weavers none of the historians well you actually can't conclude that because you have another some one else like a philosopher who could be a weaver who might be a tennis club member . . . (Hilary)*
*iv.* *. . . if you're a playwright you're always a bookworm that means you have a chance to be a stamp collector . . . (Amy)*

Although some individuals sometimes draw Euler circles when they make syllogistic inferences, we incline to Rips' (1994) view that they rely on a vestigial memory for a procedure that they encountered in school. Euler circles, however, are a legitimate hypothesis about the nature of mental models. We do not know whether those who draw Euler circles use visual images of them either to control their drawings or to reason when they have no access to paper and pencil. But, we do know that they are not powerful enough for reasoning with relational premises, and that current psychological theories based on them cannot account for the participants' tactics in our experiments.

## CONCLUSIONS

What are the main features of the reasoning strategies of logically untrained individuals? The answer is that such individuals develop strategies that are appropriate to the inferences on which they are working. For sentential reasoning, they adopt quite distinct strategies, that are compatible with the use of models at the tactical level, and which allow them to keep track, one way or another, of the relevant possibilities. The suppositional strategy pursues the consequences of one possibility, the one created by the supposition. The chain strategy likewise pursues a possibility that leads from one constituent of the conclusion to the other. The compound strategy combines pairs of premises to infer what is necessary or possible. The model strategy keeps track of all the possibilities.

In contrast, naive reasoners appear to find it impossible to envisage all the possible models of quantified premises. When reasoners are allowed to construct external models, their preferred strategies explore alternative models of the premises, though they often err by overlooking a possibility. Are they searching for counterexamples to conclusions? Certainly, they can do so when they are asked explicitly to refute conclusions. But, Polk and Newell (1995) argued that syllogistic reasoning depends on encoding and reencoding premises as mental models rather than on a search for counterexamples. They support their claim by showing that "falsification" yields little improvement in the fit of their computer program, VR, to the data. We suspect that there is little improvement because VR does some of the work of refutation in other ways. What is right about their theory, however, is its emphasis on the variety of different interpretations of the premises. What appears to be wrong is the sequences of models generated by successive reencodings of the premises. Our participants tended not to reencode the premises for one-model problems (in contrast with VR), but rather to generate sequences in the case of multiplemodel problems. In our study of syllogistic reasoning with external models, the participants generated sequences of alternative models, but whether they were searching for counterexamples is unclear (see Handley et al., chap. 12, this volume).

With hindsight, syllogisms are not an ideal test case for demonstrating a search for counterexamples. Modal reasoning is better, because the model theory predicts an obvious interaction that hinges on reasoners searching for counterexamples: It should be easier to determine that a situation is possible (one model of the premises suffices as an example) than necessary (all the models of the premises must be checked), whereas it should be easier to determine that a situation is not necessary (one model serving as a counterexample suffices) than not possible (all models must be checked). The interaction has been corroborated in reasoning both from sentential connectives (Bell & Johnson-Laird, 1998) and quantifiers (Evans, Handley, & Harper, 1998; see also Galotti, Baron, & Sabini, 1986). Hence, Polk and Newell may be right about syllogisms, but, in those tasks where counterexamples are of obvious use, reasoners appear to search for them. Indeed, as Barwise (1993) emphasized, the only way to *know* that a conclusion is invalid is by constructing a model of the premises that is a counterexample to it.

A complete theory of thinking calls for four levels. At the top level is metacognition and the sort of insightful thinking that leads to the development of a new strategy. Its manifest signs are lacking in the protocols from our experiments. Noone ever remarked, for example, "*I see now how I can solve these problems efficiently*," and then went on to describe an insightful strategy. None of our participants was an Aristotle! Metacognitive remarks, however, might be observed in other circumstance, such as a study in which the participants are explicitly instructed to develop efficient strategies and to describe them to other people. The second level is the thinking that controls a strategy. Its signs in our protocols are a sequence of organized remarks, and diagrams in the sentential reasoning experiment, from which one can infer the strategy that the reasoner was following. At the third level are the tactics from which a strategy is composed, such as drawing a diagram of a premise, or adding information from a premise to an existing model. At the fourth, and lowest level, are the processes that underlie the tactics, that is, the largely unconscious processes that support, say, making an immediate inference, or using the meaning of a premise to control a drawing. These unconscious processes are perhaps comparable to the "instruction set" of a computer chip. Is all thinking analyzable in these terms? We conjecture that goal-driven thinking is open to metacognitive insights, governed by a strategy, and depends on tactics, which in turn rely on unconscious processes.

The strategies and tactics of reasoning call for a nondeterministic theory, which, as we have illustrated, can take the form of a grammar that is used to parse reasoners' protocols. The strategies and tactics that we have observed both in sentential and quantified reasoning are not easily reconciled with formal rules of inference, but they do seem to be compatible with the unconscious processes that construct and manipulate mental models. The study of strategies in reasoning, unlike strategies in other domains (e.g., Lemaire & Siegler, 1995), has barely begun. There are three pressing goals. Future studies should delineate the "space" of possible strategies, and their effectiveness and efficiency. They should account for the sequences of strategies that reasoners pass through as they gain experience and expertise. Logic, one could say, is the ultimate strategy that some highly gifted individuals attain. But the "Holy Grail" for future research is the discovery of how logically-untrained individuals discover new strategies of reasoning.

## ACKNOWLEDGMENTS

We thank our colleague Walter Schaeken for organizing the workshop at which this chapter was presented and for his advice over the years. We are also grateful for the help of many colleagues, including Bruno Bara, Ruth Byrne, Vittorio Girotto, Paolo Legrenzi, and Yingrui Yang. The research was supported in part by ARPA (CAETI) contracts N66001-94-C-6045 and N66001-95-C-8605.

## REFERENCES

Bara, B.G., Bucciarelli, M., & Johnson-Laird, P. N. (1995). The development of syllogistic reasoning. *American Journal of Psychology, 108,* 157–193.

Barwise, J. (1993). Everyday reasoning and logical inference (Commentary on Johnson-Laird and Byrne, 1991). *Behavioral and Brain Sciences, 16,* 337–338.

Bell, V., & Johnson-Laird, P. N. (1998). A model theory of modal reasoning. *Cognitive Science, 22,* 25–51.

Braine, M. D. S., & O'Brien, D. P. (1991). A theory of If: A lexical entry, reasoning program and pragmatic principles. *Psychological Review, 98,* 182–203.

Braine, M. D. S., & O'Brien, D. P. (Eds). (1998). *Mental logic.* Mahwah, NJ: Lawrence Erlbaum Associates.

Braine, M. D. S., & Rumain, B. (1983). Logical reasoning. In J. H. Flavell & E. M.

Markman (Eds.), *Carmichael's handbook of child psychology, Vol. III. Cognitive Development* (pp. 240–263). New York: Wiley.

Bucciarelli, M., & Johnson-Laird, P. N. (1998). *Strategies in syllogistic reasoning.* Unpublished manuscript, Department of Psychology, Torino University.

Byrne, R. M. J., & Handley, S. J. (1997). Reasoning strategies for suppositional deductions. *Cognition, 62,* 1–49.

Cardaci, M., Gangemi, A., Pendolino, G., & Di Nuovo, S. (1996). Mental models vs. integrated models: Explanations of syllogistic reasoning. *Perceptual and Motor Skills, 82,* 1377–1378.

Chater, N., & Oaksford, M. (1999). Rational analysis and heuristic processes for syllogistic reasoning. *Cognitive Psychology, 2,* 191–258.

Erickson, J. R. (1974). A set analysis theory of behaviour in formal syllogistic reasoning tasks. In R. Solso (Ed.), *Loyola Symposium on Cognition, Vol. 2* (pp. 305–330). Hillsdale, NJ: Lawrence Erlbaum Associates.

Evans, J. St. B. T., Handley, S. J., & Harper, C. (1998). *Deductions about what is necessary and what is possible: A test of the mental model theory.* Unpublished manuscript, Department of Psychology, University of Plymouth.

Evans, J. St. B. T., Newstead, S. E., & Byrne, R. M. J. (1993). *Human reasoning: The psychology of deduction.* Mahwah, NJ: Lawrence Erlbaum Associates.

Fisher, D. L. (1981). A three-factor model of syllogistic reasoning: The study of isolable stages. *Memory and Cognition, 9,* 496–514.

Ford, M. (1995). Two modes of mental representation and problem solution in syllogistic reasoning. *Cognition, 54,* 1–71.

Galotti, K. M., Baron, J., & Sabini, J. P. (1986). Individual differences in syllogistic reasoning: Deduction rules or mental models? *Journal of Experimental Psychology: General, 115,* 16–25.

Girotto, V., Mazzocco, A., & Tasso, A. (1997). The effect of premise order in conditional reasoning: A test of the mental model theory. *Cognition, 63,* 1–28.

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

Johnson-Laird, P. N. (1997). Rules and Illusions: A critical study of Rips's The Psychology of Proof. *Minds and Machines, 7,* 387–407.

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

Johnson-Laird, P. N., & Byrne, R. M. J. (1990). Meta-logical problems: Knights, knaves, and Rips. *Cognition, 36,* 69–81.

Johnson-Laird, P. N., & Byrne, R. M. J. (1991). *Deduction.* Hillsdale, NJ: Lawrence Erlbaum Associates.

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

Lemaire, P., & Siegler, R. S. (1995). Four aspects of strategic change: Contributions to children's learning of multiplication. *Journal of Experimental Psychology: General, 124,* 83–97.

Oaksford, M., & Chater, N. (1991). Against logicist cognitive science. *Mind & Language, 6,* 1–38.

Polk, T. A., & Newell, A. (1995). Deduction as verbal reasoning. *Psychological Review, 102,* 533–566.

Richardson, J., & Ormerod, T. C. (1997). Rephrasing between disjunctives and conditionals: Mental models and the effects of thematic content. *Quarterly Journal of*

*Experimental Psychology, 50A,* 358–385.

Rips, L. J. (1989). The psychology of knights and knaves. *Cognition, 31,* 85–116.

Rips, L. J. (1994). *The psychology of proof.* Cambridge, MA: MIT Press.

Russell, B. A. W. (1946). *History of Western philosophy.* London: Allen & Unwin.

Stenning, K., & Yule, P. (1997). Image and language in human reasoning: A syllogistic illustration. *Cognitive Psychology, 34,* 109–159.

Wason, P. C., & Johnson-Laird, P. N. (1972). *Psychology of reasoning: Structure and content.* London: Batsford.

Wetherick, N. E., & Gilhooly, K. J. (1990). Syllogistic reasoning: Effects of premise order. In K. J. Gilhooly, M. T. G. Keane, R. H. Logie, & G. Erdos (Eds.), *Lines of thinking, Vol. I* (pp. 99–108). New York: Wiley.

Yang, Y., Braine, M. D. S., & O'Brien, D. P. (1998). Some empirical justifications of the mental predicate logic model. In M. D. S. Braine & D. P. O'Brien (Eds.), *Mental logic* (pp. 333–365). Mahwah, NJ: Lawrence Erlbaum Associates.

# DEDUCTIVE REASONING AND STRATEGIES

Edited by

Walter Schaeken
*University of Leuven*

Gino De Vooght
André Vandierendonck
*University of Ghent*

Géry d'Ydewalle
*University of Leuven*

2000
LAWRENCE ERLBAUM ASSOCIATES, PUBLISHERS
Mahwah, New Jersey London

The final camera copy for this work was prepared by the editors, and therefore the publisher takes no responsibility for consistency or correctness of typographical style. However, this arrangement helps to make publication of this kind of scholarship possible.

Lawrence Erlbaum Associates, Inc., Publishers
10 Industrial Avenue
Mahwah, New Jersey 07430

Cover design by Kathryn Houghtaling Lacey

**Library of Congress Cataloging-in-Publication Data**

Deductive reasoning and strategies / edited by Walter Schaeken ... [et al.].
p. cm
Includes bibliographical references and indexes.
ISBN 0-8058-3238-6 (cloth : alk. paper)
1. Reasoning (Psychology) I. Schaeken, Walter.

BF442.D43 1999
153.4'3 21--dc21 99-043475

Books published by Lawrence Erlbaum Associates are printed on acid-free paper, and their bindings are chosen for strength and durability.

Printed in the United States of America
10 9 8 7 6 5 4 3 2 1