Mental Logic
List of Contributors

Luca Bonatti, Laboratoire des Science Cognitives et Psycholinguistique, Paris France
Martin D.S. Braine, Department of Psychology, New York University
Shalom M. Fisch, Department of Psychology, New York University
Joshua B. Cantor, Long Island University
Maria G. Dias, Department of Psychology, Federal University of Pernambuco, Recife, Brazil
Joseph R. Hosie, Department of Psychology, Baruch College of the City University of New York
R. Brooke Lea, Department of Psychology, Bowdoin College, Brunswick, Maine
Ira A. Noveck, Centre de Recherche en Epistémologie Appliquée, Paris, France
David P. O’Brien, Department of Psychology, Baruch College and the Graduate School of the City University of New York
Guy Politzer, Centre National de la Recherche Scientifique, Saint Denis, France
Brian J. Reiser, Department of Psychology, New York University
Antonio Roazzi, Department of Psychology, Federal University of Pernambuco, Recife, Brazil
Barbara Rumain, Department of Psychology, New York University
Mark C. Samuels, Department of Psychology, New York University
Yingrui Yang, Department of Psychology, New York University, and Department of Psychology, Princeton University, Princeton, New Jersey
Mental Logic

Edited by

Martin D.S. Braine
New York University

and

David P. O’Brien
Baruch College and the Graduate School
of the City University of New York

LAWRENCE ERLBAUM ASSOCIATES, PUBLISHERS
1998 Mahwah, New Jersey London
Contents

1. Introduction: Some Background to the Mental-Logic Theory and to the Book
   David P.O’Brien  
   1

2. Why It Took So Long to Bake the Mental-Logic Cake: Historical Analysis of the Recipe and Its Ingredients
   Luca Bonatti  
   7

3. Mental Logic and Irrationality: We Can Put a Man on the Moon So Why Can’t We Solve Those Logical Reasoning Problems?
   David P.O’Brien  
   23

4. How to Investigate Mental Logic and the Syntax of Thought
   Martin D.S.Braine and David P.O’Brien  
   45

5. Logical Inferences and Comprehension: How Mental-Logic and Text Processing Theories Need Each Other
   R.Brooke Lea  
   63

6. The Theory of Mental-Propositional Logic: Description and Illustration
   Martin D.S.Braine and David P.O’Brien  
   79
7. Evidence for the Theory: Predicting the Difficulty of Propositional Logic Inference Problems
   Martin D.S. Braine, Brian J. Reiser, and Barbara Rumain

8. Further Evidence for the Theory: Predicting Intermediate and Multiple Conclusions in Propositional Logic Inference Problems
   Martin D.S. Braine, David P. O’Brien, Ira A. Noveck, Mark C Samuels, R. Brooke Lea, Shalom Misch, and Yingrui Yang

9. A Theory of If: A Lexical Entry, Reasoning Program, and Pragmatic Principles
   Martin D.S. Braine and David P. O’Brien

10. Conditional Reasoning: The Logic of Supposition and Children’s Understanding of Pretense
    David P. O’Brien, Maria G. Dias, Antonio Roazzi, and Martin D.S. Braine

11. Steps Toward a Mental-Predicate Logic
    Martin D.S. Braine

12. Some Empirical Justification of the Mental-Predicate-Logic Model
    Yingrui Yang, Martin D.S. Braine, and David P. O’Brien

13. Leveling the Playing Field: Investigating Competing Claims Concerning Relative Inference Difficulty
    Ira A. Noveck and Guy Politzer

    David P. O’Brien, Maria G. Dias, and Antonio Roazzi

15. A Dubious Premise: Suppressibility of Modus Ponens
    David P. O’Brien, Maria G. Dias, and Joseph R. Hosie

16. What the Mental Logic-Mental Models Controversy Is Not About
    Luca Bonatti
17. Pinocchio’s Nose Knows: Preschool Children Recognize That a Pragmatic Rule Can Be Violated, an Indicative Conditional Can Be Falsified, and That a Broken Promise is a False Promise

David P. O’Brien, Maria G. Dias, Antonio Roazzi, and Joshua B. Cantor

References

Author Index

Subject Index
Introduction: Some Background to the Mental-Logic Theory and to the Book

David P.O’Brien

Baruch College and the Graduate School of the City University of New York

I entered graduate school with the conviction that a mental logic is a basic part of what allows humans to have coherent experiences, although exposure to modern logic had led me to believe that this mental logic is not the classical logic that Kant had included among the a priori categories. I had thus started looking through the literature to find a reasonable theoretical account of this mental logic. The most prominent mental-logic theory at that time was Piaget’s, but his logic seemed confused and did not bear any obvious relation to the sorts of reasoning tasks he presented.¹ The most promising approach that I found was the “natural logic” of the logician Gentzen (1935/1964), but the attempts at a psychological realization of this approach (Johnson-Laird, 1975; Osherson, 1975a, 1976) were clearly initial steps. In 1978, Braine published his Psychological Review article, “On the Relation Between Standard Logic and the Natural Logic of Thought.” After reading the article I called Marty and arranged to meet him. Our meetings continued until his death in April 1996, and during these years we worked together to develop the mental-logic theory presented in this volume.

Marty’s article, as well as my first meetings with him, convinced me that he was bringing something to this search that I was not finding elsewhere. Marty

¹ Piaget asserted, for example, that formal-operational structure is equivalent to the 16 propositional operators of normal-disjunctive form (e.g., Inhelder & Piaget, 1958). The formal-operational tasks, however, present objects and their attributes that require quantification that cannot be captured by a logic at the propositional level.
already was established as a leading investigator of language acquisition. As Melissa Bowerman (personal communication) said to me, Marty approached children’s speech like a linguist studying an unknown language—he identified regularities in collocation and proposed accounting for these in terms of a simple set of grammatical rules (including his influential pivot grammar). He brought the same inclinations and talents to the issues of reasoning, and he conveyed to me two basic insights that seemed to be obviously right. First, although logicians have provided myriad systems, their interests have not been psychological. Indeed, the sorts of things logicians provide, such as metalogical proofs of completeness, or construction of a parsimonious set of operators, are often of little benefit to someone investigating mental logic. Although a theory of mental logic will bear an interesting relation to standard logic, for example, including sound schemas, it need not be the same as any particular system presented in the logic literature, and its basic parts need not have the power to derive all the theorems of some standard system. Thus, although one needs to attend to what logicians have discovered about logic, the psychological researcher should begin by observing the ways people use natural-language particles, such as if, and, not, or, some, all, none, any, and so forth, in natural speech. Second, we should not be seduced by our training in experimental methods, which can lead us to look for differences and deflect us from looking for universals. Linguistics, for example, would not have made progress if it had focused on differences; it was in looking for universals that it found its way. The mental logic theory presented in this volume reflects these two concerns: We have sought to discover what is basic about the ways these logic particles are used in speech and in reasoning.

Across the past 10 years, the question of whether there is a mental logic has become subject to considerable debate. There have been attacks by critics who believe that all reasoning uses mental models (e.g., Johnson-Laird & Byrne, 1991, 1993; Johnson-Laird, Byrne, & Schaeken, 1992, 1994) and return attacks on mental-models theory (e.g., Braine, 1995; O’Brien, Braine, & Yang, 1994). This controversy prominently invaded journals such as Psychological Review, Behavioral and Brain Sciences, Cognition, and the Journal of Experimental Psychology: General. There have also been issues between mental logic and the biases-and-heuristics approach to reasoning (as represented, e.g., by Evans, 1989), and the content-dependent theorists (e.g., Cheng & Holyoak, 1985; Holyoak & Cheng, 1995). However, despite its pertinence to current issues in cognition, few cognitive scientists really know what the mental-logic theory is, and misapprehensions are rife. Actually, there have been two mental-logic theories, ours and another developed by Rips (1994). The two mental-logic theories are similar to one another in some respects but are significantly different in others. The question of whether there
is a mental logic thus is different from the question of what that mental logic is, although this distinction often has been lost in much of the debate.

The theory of mental logic presented here has three parts. One part is the mental logic per se, and it consists of a set of inference schemas of the sort first proposed by Gentzen. For example, when one knows that two propositions of the form $p \lor q$ and $\neg p$ are true, one can assert $q$. We take the issue of which schemas to include to be empirical; for example, the schema that derives $p \lor q$ from the premise $p$, which is found in most logic textbooks, is not among our proposed schemas because the evidence does not support its inclusion (see chap. 7). The set of basic schemas that are included is not intended to be an exhaustive inventory of all of the sorts of inferences people make, but to be an inventory of those logic inferences that are made routinely.

The second part of the theory is a reasoning program that applies the schemas in lines of reasoning, and it includes both a direct-reasoning routine and some more sophisticated indirect-reasoning strategies. The direct-reasoning routine is claimed to be universally available and applied with minimal effort both in reasoning and in comprehension. The most basic prediction of the theory thus is that inferences that are made by the application of basic schemas with the direct-reasoning routine (the basic part of the theory) will be made routinely. Although there is evidence for availability of the more sophisticated reasoning strategies among college students, and presumably they often are available in other populations (see chap. 7), reasoning that requires such strategies is proposed as neither universal nor effortless, and inferences that require this part of the reasoning routine should be made far less often than those that are available on the basic part of the theory.

The third part of the theory is pragmatic. Our theory proposes that the basic meaning of each logic particle is in the inferences that are sanctioned by its inference schemas. How any particular proposition is construed, however, may go beyond what is sanctioned by the schemas. We discuss several pragmatic principles that account for some of the judgments that are beyond the basic mental logic (e.g., chap. 9), although these principles are not intended to provide a complete inventory of what would be required to account for all of the extralogical processes involved in comprehension and reasoning. A complete account would go well beyond the scope of our theory, and indeed would require not only a complete theory of pragmatics but would constitute a large part of a general theory of cognition. For example, it would need to include the processes by which information is stored in, and retrieved from, long-term memory.

Several criticisms of our approach have stemmed from misunderstandings about its pragmatic part. Some commentators have held, for example, that a mental logic does not account for the influences on reasoning of various
sorts of content (e.g., Holyoak & Cheng, 1995); such criticisms ignore the fact that our theory includes a pragmatic aspect. Others have claimed that the pragmatic part of our theory has been added on merely as an afterthought (e.g., Evans & Over, 1996); this criticism overlooks our argument that because logic refers to propositions rather than to sentences, the mental logic is embedded in a profoundly pragmatic architecture (see, in particular, chaps. 3 and 9 for discussions).

This volume was planned as a comprehensive presentation of the theory of mental logic we have worked on over the past 20 years and the evidence for it, together with its implications for cognition and development, including the acquisition of language. It includes both revisions of articles that were published previously, as well as chapters containing unpublished material. Its unity and focus come from the comprehensive presentation of the theory. Other theories are, of course, discussed. Rips is discussed in chapters 8 and 11, and chapters 13, 14, and 15 address nonlogical theories, especially the mental-models theory.

The first section presents four introductory chapters on mental logic. Chapter 2 is written by the philosopher Luca Bonatti, who has written on mental logic in Psychological Review (1994) and Cognition (1994). Bonatti discusses the development of mental logic as a scientific theory and addresses why an idea that has been available for 2,000 years is only recently becoming scientifically fruitful. Chapter 3 is a revision of an article by David O’Brien that addresses many misunderstandings of our mental-logic theory. Originally published as part of a volume about human rationality, this chapter addresses how a person with a mental logic often makes irrational judgments. This chapter also addresses some of the implications of the propositional nature of a mental logic.

Chapter 4, by Braine and O’Brien, is a revision of an article published previously that further discusses the relevance of pragmatics to mental logic and shows how mental logic is related to the notion of a syntax or language of thought, and via the syntax of thought, to language acquisition. Chapter 5 is by R. Brooke Lea, an assistant professor at Bowdoin College and a former member of the Braine laboratory. This chapter describes Lea’s work on how people spontaneously make logical inferences as they read text, and by implication, discourse; that mental logic plays an important role in discourse is a central claim of Sperber and Wilson (1986) in their relevance theory. Lea also addresses the mutual needs of mental logic and comprehension researchers to attend to one another’s findings, particularly if one wants to make online predictions.

The next five chapters deal with the most developed part of the theory—mental propositional logic—the part on which most empirical work has been done and the part for which there is the most supporting evidence. Chapter
Some Background to the Mental-Logic Theory

6 presents the most recent version of the mental propositional logic with considerable illustration. Chapters 7 and 8 are reprints (with considerable revision) of the major evidence supporting the theory at the propositional level, and they discuss other theories in relation to the data; in particular, chapter 8 discusses how Rips’s logic and Johnson-Laird’s mental-models theories fare in relation to the evidence from the experiments. Chapter 9, reprinted from Psychological Review, presents our theory of if and compares it with other theories, from both the psychological and philosophical literatures. Chapter 10, written with Maria Dias and Antonio Roazzi of the Federal University of Pernambuco in Brazil, presents a mental-logic view of children’s conditional reasoning, including how they reason towards an if statement and how they work out the consequences of suppositions made in the course of pretend play. We expect this work will be of interest to developmentalists concerned with theory of mind or with pretend play, as well as to reasoning researchers.

Chapter 11, by Marty Braine, presents a theory for a mental predicate logic and compares its predictions with those of Rips’s logic. Chapter 12, by Yang, Braine, and O’Brien, presents an initial empirical test of the theory, using methods similar to those presented in the empirical test of the propositional-logic theory in chapter 7.

The last section of the book presents comparisons of our mental-logic theory with the mental-models theory of Johnson-Laird. Chapter 13, by Ira Noveck and Guy Politzer, presents some experiments comparing the mental-models with the mental-logic treatment of or, and Chapter 14, by O’Brien, Maria Dias, and Antonio Roazzi, presents a comparison between the mental-logic and mental-models treatment of if, using conditional syllogisms to illustrate the difference. Chapter 15, by O’Brien, Maria Dias, and Joseph Hosie, the latter a colleague of O’Brien’s at Baruch College, addresses a controversy about the suppressibility of modus ponens, begun by Ruth Byrne (1989) in an article in Cognition, to which Politzer and Braine (1991) responded. In addition to discussing the issue, it presents some new data that reveal that Byrne’s methods can suppress premises when no inference at all is made, bringing into doubt whether inferences are suppressed at all, or whether the premises are. Chapter 16, by Bonatti, addresses some further issues raised in the mental-models versus mental-logic debate. Finally, chapter 17, by O’Brien, Dias, Roazzi, and Joshua Cantor, presents some new data to answer the claim made in criticism of mental logic that the mind uses some content-specific reasoning processes.

Marty Braine died shortly before the book was completed, and I have finished those parts of the book on which he was working that were unfinished at that time. In doing so, I made as few changes as possible in the material on which Marty was working. When Marty realized that his illness was terminal, he expressed to me his frustration with spending his remaining time and
energy debating our critics—he preferred to spend his energies developing the mental-logic approach. I made him the following offer: I would write all of the sections that responded to the critics so that he could spend his time working on the material presented in chapter 11, which presents the extension of the theory to a mental-predicate logic. Chapter 11 thus is authored only by Marty and the chapters primarily concerned with debate were authored primarily by me. This does not reflect a lack of participation on my part in developing the ideas in chapter 11, nor a lack of participation on Marty’s part in developing the approaches taken in the other chapters—we consulted one another constantly as we worked on our respective parts of the work. It does reflect, however, the division of work on which we had agreed.

At the time of Marty’s death, chapters 4 and 6 were almost completed and required only minor additional work. Chapter 11—the new mental-predicate-logic theory—was largely complete, although parts of the manuscript consisted of handwritten notes, some of which were quite difficult to read. I have included these notes in the published version without alteration, although an occasional word could be worked out only from its context. I am indebted to Lila Braine for her assistance in deciphering Marty’s handwriting in these few instances, although any errors that may have been made are completely my responsibility. In addition, some small amount of material that Marty and I had been discussing that was not in the manuscript has been added to chapter 11. Chapter 12, which reports an initial empirical test of the mental-predicate-logic theory, was completed by Yingrui Yang and myself, although the data collection had been completed when Marty was alive.

I am grateful to several people for their help in completing this volume. In addition to Lila Braine, I would like to thank Ira Noveck for help in translating computer files, Luca Bonatti and Gennaro Chierchia for their helpful comments about the material in chapter 11, Brooke Lea and Patty Brooks in preparing the revisions of articles that had been published previously, and Doris Aaronson, Mark Balton, Murray Glanzer, and Gay Snodgrass for a variety of helpful comments on the work in chapter 12. Finally, I would like to thank Marty Braine for his friendship and inspiration.
...at least to me, the use of technically advanced machinery in analyzing reasoning is encouraging; after all, Aristotle thought about reasoning; one would like to see clearly what one has that he did not have! (It is no comfort to know that over 2000 years have passed since his time unless one sees just how one has used the experience of these 2000 years.)

—Kreisel (1967, p. 271)

“Enough with mental logic! It has been around for three millennia, and no theory came out of it. It’s a dead research program!” You certainly must have happened to overhear speech streams very much like this. (They generally continue with the utterer praising the virtues of sexier alternatives, such as models, interconnected networks, space phases, or what have you.) In case you wanted to know what to respond, I can help you. The right response is, “Yes and no.”

However, “yes and no” is a contradiction and you will not be happy with it, because you, like me and everybody else, possess a natural logic and natural logic abhors contradictions. The real right answer is, “In one sense yes, and in another sense, no.” The “In one sense, yes” part is easy: It is a fact that the idea that we have a logic in our mind goes back 3,000 years or so. The “in another sense, no” part is a bit more complicated. I
tell a long story about how certain ideas have to be blended with the right ingredients, how the ingredients have to be carefully mixed together, and how the mixture must be stored in a warm place far from drafts in order for the yeast to grow. This paper is a memo to recall the basic steps for the recipe in order to be successful.

**BLEND LOGIC WITH PSYCHOLOGY, BUT NOT TOO MUCH: WHEN LOGIC AND PSYCHOLOGY WERE ONE**

Yes, something like the idea that there is a logic in our mind has been floating around for millennia. But could it actually develop into a real psychological thesis? Have the cake ingredients been available all along? Consider a step as trivial as this: If you want to claim that there is a relation between logic and our psychology, first, you had better be clear about the difference between logic and psychology, and, second, you had better get it right.

Yet this has not been a step easy to achieve, because the mental logic thesis was almost always meant in a very strong sense—as the thesis that logic is the science of the laws of thought. As a consequence, logic was considered part of psychology. This inclusive thesis remained surprisingly stable across the centuries, although it was clearly stunting its development by cutting its ties with the rest of mathematics.¹

To a certain extent, however, this stability should not be a surprise. As Kant noticed (1781/1966, pp. 17–18), logic also remained substantially unchanged for centuries, as if all possible forms of reasoning had been laid down once and for all by Aristotle and his medieval commentators. So, for a long time there was no reason to change the doctrine of the psychological nature of logic, because logic was stable because considered complete and psychology was stable because non-existent.

What begins to be surprising is that this thesis resisted even when the horizons of logic broadened, for the first time with the work of Boole. Boole first distinguished propositional and predicative logics, and first noticed a

¹ The fact that logic developed so late if compared with the rest of mathematics is indeed a puzzle. The puzzle is even deeper if one think that little or no mathematics was necessary to develop it, and that the necessary mathematics existed centuries before logic started its own solid development. I am suggesting that part of the explanation has to be found in the fact that the pie was cut in the wrong way. Logic was just not considered a discipline of the same nature as mathematics. The strong inclusive version of the mental-logic hypothesis bears part of responsibility for the anomalous development of logic.

² So also in Mangione (1993): “What does the revolutionary nature of Boolean logic consist in?… Boole clearly and definitely states the formal nature of the calculus in general, in the sense that also the formulation of a logical calculus is a formal construction to which interpretation is added from outside and does not constitute any longer the exclusive and primary basis from which the formal structure is abstracted” (pp. 106–107).
central element for all future developments, namely, the formal nature of logic. Yet, even if he dared to unrest its millennial stability, Boole did not abandon the idea that an investigation of logic is at the same time an investigation of “the fundamental laws of those operation of the mind by which reasoning is performed” (1854, p. 1). Moreover, Boole also made free use of another traditional ingredient that gives scarce results when added to the mental-logic cake. He considered logic as a discipline \textit{sui generis} whose laws are not to be discovered by empirical inspection, like all other sciences require, but can be known entirely by introspection. This conception may be perfectly adequate for logic—after all, this is what intuitionism claims—but obviously cannot work for psychology. Yet, because for Boole logical laws are laws of thinking, a consequence of the mixture of the two traditional ingredients was that for him the “science of the mind”—what we would today call the \textit{psychology of reasoning}—need not be grounded in experience:

The general laws of nature are not, for the most part, immediate objects of perception….On the other hand, the knowledge of the laws of the mind does not require as its basis any extensive collection of observations. The general truth is seen in the particular instance, and it is not confirmed by the repetition of instances. (1854, p. 4)

Let me stress that for centuries this view, according to which in order to achieve “knowledge of the mind” you need only to introspect your own mind, was never seriously questioned. Yet, it is as plausible as the view that you can discover the laws of your natural language parser by introspection; psycholinguistics would not go that far with the latter, nor the psychology of reasoning with the former.

Even if his view of mental logic was so much indebted to the past, Boole’s (1854) analysis did break with tradition in an important point: He introduced the idea that the “laws of thought” have a mathematical character, more precisely algebraical. And this agreement between algebra and logic of thought, he conceded, must be found a posteriori:

There is not only a close analogy between the operations of the mind in general reasoning and its operations in the particular science of algebra, but there is to a considerable extent an exact agreement….Of course the laws must in both cases be determined independently; any formal agreement between them can only be established a posteriori by actual comparison. (p. 6)

Boole also speculated that because the \textit{content} of number theory and logic are different, the “agreement” between operations of the mind and algebra must concern only the \textit{processes} involved. Mental processes and algebraic processes have something in common. What, though, and why?
He did not give a full answer. Perhaps an answer cannot be given, he suggested, because it lays beyond our intellectual capabilities:

Whence it is that the ultimate laws of logic are mathematical in their form, … are questions upon which it might not be very remote from presumption to endeavor to pronounce a positive judgment. Probably they lie beyond the reach of our limited faculties. (1854, p. 11)

This profession of necessary ignorance was too pessimistic. There was a possible explanation of what unifies mathematics and thinking. Boole himself had remarked insightfully that reasoning is a process of compositions of symbols according to laws, that symbols are arbitrary signs, and that their rules of composition are common to both logic and the science of numbers. In order to see the full answer, however, he needed some other ingredients. The missing parts of the recipe were a foundational theory of mathematics that would stress the centrality of signs and combinations of signs, and a psychologically plausible theory that would assign to signs a key role in our thought processes. Both were to be provided in our century, but only, paradoxically, when philosophers and logicians had already abandoned the program of mental logic.

**BUT NOT TOO LITTLE EITHER: LOGICIANS ABANDON PSYCHOLOGY…**

The first dissonant tune was sung by Frege and Russell. With their works, for the first time logic became theoretically separated from psychology. This is indeed good news, if I am right in thinking that confusion about the boundaries between the two disciplines was slowing the development of both. The trouble is, they got separated too much. Logic divorced psychology only to find itself married with metaphysics, and, in order to be accepted in the new family logicians, gave up any ties with the investigation of thinking processes.

Frege (1979) argued forcefully not only that logic—the science of thought, in the special sense he was conceiving it—is separated from psychology—for him, the science of thinking—but also that the latter is the mortal enemy of the former. In his colorful terms:

>[I]t is the business of the logician to conduct an unceasing struggle against psychology and those parts of language and grammar which fail to give untrammeled expression to what is logical. (p. 6)

---

3 Macnamara (1986, pp. 12–20), to which the reader is also referred, makes much the same point I am arguing in this section.
Since thoughts are not mental in nature, it follows that every psychological treatment of logic can only do harm. It is rather the task of this science to purify logic of all that is alien and hence of all that is psychological. (pp. 148–149)

The reasons motivating his calls for purification are interesting for us because they show that Frege was moved by his deep anti-idealism⁴ rather than by an opposition to the mental logic hypothesis qua psychological hypothesis. If we fail to separate logical and psychological content, Frege argued, we obliterate the difference between the strictly subjective and noncommunicable ideas and the universal logical principles, so nothing would guarantee that two persons can entertain the same content, asserting it or denying it. As a consequence, a contradiction between two persons would be impossible, and with it communication as well as science. There would be no way to explain the notion of truth, idealism could not be refuted, solipsism would be inescapable, everyone would forever be condemned to live in his or her own private world.⁵

Even for Russell (1919)—and even more directly—logic became an instrument for the discovery of the correct ontology, a position pushed to its extreme consequences by his claim that logic should not include unicorns any more than zoology does (see 1919, pp. 169–170). As Di Francesco (1991) commented, “for Russell ‘Logic’ does not mean anything linguistic or formal in nature, in our sense. Quite the contrary, what he was looking for was not a set of syntactic rules, but rather the correct description of the objective relations between entia or propositions, considered as complex groups of entia” (p. 87).

Russell and Frege were not alone. For different reasons the logical neopositivists also accepted the new dogma. They saw in the novel developments of logic an occasion to explain at the same time the necessity of mathematical truth—a real mystery for traditional empiricism—and the construction of complex experiences: Mathematics is necessary because it is as empty as logic, and experience is structured in public objects because objects are a logical composition of sense data. Notice that in both roles logic takes over functions that have no connection whatsoever with psychology or the mind. This was somewhat a forced choice for the neopositivists. They were trying to play cards on two separate tables. They wanted both to hold that the origin of meaning is entirely grounded in experience, and to free experience from any connotation of privacy. As a

⁴ “Idealism” here does not include transcendental idealism, but the mixture of psychologism and post-Hegelianism that was current in Germany and England at the end of the century.

⁵ See Frege (1979); for example, in “Logic,” and “The Thought.” See also Macnamara (1986), pp. 14–16.
result, to their ears any talk of inner rules of thought or of mental entities appeared suspect, and behaviorist or reductionist conceptions of the mind sounded more appealing. In America, where the neopositivists migrated to escape Nazi fascism, the climate was perfect for these ideas to flourish. So even Quine, the philosopher who killed the neopositivist program with his criticisms of the analytic-synthetic distinction, still maintained their radical skepticism about the mind and their wary attitude towards any use of logic for psychological purposes.

In short, among the large majority of philosophically minded logicians, showing interest in psychological processes became a sort of behavior that well-mannered people should avoid. Logic and reasoning took different routes precisely when logic was undergoing an impressive development, and precisely under the influence of the pivotal figures of this development.

However, this turn was motivated less by any substantial arguments, or by any problems with the psychological feasibility of mental logic, than by a change in the general cultural climate. This is clear especially in Frege and Russell, whose statements about the separation between logic and psychology can be seen as an epiphenomenon of the general rebellion against German and English idealism from which 20th-century analytic philosophy stemmed, but it is worth recalling that in the first decades of the century many other philosophers of different orientations, such as Moore, Meinong, and Husserl, attacked idealism in one way or another. So the antipsychological shift really was radical. However, it is important to realize that even conceding to Frege and the anti-idealists that the logic of thought is not the logic of thinking, there still is the possibility of having both: a metaphysical theory of thought and a theory of logical thinking, in which some of the logical rules discovered in the investigation of thought have a causal role in our thinking processes. For some reason, Frege and the post-Fregeans never conceived of this intermediate possibility.

So, although of extreme historical importance, the rejection of the mental-logic hypothesis by the new logicians was not really motivated. In the past, logicians had made the mistake of including logic within psychology and had made the further mistake of assuming the mentallogic point of view without giving any substantial argument for it. The new logicians corrected the first mistake by sharply separating logic from psychology. However, in one sense they made the same basic second mistake already made by their predecessors:

---

7 Macnamara (1986) makes this point especially with respect to Husserl’s view of logic, which drastically changed under the influence of Frege. But Husserl was no special case: A plethora of authors was coming to the same conclusions in the same years.
They took a stand towards the mental-logic program—a negative one in this case—without advancing any real argument.

...AND PSYCHOLOGISTS ABANDON LOGIC

Nevertheless, for independent reasons the same ideas were gaining the favor of experimental psychologists and became the standard view at least until the early 1960s. Many factors would probably need to be analyzed to explain the success of this attitude, but a major one is to be found in an unusual alliance of interests between otherwise opposite approaches. Notwithstanding the gulf separating behaviorism and cognitive psychology, at least for a certain period both schools held the same critical position about mental logic. Behaviorists were led to it by their general skepticism about talk of inner mental processes, including inner mechanisms of reasoning. In the limited interest shown for the problem, behaviorism seemed to have its own way to explain the “logical” appearance of people’s responses. Problem solution was considered a learned drive, a kind of associative behavior. To the objection that thinking did not really look like a type of behavior, the answer was that it is in effect an evolutionary sophistication of action, a sort of delayed action that does leave silent behavioral traces in the form of implicit responses whose behavioral realization is only more difficult to be detected, such as subvocal talk or other muscular movements that repeat the full normal behavioral response at a smaller scale.

The underlying idea was that a “central controller” (of which mental logic is a partial description) is not needed for mental tasks: Ultimately, those can be traced back to peripheral muscular activities. So Watson (1924) wrote:

The behaviorist advances the view that what the psychologists have hitherto called thought is in short nothing but talking to ourselves....My theory does hold that the muscular habits learned in overt speech are responsible for implicit or internal speech (thought). (pp. 14–15, italics added)§

Skinner later rejected the subvocal theory of thought, but not the idea that thinking and reasoning can be reduced to behavior. If it was impossible to explain it away by means of a single correlated behavior, for him thinking could be seen as an emergent property of complex behaviors. The basic point

§ See also Humphrey (1963), pp. 185–216.
remained the same: Thinking is not a succession of internal states, but a far consequence of environmental variables on the organism.\(^9\)

The need to dispose of mental vocabulary—and with it, by instantiation, mental logic—looks like a necessary corner into which the behaviorists had to paint themselves. Well, one can say, too bad for the behaviorists. However, paradoxically, even the newborn cognitive psychology, which had no need to put itself in such an uncomfortable position, assumed roughly the same attitude toward mental logic. Although the “New Look” psychology rehabilitated a vocabulary of inner mental processes and flows of information, initially it focused research on the plasticity of mental processes, on the interaction and the mutual influence of reasoning and perception, and on the strong effect of context on both. To this general picture of the functioning of mind the “rigidity” of logical rules was unappealing, whereas a more pragmatic view of reasoning was better suited. Thus we find what is considered the first bible of cognitive science stating that “much of human reasoning is supported by a kind of thematic process rather than by an abstract logic. The principal feature of this thematic process is its pragmatic rather than its logical structure” (Bruner, Goodnow, & Austin, 1956, p. 104). Once again, the intuition was correct but not its implicit conclusion. That pragmatics influences reasoning (as well as other mental processes) is obviously true. However, it is one thing is to say that a theory of reasoning has to take into account pragmatic factors, and another to say that because reasoning is influenced by pragmatic factors, it has no formal components and no mental logic rules. Compare the following: It is one thing is to say that language interactions among humans are heavily influenced by semantics, and another to say that because semantics has a great influence in language, there is no syntax. In both cases the former thesis does not imply the latter. New Look psychologists missed this difference and thus, even if they had no a priori reasons to reject mental logic, they treated it almost as badly as their behaviorist adversaries did.

ADD A GOOD MEASURE OF FORMALISM: HILBERT AND THINKING AS A SYMBOLIC PROCESS

Ideologically, it was again an impulse coming from logicians—not from psychologists—that put logic back in the psychological ballpark. In the rich intellectual debate generated by the problem of the foundations of mathematics, the formalist school has a special role for our story.

Let us distinguish Hilbert’s specific program for the foundation of mathematics from the general formalist approach. The former required that a consistency proof for arithmetic be obtained with finitary means, and Gödel’s second theorem killed it by showing that this cannot be done. The latter more general framework, however, according to which logic was conceived as the science of signs and their transformations, is not necessarily tied to Hilbert’s foundational program. Although this conception falls short of explaining mathematics, it might well be sufficient to explain thinking. It is interesting, and often overlooked, that Hilbert had already realized that his conception would open important new perspectives not only for mathematics, but also (and especially) for the relation between logic and reasoning. Thus when rejecting Brouwer’s scorn at this “material” aspect of logic, so cherished by formalists, Hilbert (1927) wrote:

The formula game that Brouwer so deprecates has, besides its mathematical value, an important general philosophical significance. For this formula game is carried out according to certain definite rules, in which the technique of our thinking [his italics] is expressed. These rules form a closed system that can be discovered and definitively stated. The fundamental idea of my proof theory is none other than to describe the activity of our understanding to make a protocol of the rules according to which our thinking actually proceeds [italics added]. Thinking, it so happens, parallels speaking and writing: we form statements and place them one behind another. (p. 475)

To this idea of proof theory as a description of the activity of the understanding, Hilbert added a particular interpretation of Kant’s thesis that something must be presented in intuition in order for thought to produce knowledge. In the case of mathematics, the extralogical intuitive necessary element is the symbol itself, a concrete object which, just by virtue of its shape, can trigger elementary processes of recognition, matching, concatenation, and deletion and allow a rule to be applied:

Kant already taught...that mathematics has at its disposal a content secured independently of all logic and hence can never be provided with a foundation by means of logic alone;...Rather, as a condition for the use of logical inferences and the performance of logical operations, something must already be given to our faculty of representation, certain extralogical concrete objects that are intuitively present as immediate experience prior to all thought. If logical inference is to be reliable, it must be possible to survey these objects completely in all their parts, and the fact that they occur, that they differ from one another, and that they follow each other, or are concatenated, is immediately given intuitively, together with the objects, as something that neither can be reduced to anything else nor requires reduction. This is the basic philosophical position that I consider requisite for mathematics and, in general, for all scientific thinking, understanding, and communication. And in mathematics, in particular, what
we consider is the concrete signs themselves, whose shape, according to the conception we have adopted, is immediately clear and recognizable. (Hilbert, 1925/1967, p. 376)

Liberated from its interpretive parts, the passage contains the clear suggestion that logical rules of thought can be seen as procedures activated only in virtue of the form of the signs with which they are formulated. In our terms, we may say that a thinking process can be seen as a set of transformations on symbols according to rules exploiting a high-level concrete property of the symbols—their shape.

Hilbert’s two great intuitions—that a thought process is a kind of proof, and that a proof is a transformation of symbols according to rules acting on their form—gave formalism a completely different turn. The syntactic nature of logical rules, together with the statement that our thinking is nothing but the application of rules in a certain order, was opening again the way for undertaking an investigation of the logic of thinking, but this time with a clearer understanding of the nature of the rules and of how properties of their forms could be used to generate deductions.

Thus formalism provided two essential ingredients for the mental-logic cake. Others, however, were still missing. What logic needed to be ready for psychological investigation was, on the one hand, a more intuitive presentation of formal systems and, on the other hand, a model of how a physical structure could use a formal system to carry out derivations. Gentzen provided the first ingredient, and the second one emerges from Turing’s work.

KNEAD FORMS UNTIL THEY FEEL NATURAL TO THE TOUCH: GENTZEN AND NATURAL LOGIC

Hilbert’s axiomatization of logic was meant to capture the set of theorems by using few initial sentences and the smallest possible number of inference rules. This system, however useful for handling many domains of logic and mathematics, fails to be appealing even to represent the particular kind of reasoning it aimed to represent, namely, logical and mathematical reasoning. Far less could it be proposed as a model of how human reasoning generally proceeds. Gentzen’s natural deduction took care of this problem. An alternative treatment was developed for some parts of logic, in which axioms disappeared, leaving many inference rules in their places. Gentzen (1964) consciously intended to provide a system of deduction “as close as possible to actual reasoning”:

The formalization of logical deduction, especially as it has been developed by Frege, Russell, and Hilbert, is rather far removed from the forms of deduction used in practice in mathematical proofs. Considerable formal advantages are
achieved in return. In contrast, I intended first to set up a formal system which comes as close as possible to actual reasoning. The result was a “calculus of natural deduction.” (p. 68)

He noticed that in actual mathematical reasoning, appeal to formal axioms is rare, whereas many—but not too many—forms of inferences are used. He set himself the task of specifying those inferences. To this purpose, he introduced different kinds of calculi in which each logical symbol occurred in two rules specifying when a formula containing it could be introduced and when it could be eliminated in favor of other formulas not containing that symbol. This central difference from axiomatic systems makes proving a theorem a completely different, more natural activity than proving it starting from axioms. Yet, nothing is lost in natural deduction: The crucial metatheoretical properties of axiomatic systems are all preserved.

Another difference with Hilbert-style logics is worth noticing. Ideally, an axiomatic system aims at maximal economy, both in axioms as well as in rules. Gentzen, instead, provided rules for all the logical connectives, even if they were logically redundant. There are technical reasons for it, but there is also an important philosophical motivation. Gentzen intended to provide a system as close as possible to actual mathematical reasoning, and from this point of view economy in rules could not be a value in itself. If mathematical natural reasoning seems to require even redundant rules, so be it. A related point has to do with the interdefinability of connectives. In classical logic equivalence laws, such as, for example, \((A \rightarrow B) \leftrightarrow (\neg A \vee B)\), render connectives interdefinable and thereby dispensable in principle. However, this goes flat in the face of another natural intuition, the feeling that each connective has a separate meaning. Natural deduction also provided a way to account also for this intuition. Even if one adopts classical rules for connectives, and thus the equivalence laws are still theorems, from Gentzen’s point of view such extensional equivalences do not show that one can dispense with the rule. If we use them when following a path of thought, if they appear to govern the functioning of certain special words lexicalized even in normal natural language, such as and, or, if, or all, then they are likely to correspond to separate units of meaning and they deserve a place in a system of mathematical reasoning.

10 With the exception of classical negation. In the intuitionistic case, negation is governed by the elimination rule \(\frac{A}{\bot}\) and by the introduction rule \(\frac{A}{\neg A}\) whereas to get classical logic a special extra elimination rule expressing double-negation redundancy is needed: \(\frac{\neg\neg A}{A}\) This rule for Gentzen spurs the “harmony” of a system in which to each connective just one elimination and one introduction rule correspond.
In short, Gentzen’s natural deduction achieved for mathematics what Gentzen wanted to achieve: It showed how a proof can be at the same time formal and natural. The route from formal logic to natural reasoning was opened. Gentzen did not provide a psychological system of rules for reasoning—he could not, because the psychology to go any further was not yet available—but because _mutatis mutandis_ his arguments hold for reasoning in general, he contributed another basic ingredient to its development. He shaped the form of the rules and showed how formal logic (this time, a real discipline separated from psychology) and intuitive logic may be not so distant from each other, after all.

**WHIP COMPUTATIONS UNTIL THEY ACQUIRE THE RIGHT CONSISTENCY: THE COMPUTER METAPHOR AND FUNCTIONALISM**

The other necessary step for transforming mental logic into a psychological thesis was to develop fully Hilbert’s intuitions about how rules exploit the “concrete” aspects of signs. Notoriously, it was Turing who offered the abstract model of a shape-driven machine, which showed how a physical mechanism could perform operations once considered to be typically mental. Turing’s intuitions on the role and possibilities of his new machines were of exceptional importance. They finally led to the thesis that the mind can be seen as an information processor, which in turn gave substance to the idea of a mental logic. However, in the development of this view two stages should be distinguished. The first stage, explicitly endorsed by Turing, is expressed by the thesis that _a computer can be a mind_, namely, that certain kinds of properties once attributable only to humans can also be appropriately predicated of other physical configurations. This thesis, however revolutionary, leaves completely undetermined the nature of the operations of the mind. It may imply that mental processes can be _simulated_ by a machine but leaves it open the possibility that the mechanisms and procedures intervening in the two physical beings—the _simulandum_ and the _simulans_—are totally different. All that the thesis requires is that the input/output relations typical of mental processes are preserved by the simulation device, and simulation only establishes an extensional identity between two physical configurations, but leaves it undetermined whether these share the same _psychology_.

Turing’s well-known test for intelligence, entirely based on the control of input/output relations (at least prima facie), uses only dispositional language and is thus compatible with conceptions that do without mental vocabulary, such as behaviorism or logical behaviorism. It does not require laws mentioning internal states, and if it does, it doesn’t require that they apply
to both minds and computers. To see this, imagine that a computer passes Turing’s test; it may still be the case that when we reason we exploit lawfully connected inner states, but the computer does not. Suppose, for example, that we parse the computer’s responses to our questions by using a grammar for a language. Then, a part of our psychology would be described by linguistic theory. But when the computer “reads” our questions, it consults a huge database of predefined forms and outputs answers to each question that we plausibly may ask by mere physical shape-matching, without really parsing our input. In this case, although the computer may happen to pass Turing’s test, we and it would not fall under the same psychological laws: Linguistics would be true for us, and false for the computer.

Turing did intend to go beyond this minimal claim; witness the fact that he proposed tests for machine intelligence couched in more “cognitive” vocabulary. However, his main concern was not to dispel the prejudice that we cannot be like machines, but rather to reject the reverse thesis that machines cannot be like us. For him, the question of whether the inner structure of mental processes and the computational states of machines are captured by a common set of laws was not at the forefront. He did use analogies to mental processes for finding algorithms, but this was meant to be a heuristic strategy based on introspection rather than a research program. His caution also is shown by his willingness to explore all possible sorts of machines, whether “classical,” as we would call them, or “connectionist,” with the main task of finding agreement in behavior between minds and machines, rather than agreement in laws.

Turing’s ideas were of exceptional value, but for various reasons—mostly due to the widespread behaviorist attitude in psychology and philosophy—further time had to pass before they could be elaborated into a fully coherent program. As Hegel wanted it, philosophy always comes at dawn. The new paradigm had to wait for the development of the theory of computable functions and for the first successes of artificial intelligence before being taken seriously by philosophers. The big further step—the reverse thesis that we are like computers—came with functionalism. Functionalism explicitly defended

---

11 The computer could not answer all questions we may possibly ask, because it wouldn’t include a productive language, but lack of productivity would not forbid it to pass Turing’s test. The test only requires that we could not plausibly tell a computer from a human, not that we could not possibly do it.

12 Specifically, he proposed that a machine that learns (i.e., that modifies its instruction) should be considered intelligent; see Turing (1947b, pp. 122–123).

13 For example, when outlining a program for playing chess, he wrote, “If I were to sum up the weakness of the above system in a few words, I would describe it as a caricature of my own play. It was in fact based on an introspective analysis of my thought processes when playing, with considerable simplifications. It makes oversights which are very similar to those which I make myself” (Turing, 1947b, p. 294).
the thesis that the psychological vocabulary is computational vocabulary, and that the natural kinds described by psychology are not organisms but computational devices. Fodor (1981) presented it as a fairly recent contender in the arena of the theories of mind:

The real point is that, if we want a science of mental phenomena at all, we are required to so identify mental properties that the kinds they subsume are natural from the point of view of psychological theory construction.... Now, there is a level of abstraction at which the generalizations of psychology are most naturally pitched and, as things appear to be turning out, that level of abstraction collapses across the differences between physically quite different kinds of systems....[I]f we wanted to restrict the domains of our psychological theories to just us, we would have to do so by ad hoc conditions upon their generalizations. Whereas, what does seem to provide a natural domain for psychological theorizing, at least in cognitive psychology, is something like the set of (real and possible) information processing systems. (pp. 8–9, italics added)

We are now well beyond the 1960s. Only then were logic and philosophy ready to nurture the development of mental logic as a serious psychological hypothesis. Everything was there, except the psychology.

MIX WITH EMPIRICAL PSYCHOLOGY: ALMOST GETTING THERE

Ideas must be ripe before becoming productive. I argued that although the idea of a mental logic has been around for centuries, it could not develop because the ingredients necessary to give substance to it either were not available or were seen from a wrong perspective. When, finally, all the conceptual ingredients did become available, the psychology for testing mental-logic theories was still missing. Another 20 years or so had to pass before experimental techniques were sufficiently developed to begin asking nature the right questions in the right way. It is true that Piaget and his school (e.g., Piaget, 1953) spoke a language externally akin to mental logic, but in their hands the thesis was never even spelled out clearly enough to see what it said or how it differed from other alternatives (see Bonatti, 1994; Braine & Rumain, 1983). If one eliminates Piagetianism, which did more harm than good to the mental-logic hypothesis, only a handful of articles and proposals remain that can be considered real attempts to transform it into a testable psychological theory, and until very recent years most of them fell short of being satisfactory, for one reason or another.

Henle (1962) was probably the first attempt to test experimentally a mental-logic hypothesis, but the limits of her work were important. She took natural
logic to be classical logic—not even in natural deduction form—but she gave no argument for this assumption. Her method was entirely based on direct analysis of subjects’ explicit conscious justifications, but those may not be a good indicator of the real processes involved in reasoning. She assumed that the syllogism has a special role in reasoning, but this too was an unjustified assumption. In short, the merit of Henle’s study was less in her arguments and methodologies than in the theses she presented.

It was only in the 1970s that the mental-logic hypothesis marked significant improvement towards its transformation into a psychological theory. In those years natural deduction became the standard form of the theory, and purely qualitative analysis was abandoned in favor of more quantitative methods often accompanied by computational implementations. Nevertheless, the proposed theories still lacked a sufficient psychological motivation. Osherson (1975a) and Johnson-Laird (1975) proposed models of rules that shared many common points and diverged for computational options. Whereas Osherson’s model was conclusion sensitive and applied its rules in strict order trying to minimize the difference between premises and conclusions, Johnson-Laird’s (1975) program blindly generated consequences of the premises, regardless of the form of the conclusion. Both systems included good and new ideas, but, interestingly for us, both were clearly more concerned with the algorithmic implementation of rules than with their psychological plausibility. For example, in both systems the basic rules were selected somewhat arbitrarily. Both systems took a one-way approach to the role of a conclusion in a problem, one being conclusion sensitive and the other conclusion blind, but neither provided evidence that subjects exclusively or preferentially follow one of those strategies. Likewise, although Johnson-Laird’s system introduced the distinction between primary and secondary inferences that remains central also in the system presented in this book, his way to sort the rules in the two classes was not psychologically motivated. It is revealing that at the end of his 1975 article, Johnson-Laird admitted to not being sure of how to test his model empirically, and he concluded that this was a really difficult, almost impossible, task. Clearly, psychological reality was not at the forefront.

In fact, both Osherson and Johnson-Laird could hardly have done any better. They had run into another tricky problem. They lacked another crucial ingredient for a mental logic, a really difficult one to find and to calibrate: the right balance between algorithmic implementation and psychological justification for a mental logic system. Only recently (e.g., Braine, 1978; Rips, 1983) did some researchers start tinkering with it, to find out the right proportions between the two ingredients. This work is still at its beginnings, and much is still needed to nail down the rules for reasoning, their implementations, and the flow of information involved in a reasoning process.
HAVE A TASTE OF THE CAKE: GETTING THERE

So developing a mental-logic theory was not so easy after all. Many ingredients needed to be found and their proportions carefully weighed before a good mixture could be formed. Logic had to be separated from psychology, but this only happened after Boole, with the development of modern logic. Its formal nature had to be put in focus, and the logical properties of forms had to be discovered, but this only happened when logicians started working on the foundations of mathematics, at the beginning of this century and well into the 1930s. The connection between forms and machines acting on forms had to be seen, but this was clear only with Turing and his notion of computability. The philosophical consequences of the new view about mental processes implicit in Turing’s ideas had to be drawn, but this only happened some decades later, when philosophical functionalism ripened. Only then could psychology revive the mental-logic hypothesis and look at it with a clearer vision of the claims being made. Psychology too had to make progress. It had to abandon the ancient commonplace that the laws of reasoning can be discovered by immediate introspection. Also, this step was slow to come. It had to devise experimental techniques to investigate the rules, and to test the psychological reality of the proposed algorithms by implementing them in the conditions in which they can be deployed. All of it is very recent development, and much of it has yet to be written.

It should no longer be a surprise that mental logic started developing fairly recently. In fact, it has a development as recent as that of its supposed sexier alternatives, such as models, interconnected networks, space phases, or what have you. The reason is simple: They all require pretty much the same conceptual ingredients to be developed, and they all require a well-developed empirical psychology. There is no dead research program in the area of reasoning; there are only theories in better or worse shape.

“OK, fine. But even if the long history of mental logic does not count against it, its present state does! Even now, with all the ingredients available, the mental-logic dough has not leavened. Compare with X [substitute X for models, interconnected networks, space phases, or what have you]: X has gone so much further!” Again, there is an answer. This time, the answer is “No and No.” No, it is false that there are better alternatives to mental logic, and, No, it is false that mental logic has not begun to be productive. Have a taste of the cake by reading the rest of this book.
Are people rational or are they irrational? Before starting to write this chapter I posed this question to an undergraduate class in cognitive psychology. No shortage of evidence was offered for either alternative. On the one hand, it was pointed out that people do many things that most of us would judge irrational: We fail to wear seat belts while knowing that this decreases our safety, we smoke cigarettes while knowing that this may lead to a frightful disease, we spend the weekend drinking at parties while knowing that an exam is being given on Monday, we continue to use our credit cards while knowing that we are unable to repay the debts already accumulated. On the other hand, people do many things that reflect a rational nature: We plan for the future, anticipating the effects of our actions; we have created mathematics, formal logic, complex engineering and computing systems, high technology, science, and philosophy. Ironically, our apparently most rational accomplishments can lead to the most irrational results; for example, modern physics has provided for the development of weapons that threaten to bring about our extinction. For me, the tenor of the discussion was captured when one student posed the

A popular rhetorical protasis, “If we can put a man on the moon,” and another student provided the apodosis, “why can’t we solve the THOG problem or the selection task?”

The consensus of my class was that as humans we have in our nature to be both rational and irrational, a conclusion with which I expect most readers will agree. Differences among us are apt to concern which aspects of human nature we propose are rational and which ones we propose are irrational. Evans (1993), for example, proposed that people possess a kind of rationality that enables us to make decisions so as to maximize the prospects of benefit to ourselves, but we do not possess a rationality of inherently logical thought processes. I disagree. People may be motivated to seek benefits for themselves and their families, clans, tribes, nations, and so forth, but my undergraduate students provided many examples of the ways in which people are not adept at maximizing such benefits. Furthermore, I argue, people’s thought processes are, in many ways, profoundly logical. I also believe that the empirical evidence per se does not settle the issue: In the practical realm we successfully do some things that appear to be in our self-interest and some that appear to be irrational, and in the realm of laboratory logical-reasoning tasks we do some things that appear to be logical and some that appear to be irrational.

The classical Greek view of human nature included a rationality that allows for logical reasoning. My colleagues and I argued elsewhere (Noveck, Lea, Davidson, & O’Brien, 1991) that we have no adequate reason to abandon this view and that this rationality includes a mental logic that accounts for our basic logical intuitions. In recent years the claim that human reasoning includes a mental logic has met considerable resistance, and the death of mental logic is proclaimed with some regularity (e.g., Cheng & Holyoak, 1985; Cosmides, 1989; Johnson-Laird, 1983; Johnson-Laird & Byrne, 1991; Legrenzi & Legrenzi, 1991). To paraphrase Mark Twain, news of this death is premature; theories of mental logic are alive and well, though often misunderstood. Indeed, I believe that most criticisms of the mental-logic approach stem from a misunderstanding of what it is.

The first section of this chapter describes the mental-logic approach. Theories of mental logic are not monolithic, and what counts as evidence against a particular theory does not necessarily count as evidence against all mental-logic theories. I first discuss the mental-logic approach generally, noting that a mental logic consists of propositional activities, then focus on the three-part theory presented in this volume. The second section addresses some principal criticisms of mental logic. In particular, I argue that (a) failure to solve complex reasoning problems does not count as evidence against mental logic, and (b) we should not interpret evidence of the effects of content on logical reasoning as counting against mental logic. The third section
addresses why I do not find the competing nonlogical theories compelling, and focuses on the content-bound theories proposed by Cosmides (1989) and Cheng and Holyoak (1985) and the mental-models theory proposed by Johnson-Laird and his associates (e.g., Johnson-Laird, 1983; Johnson-Laird & Byrne, 1991). Finally, I address irrationality from the perspective of the mental-logic approach.

THE MENTAL-LOGIC APPROACH

An adequate theory of human logical reasoning needs to account for both logically correct and erroneous judgments. We thus are faced with an apparent dilemma. On the one hand, we can assume that the human reasoning repertory includes a mental logic, in which case we have an explanation for those valid logical judgments that people make, but we still require an explanation for reasoning errors. On the other hand, we can assume that there is no mental logic, which provides an explanation for errors in reasoning but leaves the valid judgments unexplained. The proposal that there is no mental logic, though, is based on a misunderstanding of the nature of logic, and adoption of the first horn of this apparent dilemma—that there is a mental logic—is the rational choice.

Kant (1781/1966) proposed that human understanding is made possible in part by a mental logic, and that in the 2,000 years since Aristotle nothing had been added to this logic and nothing altered, so logic could be considered completed and perfect. As Macnamara (1986) noted, this view of logic reflects a Platonic heritage, and encourages the view of a mental logic as the manipulation of symbolic forms. This logic of forms is the sort of logic, I believe, that is being decried by opponents of the mental-logic approach. Since Frege at the end of the 19th century, however, logic has undergone a revolution, and logicians have come to view logic as propositional and intentional (Kneale & Kneale, 1962).

Propositions take truth values, that is, a proposition is either true or false. Note that sentences per se are not propositions. The sentence “I am traveling with an American passport” is neither true nor false, but is true when asserted by myself outside America and false if asserted by Margaret Thatcher. Were logic concerned with sentences, it would be concerned merely with the manipulation of symbolic forms. However, the assumptions and conclusions of logical arguments are propositions.1

1 Images are not propositional; although an image might be an accurate or an inaccurate representation, it can be neither true nor false. Propositions that refer to images, though, are true or false. Probabilistic propositions, however, take truth values. (The claim that “there is an 80% probability of showers next Tuesday” is either true or false; it is not 80% true and 20% false.)
Logical reasoning consists of propositional activities. Propositions are proposed, supposed, assumed, considered, claimed, believed, disbelieved, doubted, asserted, denied, inferred, and so forth. All of these propositional activities concern judgments about truth and falsity, for example, to believe $x$ is to believe that $x$ is true, to doubt $x$ is to doubt that $x$ is true, to deny $x$ is to claim that $x$ is false. Such propositional activities concern intentional states of affairs. When I assert that “Napoleon was in Egypt,” I refer to an historical person and his relation to a country, not to a symbolic idea in my mind that I can manipulate. This intentionality does not presuppose any claim to realism—it would be no less propositional were Napoleon merely the figment of historical imagination. Note that I also could assert that “Ahab was obsessed with a white whale,” which we also would judge true even though its intentional state of affairs is fictional. The reason that the sentence “I am traveling with an American passport” is neither true nor false until its utterer and circumstances are known is that sentences per se, unlike propositions, do not concern intentional states of affairs.

Because propositional activities refer to intentional states of affairs, they coexist with pragmatic activities, having to do with the practical consequences of propositions for their referred states of affairs (James, 1885/1978; Peirce, 1931/1958). These pragmatic activities include setting goals and understanding goals set by others. The logical processes that infer propositions and make inferences from them cohabit easily with other processes that are pragmatic and rely on knowledge of intentional states of affairs.

Propositions can be atomic or compound, that is, atomic propositions can be negated, or joined in conjunction, disjunction, conditionality, and so forth. For example, “If I am traveling with an American passport, then I must be an American citizen” supposes the proposition described earlier, and joins it with a conclusion drawn from that supposition together with other assumed propositional information. We need an account of how we reason to and from such compound propositions, that is, how we form and use them.

Forming a compound proposition requires an inference—one does not observe a disjunction or a conditional. Such connections are inferred, as both Hume and Kant noted in their different ways, in the understanding. It would be both a cruel hoax and evolutionarily disadvantageous for nature to provide us with propositional representations if our ways of connecting them and reasoning with them failed to preserve their propositional status. Because propositions are profoundly truth functional, their inference procedures ought to be truth preserving, that is, given a set of propositions assumed true, further propositions drawn from them by logical procedures also would be true. Logicians refer to this property as logical soundness, and a set of inference procedures is sound if, and only if, given a set of true propositions, the inference procedures will provide true conclusions only.
Soundness is distinct from validity; an argument is valid unless there is a possible assignment of truth values such that its premises taken conjunctively are true while its conclusion is false. Thus, in standard logic any argument with necessarily false premises is valid. This property is not part of our ordinary logical intuitions, with which we proceed from propositions assumed true. (See chap. 9 for a discussion of deliberately counterfactual suppositions.) Indeed, it is not uncommon for people to reject an argument because they do not accept its premises. Thus, ordinary reasoning proceeds not from premises, but from assumptions, that is, from premises that are assumed true (see Leblanc & Wisdom, 1976; Politzer & Braine, 1991; chaps. 7 and 9). Unlike the standard logic of textbooks, people draw no conclusions from contradictory premises—such premise sets cannot qualify as assumptions. No one ordinarily would assume a contradictory set of premises, but would see such as absurd.

To summarize, logical reasoning is profoundly propositional, and propositional activities, such as asserting, denying, believing, doubting, and so forth, require intentional states of affairs; logical inference procedures cohabit with pragmatic inference procedures that concern the practical consequences of propositions for intentional states of affairs. A mental logic is not, therefore, a matter of mere symbol manipulation but is about making propositional inferences, and ordinary reasoning applies sound inference procedures to propositions assumed true to infer propositions that inherit that truth.

**Inference-Schema Models**

Piaget (e.g., Inhelder & Piaget, 1958, 1964) proposed that the structure of concrete-operational thought corresponds to a logic of classes, and formal-operational thought is equivalent to the 16 truth-functional operators of normal-disjunctive form processed by the mathematical INRC group. This proposal, however, has been criticized on logical grounds (e.g., Braine & Rumain, 1983; Ennis, 1975; O’Brien, 1987; Parsons, 1959), and in recent years a consensus has developed among mental-logic adherents that reasoning proceeds through the application of sound inference schemas (e.g., Braine, 1990; Braine & O’Brien, 1991; Braine, Reiser, & Rumain, 1984; Johnson-Laird, 1975; Macnamara, 1986; O’Brien, 1987, 1991; Osherson, 1975a, 1975b; Rips, 1983; Sperber & Wilson, 1986). Inference schemas are procedures that specify which propositions can be derived from assumed propositions of a particular form, and sound inference schemas assure that propositions derived from true assumptions inherit that truth. Thus far, psychological models proposing inference schemas have been developed only for sentential connectives and have not yet addressed the role of quantifiers (see, however,
People make many sentential inferences that are sanctioned by standard systems of sentential logic, but routinely fail to make others. The primary task of a psychological inference-schema model is to describe those inferences that are made regularly and routinely. A secondary task is to describe those logical inferences that people make only sometimes.

Following Gentzen (1935/1964) logicians described two sorts of inference schemas: those used to introduce a propositional connective in a line of reasoning, for example, a schema for conditional proof to introduce propositions of the form *if* \( p \) *then* \( q \), and those used to eliminate a propositional connective in a line of reasoning, for example, a schema for disjunction elimination (\( p \ or \ q \), \( \neg p \); therefore \( q \)), modus ponens (\( if \ p \ then \ q \), \( p \); therefore \( q \)). Note that inference schemas address the forms of propositions, but this syntactic nature of the inference procedure does not diminish the propositional nature of either the atomic or compound propositions that are inferred. Sound inference procedures ensure that the drawn inferences will be truth preserving, that is, that only true propositions will be drawn from true assumptions.

The several varied inference-schema models that have been proposed do not make identical claims about the role of the schemas. For example, Macnamara (1986) proposed that inference schemas are used as logical checking devices, checking the soundness of inferences made by other, nonlogical, devices. Sperber and Wilson (1986) proposed that only elimination schemas are used. The models proposed by Rips (1983), Osherson (1975a, 1975b), Johnson-Laird (1975), and in this volume all use both introduction and elimination schemas to make inferences in lines of reasoning. Differences exist among these theories, however, concerning the particular schemas that are proposed. For example, the model described in chapter 6 does not include the disjunction-introduction schema, which draws propositions of the form \( p \ or \ q \) from \( p \), although the models of Rips (1983) and Johnson-Laird (1975) do include such a schema.

The approach that I advocate proposes a three-part model (hereafter referred to as *The Model*), which includes (a) a set of inference schemas, (b) a reasoning program that implements the schemas in a line of reasoning, and (c) a set of independently motivated pragmatic principles that influence interpretation of surface-structure propositions and can suggest or inhibit certain inferences and reasoning strategies. The description of The Model in the next few pages is intended to describe the present state of work—both what has been accomplished and what has yet to be done.

**The Inference Schemas and the Reasoning Program**

The inference schemas of The Model have been presented in detail elsewhere (chap. 6) so I do not do so here. The Model includes both introduction and
elimination schemas, with a set of core schemas and some feeder schemas, both of which are implemented through a direct-reasoning routine, and a set of complex schemas that require coordination through an indirect-reasoning routine. The core schemas describe a set of inferences that people make routinely and without apparent effort.

The Model holds that the core schemas are applied automatically through a direct-reasoning routine whenever the appropriate propositions are considered together, for example, when both \( p \) or \( q \) and \( \neg p \) are jointly considered, \( q \) will be inferred automatically. The feeder schemas, however, are not applied unless their propositional output feeds into a subsequent inference (see also Johnson-Laird, 1975, on auxiliary schemas), in which case they are applied automatically by the direct-reasoning routine. Lea et al. (1990) presented two reasons for this differentiation of the feeder schemas from the core schemas. One is theoretical: The feeder schemas can lead to infinite loops, for example, from \( p \) and \( q \) to \( p \), to \( p \) and \( (p \) and \( q) \), and so forth. People exhibit no tendency to make such inference strings. The second reason to designate these as feeder schemas is empirical; when subjects are asked to write down every inference they can from a set of assumptions, subjects usually omit the output of the feeder schemas while writing down the output of the core schemas, even when the output of feeder schemas is required to make the core-schema inferences.

The complex schemas require use of an indirect-reasoning routine. For example, in order to falsify a proposition \( p \), one can suppose \( p \) and seek to find a contradiction under this supposition. When this indirect-reasoning strategy does lead to a contradiction, a negation-introduction schema allows assertion of \( \neg p \). Unlike the core and feeder schemas, which are applied effortlessly and routinely, the complex schemas are in nowise claimed to be universally available, and their application depends on the effortful use of an indirect-reasoning routine.

Acquisition of the indirect-reasoning routine requires some tuition or reflection, and its use may be either encouraged or discouraged by knowledge of the referred intentional state of affairs. The direct-reasoning routine, however, is considered basic to logical reasoning. The Model thus predicts that problems requiring sophisticated reasoning strategies will not be solved readily by most people, whereas problems that can be solved through the direct-reasoning routine will be solved most of the time.

Differences exist among the inference-schema models concerning the nature of the reasoning program. Rips (1983, 1994) proposed an approach in which the line of reasoning proceeds towards a goal, either a conclusion to be evaluated or a lemma required to evaluate a conclusion. Hence, if subjects are presented a set of assumptions with no conclusion to judge, the model generally would draw no inferences. However, several investigations have
reported a wide variety of problems without any conclusions to be evaluated on which subjects have demonstrated no difficulty in drawing logically appropriate inferences (e.g., Lea et al., 1990; O’Brien & Lee, 1992).

As described earlier, The Model proposes that both the core and feeder inference schemas are applied automatically both in processing discourse and in reasoning. This claim of automaticity is problematic, though, when one considers that the inferences of the feeder schemas are drawn only when they provide the input for drawing further inferences. This seems to indicate that people look ahead to see what inferences are needed before they make these inferences—hardly what one expects of an automatic process.

An additional, empirical reason to think that people look forward when drawing inferences is found in comparison of problems with and without conclusions to be judged. Consider the following two problems that I gave to some undergraduates recently (about toy animals and fruits in a box). The first problem presents assumptions of the form \( p \lor q, \text{ if } p \text{ then } r, \text{ and if } q \text{ then } r \). This problem presents no conclusion to be evaluated but requires subjects to write down everything that can be inferred. Most subjects write down only \( r \), which follows directly through one of the core inference schemas. On the second problem subjects are given the same set of assumptions, but are asked to evaluate as a conclusion \( \text{if not-}p \text{ then } r \). On this problem subjects usually write down first \( q \), and then \( r \), before judging the conclusion as true—a line of reasoning that follows from The Model’s schema for conditional proof. Clearly, were subjects not looking forward to consider the conclusion, the line of reasoning on the second problem should be the same as on the first problem. At least some of the time, people seem to look forward when they are applying the basic schemas of The Model.

As yet, the reasoning program of The Model does not capture adequately when subjects look ahead and when they do not, but then, neither does the reasoning program proposed by Rips (1983), with its focus on goal-attaining inferences. Relative to the schemas, little empirical work has been done on the reasoning program, but the need for further investigation of how the inference schemas are implemented becomes apparent a fortiori with the realization that lack of sophistication in using the reasoning program is a principal source of reasoning errors. Clearly, future developments in describing a reasoning program must include some forward-looking as well as some automatic processes.

Several possible solutions could be suggested. One possibility is that the feeder schemas are applied automatically but are under processing constraints and subject to a response filter. The processing constraint might limit each feeder schema to a single application for a set of assumptions, and as a function of a filter its output would not be noticed unless it feeds a subsequent inference. A second possibility is that the reasoning program always looks
ahead to seek potential goals—either logical or pragmatic. When a clear goal is discovered, reasoning proceeds towards that goal; when no goal is presented, the core schemas are applied automatically.

As the investigation proceeds, the description of the reasoning routines, as well as of the schemas, will rely on empirical investigation. The basic part of The Model is intended to describe real-time processes, and these should be open to real-time measurement. Brooke Lea, as part of a doctoral dissertation with Martin Braine, investigated the core and feeder schemas, both in logical-reasoning problems and in text comprehension, measuring reaction times to investigate when subjects make these inferences, and I believe that this sort of investigation, among others, is needed to address the matter (see chap. 5).

**Pragmatic Principles**

That propositions are not identical to their surface-structure expressions is well known among memory and text-comprehension researchers (e.g., Bransford & Franks, 1971; Bransford, Barclay, & Franks, 1972). Likewise, the logical connectives in a mental logic are not identical to the natural-language particles used to express them, although the two should be in close correspondence, so that certain words in a natural language would provide regular ways of expressing certain sorts of propositional connectives, such as the English-language words *and* for conjunction, *or* for disjunction, *not* for negation, and *if* for conditionality.

Natural-language logic particles have meanings that allow people to solve problems and draw inferences. Suppose we are given a problem providing assumptions of the form *if p or q then r* and *p*; we would conclude *r*. Given instead *if p and q then r* and *p*, we would conclude that nothing follows. The two problems differ only in the use of *or* in one problem and *and* in the other, so the difference in responses must be based only on the meanings of these words. Braine and O’Brien (chap. 9) proposed that the basic meaning of a logic particle—its lexical entry—is provided by its basic inference schemas. For example, the basic meaning of *if* is provided by modus ponens and a schema for conditional proof.

The basic meaning of a logic particle, given by the basic schemas, can be extended by invited inferences. Geis and Zwicky (1971) provided an example of an invited inference that derives *if not-p then not-q* from *if p then q*. Another example of an invited inference for *if* is found in the pragmatic-reasoning schemas of Cheng and Holyoak (1985), who claimed that modus-tollens inferences are provided by conditionals that express permissions and obligations.

Invited inferences can be encouraged or discouraged by knowledge of the referred intentional state of affairs. For example, the Geis and Zwicky invited
inference is encouraged by promissory content, for example, “If you mow the lawn, I’ll give you five dollars” invites the listener to infer that “If I don’t mow the lawn, I’ll not receive the 5 dollars.” O’Brien, Costa, and Overton (1986) found that the Geis and Zwicky inference is more likely on problems within a mechanical domain than within a biological domain, where knowledge of spontaneous remissions discourages the invited inference. Staudenmayer (1975) reported that this inference is encouraged when the problems express a causal connection.

Such invited inferences, as supplements to the basic inferences, extend rather than restrict the available inferences. Unlike the basic inferences, invited inferences may or may not be sound. Invited inferences can lead to appropriate responses, as with the pragmatic-reasoning schemas of Cheng and Holyoak (1985), or to logically inappropriate responses, such as those that follow from the invited inference of Geis and Zwicky (1971), and lead to the fallacies of the conditional syllogisms.

An invited inference that is not logically sound, however, is not necessarily irrational. The inference that one will not receive the 5 dollars if the lawn is not mowed is not sanctioned by logic, but is sanctioned by knowledge of the intentional state. Invited inferences are inherently pragmatic because they concern the practical consequences of the considered proposition for its referred intentional state of affairs.

Knowledge of an intentional state of affairs also can suggest alternatives or suppositions to be considered. For example, a mechanic faced with a motor that fails to start might draw on intentional knowledge and infer that the problem is either in the electrical system or the fuel system. If testing the electrical system reveals no problem, the mechanic would conclude that the problem must be with the fuel system. Logical inference schemas and pragmatic sorts of inference-making processes cohabit easily within a single line of reasoning, with the output of one sort of process feeding into the inferences made by the other. Generally, it seems unlikely that an inference is marked for its source—whether the inference stems from a logical inference schema or from intentional knowledge. Although the source of the inference is of interest to a cognitive investigator, there is no reason to think it is important to the mind engaged in the line of reasoning.

Some Evidence for The Model

Direct Evidence. Several studies have provided direct tests of The Model’s claim that the core and feeder schemas together with the direct-reasoning routine are readily available. On one type of problem subjects were provided propositions that refer to letters written on an imaginary blackboard, for example, “On the blackboard there is either a T or an X.” On another type of
problem the propositions refer to boxes containing toy animals and fruits, for example, “In this box there is either a lion or an elephant.” Note that these blackboard and box problems present materials that allow subjects to refer the propositions to an intentional state of affairs, although not to one that would provide the necessary inferences for solution on a basis other than the meanings of the logic particles.

The problems in chapter 7 presented assumptions together with conclusions to be evaluated, and The Model predicted successfully which problems were solved correctly, response times on simple problems, and subjects’ judgments about relative problem difficulty. Lea et al. (1990), Fisch (1991), O’Brien and Lee (1992), O’Brien, Braine, and Yang (1994), and chapter 8 presented problems with conclusions to be evaluated on which subjects were asked to write down every intermediate inference they drew on the way to evaluating the conclusion, and problems that presented assumptions without any conclusions, on which subjects were asked to write down everything they could infer from the assumptions. On both sorts of problems, The Model predicted successfully which inferences subjects wrote down, and the order in which they were written down. Subjects almost always wrote down the output of the core schemas in the order predicted by the model, but they almost never wrote down the output of the feeder schemas, even though the output of the core schemas often depended on the previous output of the feeder schemas. A few subjects, perhaps responding to the instructions to write down everything, wrote down the output of the feeder schemas, and when they did, this output was in the order predicted by The Model.

The inferences that The Model predicts should be made effortlessly and routinely were made routinely and with little apparent effort. These findings are not limited to American undergraduate students. Fisch (1991) found that 9- and 10-year-olds make the basic Model inferences as easily as do adults, and O’Brien and Lee (1992) found the same results when American college students were presented problems in English and Hong Kong college students were presented the same problems in Chinese. Although there may be other inferences also made routinely, those included in The Model appear secure.

In an investigation of The Model’s predicted inferences in text comprehension, Lea et al. (1990) and Fisch (1991) presented story vignettes of four or five sentences each, and required subjects to judge whether or not a final sentence makes sense in the context of the story. These judgments required integration of logical information in reading the stories—corresponding to the introduction and elimination inferences of the core and feeder schemas of The Model. The stories were isomorphic in logical form to a parallel set of box and blackboard problems, on which subjects made the inferences predicted by The Model. Almost all subjects made the appropriate judgments on the story vignettes, demonstrating that they must have made the basic logical
inferences described by The Model. Following this, on each story subjects were asked to judge each of three different sorts of statements. One was a paraphrase of information in the story, one was the output of a core inference of The Model, and a third was the output of a valid inference of standard logic but was not predicted by The Model. Subjects were asked whether the information in each of the test sentences was presented in the story or had to be inferred from other information in the story. Whereas the non-Model logical inferences were judged as requiring an inference, both the paraphrase items and the items predicted by the core schemas of The Model were judged as having been presented in the story. Fisch (1991) found that subjects, including 9 and 10-year-olds, judged Model-predicted core items as having been presented in the stories even when these inferences were not required to comprehend the story. Thus, the inferences predicted by The Model are made in text comprehension so effortlessly that neither school children nor adults were aware that they made the inferences.

In sum, people seem to behave in the ways that The Model predicts both on logical-reasoning problems and in text comprehension. The Model successfully predicts which problems subjects solve, the relative perceived difficulty of the solved problems, response times on simple problems, and the order in which inferences are written down. I know of no competing model that has had this sort of empirical success.

**Indirect Evidence.** The basic schemas should be available across languages and cultures. All natural languages should have regular ways of expressing conjunction, disjunction, conditionality, and negation, corresponding to such English language words as *and, or, if,* and *not.* These words should enter early in language acquisition, and the early usage should be like that of adults. Although no exhaustive search has been made of which I am aware, all of those languages that have been surveyed do have such expressions. Across the half-dozen languages surveyed, *and* and *not* appear in speech in the second year, and *or* and *if* appear in the third year (Bates, 1974; Bloom, Lahey, Hood, Lifter, & Feiss, 1980; Bowerman, 1986; Kuczaj & Daly, 1979; Lust & Mervis, 1980; Pea, 1980; Reilly, 1986). These studies show that early use of these particles is like that of adults, these particles are applied across a wide variety or situations and content from the beginning of their use, and the particles are used in ways that are consistent with the basic schemas of The Model.

**A REPLY TO SOME ARGUMENTS AGAINST MENTAL LOGIC**

No shortage of reasons have been put forth to deny the existence of a mental logic, and those discussed here are not intended to be exhaustive, but rather
to be instructive. One argument against mental logic stems from the failure of most people to solve a variety of laboratory logical-reasoning tasks. Most notable are the algebraic-content versions (i.e., about arbitrary letters, numbers, shapes, colors, etc.) of Wason’s selection task and the THOG task, a failure that has been interpreted as an impeachment of mental logic, particularly when compared with successful solution on some meaningful content versions (e.g., Cheng & Holyoak, 1985; Cosmides, 1989; Griggs & Cox, 1982; Johnson-Laird, Legrenzi, & Legrenzi, 1972).

On Wason’s selection task (Wason, 1968) subjects are presented four cards showing, for example, A, D, 4, and 7, respectively. They are told that each card has a letter on one side and a number on the other, and are presented with a conditional rule for the four cards, such as, “If a card has a vowel, then it also has an even number.” Finally, they are told that the rule may be true, but could be false, and are asked to select those cards, and only those cards, one would need to turn over for inspection to test the truth status of the rule. Typically, few people are able to select correctly only the cards showing A and 7 (see Evans, 1982, for a review).

Cheng and Holyoak (1985) noted that in some studies up to 20% of subjects can fail to select the card showing A, which they interpret as evidence against the ubiquitous availability of modus ponens. However, consider the line of reasoning that is required to select this card. To solve the problem one begins by supposing that the rule is true. The card showing an A provides a satisfying instance of a vowel, and taken together with the supposition of the rule, by modus ponens it follows that the other side of the card must show an even number. At this point there is still no reason to turn over the card; this realization requires the reasoner now to consider the possibility of there being instead an odd number, in which case the supposition that the rule is true could be falsified by reductio ad absurdum, because there might be an odd number where there has to be an even number. Note, then, that selection of the card showing an A goes well beyond making a modus-ponens inference. Indeed, it exceeds the basic skills of The Model. The other potentially falsifying card, showing the number 7, requires an even more complex line of reasoning, with an additional reductio embedded under the original supposition of the rule.

With the exception of Piaget (see Beth & Piaget, 1966, p. 181, which seems to indicate that the selection task should be solvable), none of the theories of mental logic predicts that such problems will be solved, and the basic parts of The Model, including the core and feeder schemas together with the direct-reasoning routine, are nowhere near being sufficient to solve such problems. Thus, failure to solve these complex problems does not count as evidence against The Model.

The literature now includes an algebraic-content version of the selection task that many subjects solve (Griggs, 1989). This problem presents four
cards drawn from two decks of ordinary playing cards, one deck with red backs and the other with blue backs. The rule states that “If a card has a value greater than 6, then it must have a blue back.” Perhaps subjects perform better on this task version because they can anticipate more easily what might be on the other side of the card. Whatever the reason, these correct responses are difficult to explain unless one assumes some logical reasoning.

Another complex logical reasoning problem that few people are able to solve is the THOG task (Wason & Brooks, 1979), and its solution also requires a reasoning strategy that goes will beyond the basic parts of The Model. Recently, several versions of the THOG task have been reported that many people solve (Girotto & Legrenzi, 1989; O’Brien, Noveck, et al, 1990). None of these task versions seem solvable by nonlogical means, and the finding that some algebraic-content versions of the selection and THOG tasks can be solved indicates that many adults have developed considerable logical-reasoning skills that go well beyond the core and feeder schemas and the direct reasoning program of The Model.

I discuss the claim that some realistic-content versions of the selection task have led to successful solution later in this chapter. For now I limit my comments to how we should interpret the effects of content on reasoning. Cheng and Holyoak (1989) stated that “content effects cannot be explained by theories based solely on formal rules” (p. 286), and I agree. Likewise, theories based entirely on content-bound processes cannot explain success on algebraic-content problems, such as the box and blackboard problems discussed earlier. This is why I advocate the three-part theory. A finding of a content effect presents no problem for The Model, because it does not claim exclusivity for the logical inference schemas. The reason that some algebraic-content problems are more difficult than their realistic-content isomorphs is that the algebraic-content versions are intentionally impoverished—they do not readily evoke an understandable state of affairs and are disconnected from the ordinary concerns of propositional activities.

Not all errors on laboratory logical-reasoning tasks are made on complex problems. For example, on the same problem sets on which subjects (both adults and children) routinely make modus-ponens inferences, they also often commit the denial-of-the-antecedent fallacy. For example, Rumain, Connell, and Braine (1983) reported that given the assumptions “If there is a dog in the box, then there is an apple” and “There is a dog in the box,” people almost always conclude “There is an apple,” an inference that follows straightforwardly by modus ponens. When given the same first assumption, but with “There is not a dog in the box” as a second premise, many people erroneously accept “There is not an apple.” This fallacy disappears, however, when (a) the asymmetry of the conditional is made explicit, for example, “If there is a dog in the box then there is an apple, but if there is not a dog,
there may or may not be an apple,” and (b) when an additional conditional assumption is provided, for example, “If there is a lion in the box, there is an apple” (see also Markovits, 1984, 1985). Inference-schema theorists, such as Rumain and colleagues, have interpreted the finding that such fallacies can be suppressed as indicating that the fallacies are invited rather than basic inferences.²

An additional empirical argument against mental logic concerns the demonstrated effects of a variety of nonlogical heuristics, most notably the use of matching (e.g., Evans, 1982; Evans & Lynch, 1973). The matching hypothesis holds that many responses on a variety of logical-reasoning tasks are merely matches to the values named in the task’s rule, for example, selecting the cards showing a vowel and an even number in the original selection task. Some empirical evidence has been reported that supports the claim, particularly the manipulation of negatives with the selection task. However, the matching hypothesis has not yet been worked out in sufficient detail so that its predictions are always apparent. Although there is agreement concerning the matching-hypothesis predictions for the selection task, this is not so for the THOG task. The problem presents four designs—a black triangle, a white triangle, a black circle, and a white circle. Subjects are told that the experimenter has written down one of the shapes and one of the colors, and that any design is a THOG if, and only if, it has either the shape or the color written down, but not both. Told that the black triangle is a THOG, subjects are asked to classify each of the remaining designs as (a) definitely a THOG, (b) insufficient information to decide, or (c) definitely not a THOG. Responses predominantly fall into two erroneous response patterns: Pattern A, in which the white triangle and the black circle are judged as possible THOGs and the white circle is judged as definitely not a THOG; and Pattern B, in which the white triangle and the black circle are judged THOGs and the white circle is judged as definitely not a THOG. Evans (1982) used the matching hypothesis to account for Pattern A, whereas Griggs and Newstead (1983) and Girotto and Legrenzi (1989) claimed that the matching hypothesis predicts Pattern B. To date, then, the matching hypothesis has not been described in sufficient detail to know what its predictions are. Griggs and Newstead, and Girotto and Legrenzi took a strong view of matching, and assume that subjects are matching without using any logical resources. Evans, though, described a situation in which subjects use both the matching heuristic and some logical processes.

² The original version of this chapter included at this point a discussion of the claim made by Byrne (1989) that modus-ponens inferences can be suppressed. This discussion now is presented in an expanded form in chapter 15.
Evans and I are in agreement on an essential point—human reasoning consists both of logical and nonlogical processes. We differ, though, in emphasis. In Evans’s view, problems are first processed by nonlogical heuristics, which are sufficiently primitive that our fairly weak logical resources are rarely of much use. Indeed, Evans sees such little use for a mental logic that he has not described what he thinks it includes. From the perspective of mental logic, I see people using logic frequently, both in ordinary tasks and in laboratory-reasoning tasks. When mental logic can be applied straightforwardly, it will be. When the reasoner has available the strategic skills required to solve a complex problem, and grasps the problem’s requirements, mental logic will allow solution. Otherwise, the problem solver has no other recourse but to nonlogical heuristics or to pragmatic inferences.

Finally, I consider the conceptual question raised by Johnson-Laird (1983): Which logic do we propose is the appropriate model for a mental logic? After all, infinitely many possible logic systems for quantifiers and modal operators could be constructed. Part of the answer is intuitive. I assume a two-valued logic of truth and falsity rather than a three-valued logic. Esoteric logical systems are just that: esoteric. As to which schemas for a two-valued logic should be included, I think this an empirical question—those that are easily expressible across languages, that people make readily, and that appear early in language acquisition should be included. Unlike some theories, the inference-schema models make a wide variety of specific empirical predictions that can be tested and compared, and the appropriate specific features will emerge from experimental investigations motivated by the existing models.

COMPETING NONLOGICAL THEORIES HAVE THEIR OWN PROBLEMS

This section is not meant to provide an exhaustive review of the competing nonlogical theories. First I address the content-bound theories proposed by Cosmides (1989) and by Cheng and Holyoak (1985), and then the mental-models approach of Johnson-Laird (1983) and Johnson-Laird and Byrne (1991).

Content-Bound Theories

Cheng and Holyoak (1985, 1989), Cheng, Holyoak, Nisbett, and Oliver (1986), and Holland, Holyoak, Nisbett, and Thagard (1986) proposed that people reason typically not using logical inference schemas, but using inductively learned rules defined in terms of classes of goals, such as taking desirable actions or predicting future events. Thus far, their theory has described only two rules, one for permission and one for obligation. The
permission rule holds that: If the action is to be taken, then the prerequisite must be fulfilled.

Cosmides (1989) also argued that people rarely reason “according to the canons of logic” (p. 191), and proposed that because 99% of human bioevolutionary history has consisted of hunter/gatherer activities, our biological endowment includes special abilities to reason about social contracts and their associated costs and benefits. These social contracts have a conditional form: If one takes a benefit then one must pay the cost. Social-contract rules appear to be a subset of permission/obligation rules (paying a cost is a special case of fulfilling a prerequisite, and taking a benefit is a special case of taking an action). Thus, supporting evidence for the social-contract theory would also be supporting evidence for the pragmatic-schemas theory. Cosmides would need to show that only problems presenting social-contract rules are solved, and the larger class of other permission/obligation problems are not. This does not seem to be the case. Pollard (1990), Cheng and Holyoak (1989), and Manktelow and Over (1990) showed that subjects are not influenced by the degree of costs or benefits, and that subjects can solve some permission task versions that have neither perceivable costs nor benefits.

The empirical evidence both for the social-contract and for the pragmatic-reasoning-schemas theories thus far has been limited to performance on some quasi selection-task versions, such as the “Drinking-Age Problem” introduced by Griggs and Cox (1982). The subject is told to imagine being a policeman enforcing the rule that “If a person is drinking alcohol, that person must be at least 21 years old.” Four cards are presented, each with a person’s age on one side, and their beverage on the other. Subjects are instructed to turn over those cards, and only those cards, that might lead to the discovery of a rule violator. Most people are able to select the logically appropriate cards (the card showing someone who is underage and the card showing an alcoholic beverage).

Note that the realistic-content versions of the selection task that have led to solution, such as the “Drinking-Age Problem,” are structurally distinct from the selection task. On the pragmatic-schemas problems, the rule is assumed true and can be used directly to draw a conclusion. This is not a trivial difference, because the original selection task is a metalogical rather than a logical reasoning problem. After all, the title of Wason’s (1968) seminal paper was “Reasoning About a Rule,” not “Reasoning From a Rule.” On these quasi-selection tasks, the need to turn over the card corresponding to A in the original task follows straightforwardly by modus ponens. The problem is inherently easier.

The most impressive evidence for the pragmatic-reasoning-schemas theory has been Cheng and Holyoak’s (1985) abstract permission problem. On this
problem subjects are told to imagine that they are working in a company enforcing the rule, “If a person takes Action A, then that person must first fulfill Prerequisite P.” The four cards show: “has taken Action A,” “has not taken Action A,” “has fulfilled Prerequisite P,” and “has not fulfilled Prerequisite P.” This problem differs from the original selection task in three crucial ways: First, it requires a search for a rule violator rather than a test of the rule’s truth status. Second, it presents what Jackson and Griggs (1990) referred to as a checking context, that is, subjects are asked to assume the role of a rule enforcer. Third, the cards present explicit rather than implicit negatives—“has not taken Action A,” rather than the card showing D. Jackson and Griggs found that when any of these three task features are changed to parallel the original task, subjects fail to solve the abstract-permission problem. Recently, in work with Ira Noveck and with two undergraduate students at Baruch College, I have found nonpragmatic algebraic versions that include these three crucial features and that many people are able to solve. It may be that solution of the abstract-permission problem has nothing to do with the permissionary nature of the rules but stems from these other extraneous task features.

I agree with Cosmides that our reasoning skills are the result of our bioevolutionary history, but this history has provided us with some basic logical intuitions that make propositional language and reasoning possible. I see no a priori reason that evolution should provide domain-specific processes but not general processes. Species with overly specified behavioral traits are at an evolutionary disadvantage when their environmental situation changes. A set of content-free inference procedures would be of evolutionary benefit, providing a basis for logic particles that allow communication in a wide variety of situations. To the extent that people have some content-bound inference procedures, they coexist with a set of more general logical inference procedures.

Mental Models

Johnson-Laird (1983) and Johnson-Laird and Byrne (1991) proposed that when people process discourse, they construct internal representations, called mental models. Inferences are drawn from models by describing information explicitly represented in them. Reasoning consists of searching for alternative models that could falsify a tentative conclusion. Whereas the mental-logic approach does not claim exclusivity, the mental-models approach claims that it can account for all reasoning, and that people never use inference schemas of the sort in The Model.

Johnson-Laird and his colleagues have not provided a clear description of what a mental model is. Mental models can be images but clearly are intended to go beyond images. As was discussed earlier, images are not propositional,