9

Commonsense Reasoning, Logic, and Human Rationality

MIKE OAKSFORD & NICK CHATER

1. Introduction

In cognitive psychology it is widely assumed that our understanding of human deductive reasoning will generalize to provide an understanding of everyday, commonsense reasoning and of human thought in general (Johnson-Laird 1983; Johnson-Laird and Byrne 1991; Rips 1994). This assumption presupposes that, in everyday life, much of human reasoning is deductive. Contrary to this view, we argue that almost no commonsense inferences are deductive. Moreover, we view this fact as diagnostic of the mismatch between the normative principles of logic and actual reasoning behavior on "deductive" reasoning tasks (e.g., Evans 1982, 1989; Evans, Newstead, and Byrne 1993; Johnson-Laird 1983; Johnson-Laird and Byrne 1991; Rips 1994; Wason and Johnson-Laird 1972). This mismatch has been taken to argue that humans may be irrational (Stein 1996; Stich 1985, 1990). In contrast, we argue that behavior on these tasks reflects rational, commonsense reasoning strategies that are appropriate to reasoning about our uncertain world. We thereby throw into question the significance of current psychological theories of deductive inference for understanding commonsense reasoning. We suggest that performance on many existing laboratory reasoning tasks can be better understood in nondeductive, commonsense terms.

Our argument is as follows: The most important issue for the cognitive science of reasoning is whether deduction provides a *computational*-level theory (Marr 1982) of a substantial amount of everyday, commonsense thought. Three sources of evidence appear to have the potential to settle the question decisively. These are the analysis of corpora of commonsense inferences; direct tests of performance on "deductive" reasoning tasks; and a priori considerations from computer science. We show that despite their superficial plausibility, none of these sources of evidence decides the question at issue. However, a more sophisticated interpretation of these sources of evidence can provide answers. First, the analysis of corpora circularly presupposes a standard logical analysis of individual arguments. It is more appropriate to investigate directly the kinds of reasoning that underpin organized systems of knowledge in which particular natural-language arguments (such as would be found in a corpus) are embedded. This is the subject matter of epistemology and the philosophy of science. Second, the study of reasoning tasks circularly presupposes that people interpret the tasks as deductive reasoning problems. It is more appropriate to contrast deductive and nondeductive characterizations of *both* interpretation and reasoning and see which provides the best account of the empirical data. This is the domain of the psychology of reasoning. Third, a priori considerations from computer science (suggesting that all computations are, in a sense, deductive) are too abstract to bear on issues relevant to the cognitive science of reasoning. It is more appropriate to consider whether computational systems based on deductive reasoning, in the sense employed in cognitive science, can provide the basis for systems that can reason about the everyday, commonsense world. This project has been extensively considered in AI. We argue that these lines of argument, from three different fields of inquiry—epistemology, AI, and psychology of reasoning—converge on the conclusion that deduction plays no significant role in commonsense reasoning about the everyday world.

This chapter is organized as follows: We begin by outlining what deduction is, in abstract terms, and then consider various ways in which it can be related to human reasoning, using the framework of D. Marr's (1982) levels of explanation. We argue that the fundamental issue for the cognitive science of reasoning is whether deductive logic provides a computational-level theory of human reasoning. We then consider how this question can be addressed. We first consider and reject the three superficially plausible methods of resolving the question of the prevalence of deduction introduced earlier. We then develop three more sophisticated lines of argument, from epistemology, AI, and the psychology of reasoning. We show that each argument supports our conclusion that deduction has no significant role in commonsense reasoning. Finally, we consider the implications of rejecting deduction for the cognitive science of human reasoning.

2. What Is Deduction?

Classically, in a valid deductive argument, the conclusion must be true if the premises are true. That is, deductive inferences are certain. Such inferences guarantee that the conclusion is true, if the premises are true, independent of any other information. Therefore, if a conclusion follows deductively from a set of premises, then it must also follow deductively from that set of premises conjoined with any set of additional premises (see Oaksford and Chater 1998, chapter 1). That is, nothing can overturn a deductive inference. Suppose, for example, that the commutativity of addition (that is, $x + y = y + x$, for all x and y) can be deduced from some axiomatic formulation of arithmetic. It therefore follows that commutativity holds, *whatever* further axioms are added. This property is known as monotonicity, and it is a crucial property of axiomatic systems in mathematics. Moreover, many logicians regard this property as the defining feature of a logical system (Curry 1956).

Other modes of reasoning, which are not deductively valid, are "nonmonotonic"—adding premises can lead to conclusions being withdrawn. An important example is induction, in which general laws or regularities are inferred from particular observations. At any time it is possible that a new observation may conflict with the regularity and undermine it. For example, a new observation of a nonblack raven logically undermines the inductive inference that *all ravens are black* based on the observation of numerous black ravens. Thus, adding a new premise (a new observation) can remove the conclusion, and hence induction is nonmonotonic. Another example is abduction, which typically involves inferring causes from their effects. For example, in a detective mystery a particular set of clues might, for example, suggest that the butler is the murderer. But a new and decisive clue (e.g., the chauffeur's bloodstained shirt) might overturn this conclusion. Thus, abduction is also nonmonotonic and hence not deductive.

3. Deduction and Reasoning?

The claim that human reasoning involves deduction can be understood in a number of different ways. We can understand these different interpretations in terms of two of Marr's (1982) three levels of computational explanation.

Marr's highest level of analysis is the *computational* level, where "the performance of the device is characterized as a mapping from one kind of information to another, the abstract properties of this mapping are defined precisely, and its appropriateness and adequacy for the task at hand are demonstrated" (1982, p. 24). Marr uses the example of a cash register. The theory of arithmetic provides the computational-level analysis of this device, and its appropriateness is demonstrated by showing that our intuitive constraints on the operation of a cash register map directly onto this mathematical theory (1982, p. 22). In the case of human reasoning, psychologists have typically assumed that deductive logic plays the role that arithmetic plays for the cash register; that is, logic characterizes the inferences people draw. Whether or not this is true is clearly an empirical question, just as it is an empirical question whether or not a particular piece of machinery functions as a cash register.

Marr's *algorithmic* level describes how to compute the function specified at the computational level. This level also involves specifying the representations that the algorithm manipulates in computing the function. Thus, in the case of the cash register, using Arabic numerals as the representations involves using the standard rules "about adding the least significant digits first and 'carrying' the sum if it exceeds 9" (1982, p. 22) as an algorithm. Although the choice of algorithm is constrained by the choice of representation, it is not uniquely constrained—there may be several ways of computing a certain function that use the same representation. It could be the case that deduction provides a crucial component at the algorithmic level. In computer science, this idea is embodied in theorem provers, which are computational systems for proving logical theorems. Theorem provers can be used to reason about the everyday world, given axioms that embody everyday knowledge. They have also been used to construct general programming languages, such as PROLOG (Clocksin and Mellish 1984).

Hypotheses about deductive reasoning at the computational and algorithmic levels have been prevalent within the psychology of deductive reasoning. Many theorists argue for deduction at both levels. For example, B. Inhelder and J. Piaget (1958, p. 305) go as far as to say that human "reasoning is nothing more than the propositional calculus itself." For Piaget, attainment of the formal operational stage in cognitive development is, by definition, revealed in the ability to show logical reasoning behavior. Thus, logic is viewed as an appropriate computational-level description of mature human behavior. But, moreover, the quotation earlier reveals that the mechanism that achieves this performance is itself logic. This view is still widely advocated in current psychology of reasoning, by advocates of "mental logic" (Braine 1978; Braine and O'Brien 1991; Henle 1962; Lea et al. 1990; O'Brien, Braine, and Yang 1994; Pollitzer and Braine 1991; Rips 1983, 1994; for a collection on these issues, see Macnamara and Reyes 1994). For example, Rips (1994, p. viii) argues for what he calls the Deduction System Hypothesis: that logical "principles . . . are central to cognition because they underlie many other cognitive abilities . . . [and] that the mental life of every human embodies certain deduction principles."

Although the claims that logic has a role at the computational and algorithmic levels are often held together, they are clearly independent. It is possible that while logic characterizes the behavior of a device at the computational level, the algorithms that produce the behavior are not themselves logical. This viewpoint is explicitly advocated in the psychology of reasoning by J. Macnamara, who also places deductive logic at the center of human cognition but articulates this thesis more guardedly: "A logic that is true to intuition in a certain area constitutes a competence theory [in Chomsky's (1965) sense] for the corresponding area for cognitive psychology" (Macnamara 1986, p. 22). As Marr (1982) notes, "competence theory" is simply another way of talking about a computational-level account. Mental model theory (Johnson-Laird 1983; Johnson-Laird and Byrne 1991) explicitly takes the view that logic is part of the computational-level theory of reasoning.[1] But mental model theory is typically viewed as not involving logical inference at the algorithmic level. Instead mental model theory assumes that deductive reasoning involves the construction and manipulation of mental models.[2]

The converse position is also possible. The algorithms that underlie thought might follow deductive logic, but the behavior that results from those algorithms might be best characterized in nondeductive terms. For example, a theorem prover could implement list-handling operations or arithmetic. Therefore, the computational-level characterization of what the program is doing will involve descriptions of list manipulation or arithmetical calculation, rather than logical proof. Within the psychology of reasoning, this viewpoint has not been explicitly advocated, as far as we know.[3] However, it is reasonable to interpret influential theorists in the foundations of cognitive science, such as J. A. Fodor, as advocating this position. Thus, Fodor and Z. W. Pylyshyn (1988, pp. 29–30) argue that "it would not be unreasonable to describe Classical Cognitive Science as an extended attempt to apply the methods of proof theory to the modeling of thought," and they proceed to defend this position strongly. Because proof theory is the mechanism by which deductive inferences are made, this amounts to the claim that cognition is deductive at the algorithmic level. But Fodor (1983) also argues extensively that almost all aspects of thought are "nondemonstrative," that is, nondeductive, in character (and we shall outline some of these arguments later). Therefore, Fodor seems to reject deduction as a computational-level theory of reasoning but embraces logic as an algorithmic theory.

In this chapter, we shall focus primarily on whether deductive logic provides an appropriate *computational*-level description of human reasoning, rather than dealing with the algorithmic level. As we have seen, the assumption that deduction does provide a computational-level description for much human inference is shared by many contemporary researchers on reasoning, including advocates of mental logics and mental models. Of course, no theorist would propose that deductive logic could provide a computational-level theory of all aspects of human reasoning, such as reasoning under uncertainty (Tversky and Kahneman 1974), decision making (Baron 1994; Tversky and Kahneman 1986), and abductive (Gluck and Bower 1988), and inductive (Gorman and Gorman 1984; Wason 1960) reasoning. Instead, deductive logic is assumed to provide a computational-level account of an important class of human reasoning. Moreover, deductive logic is assumed to play at least a partial role in almost every other aspect of cognition (Johnson-Laird and Byrne 1991; Macnamara and Reyes 1994; Rips 1994). For example, P. N. Johnson-Laird and R. M. J. Byrne (1991, pp. 2–3) argue for the centrality of deduction

> because of its intrinsic importance: it plays a crucial role in many tasks. You need to make deductions in order to formulate plans and to evaluate actions; to determine the consequences of assumptions and hypotheses; to interpret and formulate instructions, rules and general principles; to pursue arguments and negotiations; to weigh evidence and to assess data; to decide between competing theories; and to solve problems.

Thus, the idea that a deductive competence theory is central to human cognition both has a long pedigree and is widely held by many leading figures in the psychology of reasoning.

In this chapter, we argue against this tradition in the psychology of reasoning. We claim that almost no commonsense human reasoning can be characterized deductively or has any significant deductive component. Although many theorists have argued that deduction is at the core of cognition, we argue that it is at the periphery. From this point of view, we shall argue that the "errors and biases" observed in "deductive" tasks in psychological experiments should be understood not as failed deductive reasoning but as successful nondeductive reasoning. Consequently, these "biases" do not provide evidence for human irrationality; rather, they reveal the nature of people's commonsense reasoning strategies. Other theorists have argued for a similar position (e.g., Cheng and Holyoak 1985; Cosmides 1989; Fischhoff and Beyth-Marom 1983; Gigerenzer and Hug 1992; Holland et al. 1986; Holyoak and Spellman 1993; Klayman and Ha 1987). But these theorists have focused on whether particular tasks are deductive, rather than on whether deduction provides a satisfactory computational-level description of real, commonsense reasoning.[4] It is this wider question that is the central concern of this chapter.

4. Deduction and Common Sense: False Trails

There are three methods of investigating the question of how much commonsense reasoning is deductive at the computational level, which initially appear to provide decisive answers. First, the collection and analysis of corpora of commonsense arguments promises to reveal the statistical prevalence of deductive reasoning directly. Second, presenting people with deductive reasoning tasks should directly tap any underlying deductive competence. Third, a priori considerations from computer science appear to decide the question, before any empirical investigation is carried out: specifically, any computational process whatever can be viewed as deductive. In this section, we show that none of these considerations can decide the question. In the next section, we show that when these sources of evidence are viewed from a more sophisticated perspective the question can be genuinely addressed. The three sources of evidence provide three criteria of adequacy on theories of reasoning, from epistemology, AI, and the psychology of reasoning, none of which can be met by a cognitive science of reasoning using deduction as its computational-level theory.

Looking at Commonsense Argument

This strategy involves collecting and analyzing corpora of commonsense natural-language arguments and deciding what fraction of them are deductively valid. If deduction is prevalent in commonsense argument, it is also likely to be prevalent in reasoning, on the reasonable assumption that what people say is closely related to what they think.

The problem is that whether or not a natural-language argument is deductive cannot be straightforwardly ascertained by purely logical analysis. Consider, for example, the argument:

Birds fly.
Tweety is a bird.
Therefore, Tweety flies.

One way of assessing the validity of this argument is to translate it into a logical formalism, such as the predicate calculus, as follows:

$$\begin{array}{l} \forall x(\mathit{bird}(x) \rightarrow \mathit{flies}(x)) \\ \mathit{bird}(\mathit{Tweety}) \\ \hline \mathit{flies}(\mathit{Tweety}) \end{array}$$

According to the logical properties of the predicate calculus, this is a deductively valid argument: in logical terms, there is no model in which the premises are true, but the conclusion is false.

But this only means that the original argument is deductively valid if the *translation* from natural language into the logical language is accepted as capturing the "logical form" of natural-language statements (Haack 1978). In practice, this step is frequently highly controversial. For example, even the logical terms $\neg$, $\wedge$, and $\vee$ are notoriously distant relatives of their natural-language counterparts *not, and* and *or* (e.g., Hodges 1977; Horn 1989; Lemmon 1965). The relation between quantifiers $\forall$ and $\exists$ and universal and existential quantification in natural language is even more complex (Barwise and Cooper 1981). This is particularly true when, as in our example sentence "Birds fly," the quantification is not explicit. Should this sentence be treated as meaning that *all* birds fly? Or is a better interpretation that *most* birds fly, that *normal* birds fly (McCarthy 1980), or perhaps that it is reasonable to assume that a bird flies unless there is reason to believe the contrary (Reiter 1980, 1985)? On any of these latter interpretations, the conclusion of the preceding inference does not follow deductively—Tweety may be one of the exceptional nonflying birds. Therefore, whether or not a natural-language argument is deductive depends on how the premises and conclusions are translated into logic—and this translation is a highly controversial matter.

Moreover, both philosophers (e.g., Davidson 1984; Quine 1960) and psychologists (Smedslund 1970) have pointed out that there is a circularity in the relationship between studying reasoning and studying the meaning of what people say. That is, which translation of a natural-language statement is correct depends on how people *reason* with that statement.[5] For example, is the statement rejected as soon as a counterexample is found? Will a person wager an arbitrarily large sum of money against the possibility that the premises are true but the conclusion false? The logical form of a statement is intended to capture the patterns of reasoning in which it figures, and

hence the nature of reasoning with the statement constrains the choice of translation. But, of course, discovering how people reason with a statement is a question in the psychology of reasoning. So psychologists cannot look to a purely logical analysis of natural-language arguments as a neutral way of assessing how much deduction people do, because the appropriateness of any particular logical analysis itself depends on how people reason.

The fact that logical analysis of natural-language arguments itself depends on psychological considerations rules out the obvious methodological strategy of collecting and analyzing a corpus of commonsense arguments.

Using Deductive Reasoning Tasks

Instead of collecting a corpus, a more appropriate strategy might be to experimentally test people's ability to solve deductive reasoning problems. However, the problem that logical analysis is not psychologically neutral also applies to the interpretation of materials in psychological experiments (Smedslund 1970). In setting up putative "deductive reasoning problems," psychologists have typically assumed a particular logical analysis of the natural-language statements that people are asked to reason over. This analysis is typically accepted uncritically from logical texts, and the psychological presuppositions involved in adopting a particular analysis are not examined. But if the logical analysis chosen turns out to be psychologically inappropriate (for example, a quantifier is interpreted as meaning "all" when it should be interpreted as "almost all"), then serious consequences may follow. First, the psychologist may not be studying deductive reasoning at all. Second, the logical misanalysis will lead the psychologist to assume that people are making reasoning errors (in terms of the wrong logical analysis), while they may be reasoning correctly (in terms of the psychologically appropriate analysis). Third, the psychologist will postulate biases and other reasoning limitations to explain the apparent reasoning errors and will be unable to discover any rational pattern in people's performance. In our discussion of the psychology of reasoning later, we argue that these problems lie at the heart of current theories of the psychology of deductive reasoning. We argue that most "deductive" reasoning tasks are not deductive at all and that human data can be rationally explained as involving uncertain, commonsense reasoning, rather than deduction.

At the heart of this issue is the question "what is the appropriate computational-level description of human behavior on 'deductive' reasoning tasks?" Although logic may have inspired a particular task, it is an empirical question whether people interpret the task deductively. We illustrate this point by an analogy with discovering the function of an unknown device. Suppose it has a keypad on it, a digital readout, and a coin slot. You therefore form the reasonable hypothesis that it is a public pay calculator; that is, its function is to perform arithmetic calculations for you when you put a coin in the slot. To test this reasonable hypothesis you may enter a coin and punch in some numbers in the following sequence: "2," "*," "4," "=." Your hypothesis leads you to predict that the digital readout some short time later will be "8." Three possibilities arise: (1) Suppose that the readout is indeed "8," which seems to confirm your hypothesis. It would not take many more such successful tests before you were reasonably confident that this device was indeed a pay calculator. (2) Suppose, however, that you adopt the same hypothesis, but now, after placing your coin in the slot and pressing the appropriate keys in the preceding test, the digital readout was "10." This would seem to suggest that the device was not a calculator after all. However, you continue to test the device in the same way—perhaps you think quite reasonably that even machines can make occasional mistakes—and still find no consistent arithmetic relation between the inputs and the outputs. At this point you should become reasonably convinced that the device is not a calculator. (3) You then notice that the number that comes up at each trial in fact corresponds to the value of the coin you insert into the device. You therefore come to the conclusion that the device is actually a pay phone.

We suggest that contemporary psychology of reasoning theory is in situation (2) earlier. Researchers have assumed that the tasks they present people are deductive but then fail to elicit logical performance. However, rather than abandon the view that people interpret these tasks logically, researchers have inferred that people have error-full deductive mechanisms. We will argue that reasoning researchers should instead search for more appropriate computational-level theories—specifically, they should explore computational-level theories of nonmonotonic, uncertain reasoning, rather than computational-level theories based on deduction.

In sum, the strategy of presenting people directly with deductive reasoning problems confronts the same problem of interpretation as studying corpora—you can never tell which arguments in the corpus, or which experimental tasks, are indeed deductive. Moreover, illogical performance on a task need not bear on whether people are rational, because they may interpret the task in such a way that some other normative theory applies by whose standards their reasoning is perfectly sound.

An Answer from Computer Science?

Perhaps all these empirical considerations are wholly unnecessary. Cognitive science is built on the assumption that cognition is computation. And there are deep results in computer science that suggest that all computations can be viewed as deductive, that is, logical, inferences.

All computational devices are equivalent to Turing machines (assuming the Church-Turing thesis). It is simple to write logical axioms that perfectly model the behavior of any Turing machine (Boolos and Jeffrey 1980). Thus, logic (assuming the appropriate axioms) provides a computational-level theory of the behavior of the device. In this sense, logic can provide a computational-level account of any computational process, including any postulated in cognitive science. Furthermore, the steps that underlie the Turing machine's behavior each correspond to a step in a logical derivation.

Therefore, logic also provides an algorithmic-level account of the Turing machine's behavior.

Similarly, programming languages are typically given two kinds of semantics, according to which they can be understood as performing logical inferences. Operational semantics maps the programming language onto machine operations. Denotational semantics typically maps the programming language onto some abstract mathematical structure, according to which, for example, algorithmic correctness can be proved (Girard 1989; Scott and Strachey 1971). According to both approaches, logic can provide a computational- and an algorithmic-level theory for any programming language.

It would therefore appear that the question of how much deduction there is in human cognition is decided the moment we adopt the assumption that cognition is computation. Any behavior that we can explain in computational terms must automatically have a logical interpretation both at the algorithmic and at the computational level. Consequently, it appears that *all* cognitive processes are ipso facto deductive.

However, there is clearly something seriously wrong with this line of argument. If it is accepted, then any cognitive task is necessarily a logical task. This would mean that arithmetic, reading, probabilistic reasoning, motor control, perceptual processing, syntactic analysis, and so on are all examples of logical reasoning. Furthermore, this means that all mundane computations, such as doing spreadsheets, word processing, solving differential equations, and doing actuarial calculations, have deductive logic as their computational-level theory. But this is close to a reductio ad absurdum of this line of reasoning. The point of computational-level theory is that it describes both the purpose of the computation and the objects and relations it is *about*. Thus, a computational-level theory of a cash register must involve numbers and numerical operations; a computational-level theory of syntactic analysis must deal with words, phrases, and sentences of natural languages; and so on. However, the objects utilized in providing the overarching logical analyses outlined earlier involve states and possible state transitions of a Turing machine or mathematical objects in abstract function spaces. A logical description in terms of these objects provides a computational-level description of a sort, but at such an abstract level of specification that it says nothing about the *point* of the computation. At this level of description, it is not possible to discriminate a computational process that carries out actuarial calculations from one that does word processing or produces natural-language utterances. The whole point of Marr's (1982) computational-level analysis is to make exactly these distinctions, to which these highly abstract analyses are insensitive.

In the case of human reasoning, it is crucial that a computational-level theory view human reasoning as about the objects and relations in the everyday world. The claim that an inference from "Birds fly" and "Tweety is a bird" to "Tweety flies" is (or is not) deductive only makes sense where we interpret these statements as referring to birds, flying, and Tweety, not to either states of a computational device or to objects in an abstract function space.

Therefore, the apparently decisive a priori arguments from computer science turn out to be entirely beside the point in determining whether or not human inference should be understood deductively. The results from computer science only apply to a level of analysis so abstract as to be of no practical value in constructing computational- or algorithmic-level theories of human reasoning.

Summary

The three possible ways of answering our question concerning the relationship between commonsense and deductive reasoning turned out, on closer inspection, to be inadequate. The first and second cases, analyzing corpora of inferences or directly testing people with "deductive" reasoning tasks, both founder on the same problem. This is that the interpretations of arguments and of tasks both circularly involve assumptions about how people reason. Moreover, a priori concerns from computer science are too abstract to differentiate deductive from nondeductive reasoning in any sense interesting to cognitive theory.

Each of these sources of evidence can, however, when viewed from a more sophisticated perspective, bear on the prevalence of human deductive reasoning at Marr's computational level of description. First, we have argued that analyzing natural-language corpora of inferences piecemeal would be futile, because of the problem of interpretation. This issue is directly addressed in philosophy, by classical epistemology and its contemporary counterpart, the philosophy of science. These philosophical projects attempt to characterize the inference patterns that underpin whole bodies of beliefs (such as scientific theories), in which specific arguments and inferences are embedded. Second, we have argued that simply testing people on "deductive" reasoning tasks confronts the same problem—that how participants interpret the task is unclear. But empirical psychology can answer this question, by providing theories of how tasks are interpreted *and* how to characterize people's reasoning given these interpretations, that is, by providing a computational-level explanation. Competing explanations at this level (e.g., involving or not involving deduction) can then be compared with the data to see which provides the best characterization of human reasoning behavior. Third, we have argued that a priori considerations from computer science are too abstract to be useful. But computer science can inform the question of how much deduction there is in commonsense reasoning via the practical attempt to build intelligent reasoning systems using logical methods (where logic is employed at the level relevant to psychology). This is not merely an abstract possibility but has been a popular research strategy in much of AI (Charniak and McDermott 1985). Therefore, the success or failure of this strategy can inform the viability of a deductive characterization of human thought. We now consider each of these more sophisticated approaches and argue that each reveals that little or no commonsense human reasoning is deductive.

5. Common Sense and Deduction: Lines of Evidence

If real human reasoning is based on deduction, then three criteria of adequacy can be adduced, which correspond to three more sophisticated research strategies mentioned earlier. First, unless we are to give up the claim that human reasoning has any rational justification, then a deductive account of how people *should* reason about the world must be viable. That is, deduction must be central to our theories of epistemology, which define the very standards of rationality against which we measure actual human reasoning performance. Second, any psychological theory must be computationally viable. Therefore, if deduction is the foundation of human thought, then it must be possible to design and successfully implement AI systems that reason about the real world using deduction. Third, a good psychological theory must, of course, be consistent with the data. Psychological theories of reasoning based on deduction should provide better fits to the data than accounts that reject deduction.

We argue that the claim that human reasoning is based on deduction at the computational level fails on all three counts: it represents an epistemologically outmoded tradition; it has not proved viable in AI; and, moreover, nondeductive accounts of human reasoning provide better fits to the empirical data.

Epistemological Adequacy

In this section our argument is in three parts. First, we argue that the view that deduction is the foundation for reasoning about the world reflects an outmoded epistemological tradition. Second, we consider what aspects of reasoning might be deductive and conclude that the obvious candidates turn out to be nondeductive in character. Third, we briefly consider and reject a possible defense of the centrality of deduction in psychology, which seeks to exploit the fact that epistemology considers how we *should* reason, whereas cognitive science is concerned with how people actually *do* reason.

DEDUCTION: A FOUNDATION FOR THOUGHT? Euclid provided the first systematic exploration of deductive reasoning (Coolidge 1940). Beginning with definitions and apparently self-evident axioms, he showed how purely deductive argument could establish a large class of geometrical truths. The Euclidean method has proved to be enormously productive not just in geometry but throughout mathematics.

But mathematical reasoning does not seem, superficially, to have much in common with commonsense thought. In particular, mathematics appears to be about establishing certainties that concern abstract objects (see Putnam and Benacerraf 1983 for discussion). In contrast, commonsense thought appears to be about making the best sense possible of an ill-defined, concrete external world, in which certainty is rarely, if ever, encountered

(Barwise and Perry 1983). Can deduction extend beyond the mathematical realm and provide a route to knowledge that concerns the external world? This question is crucial for the psychology of reasoning, for it concerns the scope that human reasoning might have. It is also a central question in the history of epistemology (Russell 1946).

Euclid's astonishing successes in geometry and the absence of any comparably impressive achievements that use other reasoning methods suggested that deductive reasoning could also provide a foundation for knowledge of the external world. From Plato to Kant, an influential line of philosophers has attempted to establish nonmathematical knowledge using deductive argument. Attempts to model scientific inquiry on Euclid's deductive model had a profound influence on Greek and medieval science (Russell 1946). Spinoza even went as far as using the Euclidean method in his *Ethics*, with definitions, axioms, and "proofs." It would not be unreasonable to suggest, then, that psychologists of reasoning have simply taken over a preoccupation with deduction that has been evident more widely in Western thought since Plato.

However, more recently, the view that science derives knowledge of the world by deduction from self-evident foundations has fallen into disrepute (Lakatos 1977a, 1977b; Popper 1959). Contrary to the Euclidean picture, science appears to proceed by plausible conjecture on the basis of observation, not by deductively certain inference. For example, Bacon explicitly advocated alternative inferential methods for what he called inductive reasoning (James 1975). In the twentieth century, it has become increasingly accepted that people derive knowledge of the world by different means from knowledge of mathematics (see, for example, Russell 1919). The nondeductive origin of scientific knowledge is a common thread that links diverse views in modern philosophy of science (e.g., Glymour 1980; Howson and Urbach 1993; Kuhn 1962; Lakatos 1970; Putnam 1974; Thagard 1988; Toulmin 1961; Van Frassen 1980). In epistemology more generally there is agreement that knowledge of the world does not have a deductive basis (see, e.g., Goldman 1986; Lehrer 1990; Pollock 1986; and Thagard 1988).

The deductive picture has, however, remained influential as an apparently unattainable standard against which philosophers may assess other knowledge-gathering methods. In particular, skepticism concerning knowledge of the world, from Descartes onward, has its roots in the distance between the certainty that deduction can assure and the lack of certainty that nondeductive empirical methods of inquiry provide (Burnyeat 1983). That no one has met the skeptical challenge to provide a certain grounding for knowledge reinforces our thesis that science does not obtain knowledge by deductive means.

So it seems that we have two paradigms of thought: mathematical knowledge, where deductive inference appears to be of primary importance; and empirical, scientific knowledge, in which deductive inference plays at most a secondary role. Both within epistemology and in psychology, there is agreement that we should view human thought as generally analogous to scientific, rather than mathematical, inquiry (Fodor 1983). For example, many

philosophers advocate the view that common sense and science are parts of the same general project of understanding the world, differing only by degree of systematicity and rigor (e.g., Goodman 1951; Quine 1960, 1990). Developmental psychologists frequently view the child as a "naive scientist" (Carey 1988; Karmiloff-Smith 1988); accounts of causal reasoning in adults also use the naive-scientist metaphor (Jaspars, Fincham, and Hewstone 1983; Kelley 1967). Further, psychologists view learning from experience in any domain as involving inductive inference (Holland et al. 1986), and so on.

As mentioned earlier, Fodor (1983) argues strongly for the nondeductive character of thought. He notes that the perceptual system attempts to infer the causes in the external world of the inputs to the sensory receptors and that such reasoning is an instance of inference to the best explanation (Harman 1965). This pattern of inference, like induction, is uncontroversially nonmonotonic. From a distance, the perceptual system may misclassify sensory input as being generated by a horse. However, on moving closer it may be apparent that it is actually a cow. The addition of new information overturns the original conclusion—additional information that indicates another explanation is better overturns what was previously the best explanation. Thus, perception involves nonmonotonic and hence nondeductive reasoning. Moreover, Fodor argues that what he calls "central" cognitive processes, of belief revision and commonsense thought, face a problem analogous to scientific inference. For reasons similar to those we have outlined earlier, Fodor believes that scientific inference is nonmonotonic and that it is not deductively formalizable. He therefore concludes that central cognitive processes will likewise be nondeductive in character.

To sum up so far: There is a long tradition in epistemology, initially inspired by Euclidean geometry, that attempts to provide deductive foundations for nonmathematical knowledge. According to this model, it seems quite reasonable to postulate that deduction is the foundation of human thought. However, the rise of science has involved plausible but uncertain inferences that do not fit this deductive pattern. Furthermore, human cognition is related to empirical, nondeductive inquiry, rather than deductive, mathematical inquiry. So the view that deduction provides the foundation for human thought may be unworkable, a vestige of an outmoded epistemological tradition.

DEDUCTIVE ASPECTS OF THOUGHT? The conclusion that deduction is not the foundation for thought does not, of course, imply that no thought is deductive. From the point of view of the psychology of reasoning, the study of deduction may still be of wide significance to psychology, if some substantial and important aspects of human reasoning are deductive. Indeed, many psychologists of reasoning appear, at least in some passages, to advocate this relatively modest position. Johnson-Laird and Byrne (1991), who advocate a central role for deduction, nonetheless concede that much human reasoning is not deductive in character. For example, they point out that even Sherlock Holmes, whose "powers of deduction" are legendary, does not really solve problems deductively at all. Johnson-Laird and Byrne note that Holmes's inferences are plausible conjectures, which although ubiquitous in everyday life are not the consequences of deductively valid arguments.

Pursuing the analogy with science, it is interesting that although philosophers of science have generally abandoned the view that scientific inference might be deductive, some continue to advocate the view that certain aspects of science involve deductive inference. Roughly, the view is that although forming theories does not involve deductive inference, deduction is crucially involved in prediction, explanation, and theory testing. Most notably the hypothetico-deductive account of prediction and hypothesis testing (Popper 1959) and the deductive nomological view of scientific explanation and prediction (Hempel 1965) advocate this view. If this view is right, we can conjecture that deduction plays an analogous role in human thought as in science. Thus, perhaps deduction does have an important, if not exclusive, role in human cognition.[6]

However, deductive views of science provide little support for the deductively inclined psychologist of reasoning. Contrary to the views of C. Hempel and K. R. Popper, recent philosophy of science has suggested that prediction, explanation, and hypothesis testing in science are not really deductive in character. Taking prediction as an example, according to the deductive view a prediction is a deductive consequence of a theory or a hypothesis together with some initial conditions. So, according to Hempel's view, we have the following picture:

(1) $$\text{Theory (T)} \wedge \text{Initial Conditions (I)} \models \text{Prediction (P)}$$

where the theory and initial conditions form the antecedent of an entailment and the prediction the consequent. For example, Newton's laws (T) together with information (I) about the state of the solar system at time, t_0, deductively imply particular trajectories (P) for the planets at subsequent times $t > t_0$. However, as critics have pointed out (Duhem 1954; Lakatos 1970, 1977a, 1977b; Putnam 1974; Quine 1953), this conclusion only follows *all other things being equal*. Real physical systems are not causally sealed off from external forces; they are open systems. Consequently there are limitless possible intervening forces and factors, that is, "auxiliary hypotheses" (Putnam 1974), that could intervene to make the prediction fail—unexpected frictional, electromagnetic, or as yet undiscovered forces, changes in physical constants in different parts of space/time, and so on.

The nondeductive character of scientific prediction becomes clear when the prediction does not fit with observation. If the prediction followed deductively from the theory, then the theory would automatically be falsified. However, in practice, the theory may be fine, because auxiliary hypotheses such as those noted earlier may be the cause of the mismatch between prediction and observation. H. Putnam (1974) notes, for example, that Newton's laws are entirely compatible with square orbits, given appropriate additional forces.

One possible defense for the deductive view is to argue that the predictions follow deductively from the conjunction of the theory under test and the auxiliary hypotheses. This suggestion involves modifying the antecedent of the entailment in (1) as follows:

(2) $$T \wedge I \wedge AH_1 \wedge AH_2 \wedge \ldots \wedge AH_n \models \text{Prediction (P)}$$

If the prediction turns out to be false, then the scientist may reject one of these auxiliary hypotheses, not the theory under test, thus bringing prediction and observation back into line. However, this requires a complete enumeration of all the auxiliary hypotheses, which appears to be impossible, even in principle (Putnam 1974; Quine 1953). Science is concerned with causally open systems where there are indefinitely many unexpected external factors. It is impossible to exhaustively enumerate these other possible factors, and hence scientific prediction cannot be made to fit into the deductive mold.

We have considered prediction in science at length because precisely analogous considerations arise in attempting to model commonsense reasoning in deductive terms (e.g., Schiffer 1987), as we shall see in the discussion of default rules in AI later.

We have so far argued that deduction is not appropriate for formalizing predictive and explanatory reasoning: that is, the derivation of empirical predictions about the world does not proceed by deductive inference. The final position we consider is that deductive reasoning leads not to empirical predictions but to conceptual or analytic truths. For example, perhaps it follows deductively from what it is to be a bachelor that bachelors must be male. Kant's program for establishing the properties of space, causality, and the like pursued this line: the idea was that these properties are not really properties of the external world but conceptual necessities of human thought (Scruton 1982). Perhaps this kind of inference is deductive.

Modern epistemology suggests that it is not, because the distinction between conceptual and empirical claims cannot be maintained. For example, W. V. O. Quine (1953) has forcefully argued that there is no noncircular way to characterize the distinction between analytic, conceptual truths and synthetic, empirical truths. According to contemporary epistemology, all statements are revisable in the light of experience—none are purely conceptual truths derived by deductive argument alone (although see Katz 1990 for an opposing point of view). Furthermore, the development of mathematics and physics since Kant has shown that "conceptual necessity" is unexpectedly flexible. For example, Kant took the Euclidean character of space to be conceptually necessary, but modern science has shown that space is in fact non-Euclidean (Putnam 1975). Hence, what appears to be conceptual knowledge, which might potentially be the result of deduction, may be empirical and thus not derived by deduction.[7] In sum, epistemological considerations suggest that deduction is not central to human thought in either making empirical predictions or establishing conceptual truths.

From the first part of the epistemological argument, we concluded that deduction cannot be the primary means of acquiring knowledge of the world. In the second part, we have searched for a substantial residual role for deductive reasoning and failed to find one. From an epistemological point of view, one might suspect that deduction plays only a small role in human reasoning.

We have shown that deduction does not provide an adequate epistemology. If deduction is never or rarely *justified*, then it seems unlikely to provide a useful design principle for intelligent systems, whether human or artificial. This suggestion is reinforced by the next two arguments, which concern AI and the empirical data.

A POSSIBLE DEFENSE? Epistemology and the philosophy of science are concerned with how people *should* reason, rather than how they *do* reason: it is *normative* rather than descriptive. This appears to leave open the possibility that the epistemological arguments described earlier may be accepted, but that they imply only that people *should not* use deduction in everyday life. This seems not to preclude the possibility that people actually *do* reason deductively about the everyday world.

But this defense fails for a number of reasons. First, notice that epistemological considerations show that interesting conclusions never (or almost never) follow deductively from known premises about the real world. This means that if people do reason deductively, then the conclusions that they will be able to draw will be entirely uninteresting: they will not support induction of general rules, allow predictions about or explanations of the everyday world, or even reveal conceptual "analytic" truths. The arguments from contemporary epistemology earlier show that reasoning deductively about the world would not yield conclusions of any interest, despite a long philosophical tradition to the contrary. Therefore, it would seem bizarre, to say the least, to suppose that deduction is central to human thought, even though the results of deduction would never be useful.

Second, it is clear that people *do* predict, explain, and find regularities in the scientific and everyday worlds and use this knowledge as a basis for decisions and action. These abilities are of fundamental cognitive significance. So even if one were, in desperation, to maintain that much reasoning is perversely based on deduction (and reaches no useful conclusions about the everyday world), it would still be necessary to grant that reasoning that does lead to substantial and interesting conclusions about the everyday world is not based on deduction. But given that these concessions must be granted, deductive reasoning is clearly marginal, rather than central, to cognition, which is the conclusion for which we are arguing.

Third, the attempt to drive a wedge between how people *should* reason and how they *do* reason is beside the point in this context, because the epistemological arguments that we discussed earlier are themselves derived from the actual practice of scientific reasoning. The tendency to make inductive leaps with no deductive justification, the Quine-Duhem thesis, the failure to automatically reject hypotheses when their predictions are disconfirmed, and so on are not merely abstract methodological recommendations. They are manifest in the history of science (e.g., Kuhn 1962). Indeed, in philosophy of science and contemporary epistemology more generally the constraints between accounts of how people *should* and *do* reason are so tight that many philosophers have argued that they cannot be separated (Kornblith 1994; Quine 1969; Thagard 1988).

A final defense might be that people do derive interesting conclusions about the world (e.g., prediction, explanation, and the like), that they use deduction to do this, and that epistemology simply shows that these deductions are not valid. But admitting that people's actual deductions are invalid amounts to giving up deduction as a *computational-level explanation* of human reasoning. This is because, by assumption, people's actual deductions are not valid, and hence their reasoning behavior cannot be characterized by deductive logic. Moreover, the degree to which human reasoning about the world is successful is left entirely mysterious on the view that human reasoning is merely bungled deductive logic.

We have seen that the view that deduction characterizes people's reasoning about the world at the computational level is epistemologically unviable. Reasoning about the everyday world *should not be* and *is not* deductively valid. This conclusion is reinforced by considerations from our two other sources of evidence: AI and the psychology of reasoning.

Artificial Intelligence

A cognitive scientific approach to reasoning assumes that reasoning is a kind of computation. Any adequate cognitive theory must therefore be computationally viable. But, we shall argue, this minimal condition is violated by theories of reasoning that use deduction as their computational-level theory. The fundamental reason should already be clear from our discussion of epistemology: that real-world inferences are not deductively valid and must be outside the scope of any theory based on deduction. But AI provides an ideal testing ground for the hypothesis that reasoning is deductive, because it has adopted the practical project of attempting to formalize (fragments of) human knowledge and build computational systems that reason using this knowledge (e.g., Charniak and McDermott 1985). As we now see, research in AI reinforces the conclusion from epistemology that deductive inference has little or no role in reasoning about the everyday world.

First, let us briefly consider inductive reasoning and inference to the best explanation, as they are studied in AI. According to Hempel's (1965) deductive-nomological view of explanation and Popper's (1959) hypothetico-deductive approach to prediction and theory testing, deduction might be expected to play an important role in computational systems that perform such reasoning. Similarly, according to Rips (1994) and Johnson-Laird and Byrne (1991), who argue that deduction plays an important role in almost all cognitive processes, it might be expected that deduction would have an important role to play. But in fact, as might be expected in the light of the epistemological arguments outlined earlier, AI has found no useful role for deduction in inductive and abductive reasoning. Induction, which is studied in the AI and engineering literatures on machine learning (e.g., Michalski, Carbonell, and Mitchell 1983), pattern recognition (e.g., Duda and Hart

1973), and neural networks (e.g., Hertz, Krogh, and Palmer 1991), has no place for deduction. Inference to the best explanation, which we mentioned in the previous section, is known as abduction in AI. AI systems for abductive inference are generally nondeductive in character (Josephson and Josephson 1994). Furthermore, making predictions, where deduction at least seems prima facie to be appropriate, is no more deductive in the domain of commonsense reasoning than we found it to be in science.

Consider, for example, the prediction that if you drop an egg it will break, based on knowledge about eggs, the floor surface, the height from which you drop the egg, and so on. This inference is uncertain: you may catch the egg before it lands, you may have hardened the shell by artificial means, and so on. As an alternative, reconsider our inference that Tweety flies, from the general proposition that birds fly and the knowledge that Tweety is a bird. This inference, too, is uncertain, since Tweety may be a penguin, have an injury, be newborn, have clipped wings, and so on. Different areas of cognitive science, from cognitive psychology and philosophy to AI, give many different labels to this phenomenon. Commonsense inference is "context-sensitive" (Barsalou 1987), "holds only relative to background conditions" (Barwise and Perry 1983), is "defeasible" (Minsky 1977), "admits exceptions" (Holland et al. 1986), "lacks generality" (Goodman 1983), and has categories that are "intention-relative" (Winograd and Flores 1986). Borrowing a term from AI, we shall call inferences that use rules that allow exceptions *default* inferences.

How can deductive logic, the calculus of certainty, be used to model the uncertainty of default inference? An initial suggestion is to deny that prediction really is uncertain; instead, the conclusion follows deductively from the premises and fails only when one or more of the premises do not apply. According to this view, prediction only appears to be uncertain because some of the premises are left unstated. In the case cited earlier, for example, additional premises such that the egg falls unimpeded, no one has artificially tampered with the shell, and so on are required to deduce the conclusion that the egg will break. If those premises are true, so the story goes, the conclusion follows with certainty.

We saw that this approach does not appear to be successful in the context of scientific reasoning, where we noted that because systems under study are open there will always be indefinitely many unexpected factors that can defeat our predictions. Open systems are also the concern of commonsense reasoning about the everyday world, and so here, too, similar problems arise. In formalizing commonsense reasoning, as in formalizing science, it is not possible to restore certainty by including all these possible additional factors as extra premises in the argument. Even if we rule out the possibility that the dropped egg has an artificially hardened shell or that you catch the egg before it lands, there remain possibilities, such as that room is in free fall or is flooded, that the egg is caught by a net, and so on, that defeat the conclusion that the egg will break. Whatever additional premises we add, there are always further additional factors, not ruled out by those premises, that will overturn the conclusion.[8]

It seems that the majority of inferences about the real world, whether commonsense or scientific, are uncertain rather than deductive. In particular, we have seen that prediction in both science and common sense is nondeductive. Recent research in AI, in contrast to the psychology of reasoning, has abandoned the attempt to model commonsense reasoning purely deductively and has recognized the need for a calculus of default reasoning that goes beyond deduction (Ginsberg 1987).

There are three broad approaches to dealing with uncertainty in AI, none of which maintains that commonsense inference is deductively valid:

1. *The logicist approach.* This approach attempts to develop nonmonotonic logics (or related methods) where future premises can overturn conclusions; that is, they sacrifice deductive validity (McCarthy 1980; McDermott 1982; McDermott and Doyle 1980; Reiter 1980, 1985). Within logic, there have been other attempts to extend logical methods so that they handle the uncertain, nondeductive character of inference that concerns the real world, situation theory (Barwise and Perry 1983) being a notable example. We may view these approaches as broadening the notion of deduction, rather than abandoning it. We should not underestimate the magnitude of the change, however. From an epistemological point of view, giving up certainty is, of course, of fundamental significance; from a formal and computational point of view, nonmonotonic reasoning has different properties from standard monotonic reasoning, partly because these systems must continually reevaluate past conclusions in the light of new information, to see if they still follow (Brachman and Levesque 1985; Ginsberg 1987; Harman 1986; Oaksford and Chater 1991). As noted earlier, in this chapter we restrict deduction to monotonic reasoning, as this is the sense used in the psychology of reasoning; according to this usage, these nonmonotonic reasoning schemes are not deductive. Indeed, psychologists of reasoning have generally rejected the use of nonmonotonic logics (Johnson-Laird 1986; Johnson-Laird and Byrne 1991; Rips 1994). Indeed, Johnson-Laird and Byrne (1991) and A. Garnham (1993) suggest that the psychology of reasoning does not need to appeal to nonmonotonic logics to understand how people carry out nonmonotonic reasoning (although see Chater 1993; Chater and Oaksford 1993; Oaksford 1993; and Oaksford and Chater 1991 for critiques of such approaches).
2. *The probabilistic approach.* This approach to uncertainty in human reasoning uses the mathematical theory of uncertainty, probability theory, or related formalisms (Dempster 1967; Pearl 1988; Shafer 1976). This approach reconstructs commonsense inferences as establishing that a conclusion is probable, rather than deductively certain, given a set of premises. In cases in which conditional probabilities are close to 1, then the probabilistic style of reasoning becomes increasingly close to logical inference. Pearl (1988)

exploits this fact in developing a nonmonotonic logic that he justifies as a limiting case of probabilistic inference. This approach has the advantage of dealing with defeasibility naturally without many of the problems inherent in the nonmonotonic logic approach (Oaksford and Chater 1998). Moreover, it provides a set of normative principles against which to assess human rationality. Later on we will show that this approach permits us to reinterpret some important results in the psychology of reasoning that had previously been thought to impugn human rationality. We show that these results conform to the normative principles of probability theory (Oaksford and Chater 1994a).

3. *The proceduralist approach.* This approach involves devising procedures that solve particular inference problems, but without attempting to ground such procedures in any formal theory of inference (McDermott 1987). A fortiori, this approach rejects the reconstruction of commonsense inference as *deductive* inference.

We have seen that none of these three approaches to uncertainty rely on deductively valid inference. Thus, we might expect that cognitive psychologists concerned with computational models of the mind would similarly have rejected deductive inference as the central mechanism for human thinking. Indeed, it appears that cognitive psychology has implicitly recognized this point, although the implications of this fact do not appear to have touched the psychology of reasoning.

Outside the psychology of reasoning, cognitive psychology has used AI models to provide theoretical proposals about how to organize world knowledge to support commonsense reasoning. We now discuss three approaches that have been influential in psychology: semantic networks, schemas, and production systems. All of these approaches contain mechanisms for dealing with default inference. We shall consider how these theories relate to the three-way classification of AI approaches to default reasoning, arguing that they are all examples of the proceduralist approach.

Semantic networks (e.g., Collins and Loftus 1975; Collins and Quillian 1969) use a hierarchical organization of knowledge and associate properties of objects with nodes in the hierarchy. Suppose that the BIRD node is associated with the property of FLYING, but that the PENGUIN node is associated with the property of NOT_FLYING. On encountering Tweety, a specific penguin, the system can infer both that Tweety cannot fly (because Tweety is a penguin) and that Tweety can fly (because Tweety is a bird). To resolve the conflict the system assumes that information lower in the hierarchy (i.e., more specific information) takes precedence. Thus, the system infers that Tweety does not fly. This method involves a particular approach to a certain kind of default rule—rules that apply to most of the objects in a class but not to some subclasses of that class (rather than default rules that intervening external factors may override, such as those we discussed in the section on epistemology). However, as it stands, the approach is very unconstrained. For example, the label NOT_FLYING may apply to all varieties of bird, although FLYING may still attach to the class BIRD, and other bizarre possibilities (see Woods 1975 for related discussion).

Schema theories and production systems use essentially similar mechanisms for dealing with default inference. In schema or frame theories (e.g., Minsky 1977; Schank and Abelson 1977), incoming information fills "slots" in rules that are organized into domain-specific compartments or "schemas." Slots have associated default values, which further information can override. For example, Tweety may fill the BIRD slot in a rule such as IF **hears**(*BIRD*, *bang*), THEN **flees**(*BIRD*) (i.e., if a bird hears a bang it flees). The slot for BIRD will carry the default assumption that birds fly and hence that Tweety can fly. This may lead, for example, to the inference that Tweety will flee by flying away on hearing the bang. But if you know Tweety is a penguin, then this will override the default using much the same mechanism as in the semantic network, and you may infer that Tweety will waddle rather than fly away. Production systems (e.g., Anderson 1983; Newell 1990) encode knowledge in conditional rules much as in the preceding example. Default inferences arise out of conflicts between different conditional rules, for example, between the rule that IF **bird**(x), THEN **flies**(x) and IF **penguin**(x), THEN NOT(**flies**[x]). On encountering Tweety, to which both rules apply, the system resolves the conflict by choosing the most specific rule, as with semantic networks. Production systems also embody a variety of other procedures, in addition to specificity, for resolving conflicting defaults, including use of production *strength*, goodness of match with the antecedent of the conditional, and so on (Anderson 1983).

Semantic networks, schemas, and productions are all procedural approaches in the sense discussed earlier, specifically in regard to their approach to default reasoning. They handle defaults by simple procedural strategies, such as preferring rules whose antecedents are at a lower level in a default hierarchy. Although the nondefault aspects of some of these systems can be formalized using standard monotonic logic (for example, regarding semantic networks see Woods 1975; for schemas see Hayes 1979), the default inferences are simply treated as procedures without any logical justification. For this reason, none of these systems simply implement logical inference. Hence, since almost all knowledge is defeasible, as we have already discussed, the majority of inferences drawn will be nondeductive in character.

It seems that from the perspective of constructing practical AI systems, as well as from the point of view of epistemology, thought about the world (rather than about mathematics) does not appear to be deductive in character. Thus, AI, as well as epistemology, strongly suggests that psychological theories should not place deductive reasoning at the center of human thought. We now turn to a more direct source of evidence for this view from experimental studies in the psychology of reasoning.

The Argument from Psychology

It is symptomatic of the focus on deduction in the psychology of reasoning that the defeasible, nondeductive character of commonsense inference has not been a major focus of research. In other areas of psychology, though, discussion of such inferences is ubiquitous. We have already noted the popular views that cognitive development is akin to scientific theory change, that commonsense inference is equivalent to the problem of confirmation in science, that perception is inference to the best explanation, and that learning from experience, in whatever domain, involves induction. We have also observed that theories of the organization of knowledge and commonsense inference borrowed by psychologists from AI, including semantic networks, schemas, and production systems, were specifically developed to deal with nondeductive default inference. It is easy to add more examples of aspects of psychology where default inference is central: in text comprehension, the wealth of "bridging" and "elaborative" inferences (e.g., Clark 1977; Garrod and Sanford 1977; Kintsch and van Dijk 1978; O'Brien et al. 1988) required to understand a text do not follow deductively from what is said; similarly, pragmatic inferences about speakers' intentions do not follow deductively from what is said but are a special case of inference to the best explanation (e.g., Levinson 1983; Sperber and Wilson 1986). In memory, elaborative inferences, which importantly affect memory performance, are not deductive inferences from what people must recall—rather, they go beyond and elaborate what people hear (e.g., Craik and Lockhart 1972; Craik and Tulving 1975). Research on decision making and probabilistic reasoning has directly focused on nondeductive forms of inference (e.g., Kahneman, Slovic, and Tversky 1982). Moreover, within the psychology of reasoning itself there has been interest in inductive reasoning tasks (e.g., Klahr and Dunbar 1988; Mynatt, Doherty, and Tweeny 1977; Oaksford and Chater 1994b; Wason 1960).

Mainstream psychology of reasoning has, however, concentrated on what are regarded as deductive reasoning tasks. Our discussions of epistemology and AI suggest that deductive reasoning is rather rare in everyday life; we submit that it is also rather rarer in the laboratory than many reasoning theorists have suspected. At a superficial level, this is because subjects do not appear to perform many deductively valid inferences in laboratory tasks and draw many inferences that are not deductively valid (e.g., Evans, Newstead, and Byrne 1993). In the psychology of reasoning, however, this mismatch between human performance and deductive expectations has not discouraged researchers from proposing theories of reasoning that postulate a core deductive component and explaining away the experimental data as performance errors. We argue that there is a deeper reason that reasoning researchers rarely observe deduction in laboratory reasoning tasks. This is because many tasks that experimenters assume tap deductive reasoning are not really deductive tasks at all. Moreover, experimental data on these tasks are better modeled as uncertain, nondeductive reasoning, rather than as error-full deductive reasoning. We consider two examples that have been of central importance in the psychology of reasoning: P. C. Wason's selection task and conditional inference tasks.

WASON'S SELECTION TASK Wason's selection task (1966, 1968) is perhaps the most intensively studied task in the psychology of reasoning. Subjects must assess whether some evidence is relevant to the truth or falsity of a conditional rule of the form *if p then q*, where by convention *p* stands for the antecedent clause of the conditional and *q* for the consequent clause. In the standard abstract version of the task, the rule concerns cards, which have a number on one side and a letter on the other. The rule is *if there is a vowel on one side* (*p*), *then there is an even number on the other side* (*q*). Four cards are placed before the subject, so that just one side is visible; the visible faces show an "A" (*p* card), a "K" (*not-p* card), a "2" (*q* card), and a "7" (*not-q* card). Subjects then select those cards they must turn over to determine whether the rule is true or false. Typical results were: *p* and *q* cards, 46 percent; *p* card only, 33 percent; *p*, *q*, and *not-q* cards, 7 percent; and *p* and *not-q* cards, 4 percent (Johnson-Laird and Wason 1970).

The task subjects confront is analogous to a central problem of experimental science: the problem of which experiment to perform. The scientist has a hypothesis (or a set of hypotheses) that he or she must assess (for the subject, the hypothesis is the conditional rule) and must choose which experiment (card) will be likely to provide data (i.e., what is on the reverse of the card) that bears on the truth of the hypothesis.

In the light of the epistemological arguments we have already considered, it may seem unlikely that this kind of scientific reasoning will be deductive in character. Nonetheless, the psychology of reasoning has viewed the selection task as paradigmatically deductive (e.g., Evans 1982; Evans, Newstead, and Byrne 1993), although a number of authors have argued for a nondeductive conception of the task (Fischhoff and Beyth-Marom 1983; Kirby 1994; Klayman and Ha 1987; Rips 1990).[9]

The assumption that the selection task is deductive in character arises from the fact that psychologists of reasoning have tacitly accepted Popper's hypothetico-deductive philosophy of science. Popper (1959) assumes that evidence can falsify but not confirm scientific theories. Falsification occurs when predictions that follow deductively from the theory do not accord with observation. This leads to a recommendation for the choice of experiments: to only conduct experiments that have the potential to falsify the hypothesis under test.

Applying the falsificationist account to the selection task, the recommendation is that subjects should only turn cards that are potentially logically incompatible with the conditional rule. When viewed in these terms, the selection task has a deductive component, in that the participant must deduce which cards would be logically incompatible with the conditional rule. According to the rendition of the conditional as material implication (which is standard in the propositional and predicate calculi; see Haack 1978), the only observation that is incompatible with the conditional rule *if p then q* is a card with *p* on one side and *not-q* on the other. Hence, subjects should select only cards that could potentially falsify the rule. That is, they should turn the p card, since it might have a *not-q* on the back, and the *not-q* card, since it might have a p on the back.

This pattern of selections is rarely observed in the experimental results outlined earlier.[10] Subjects typically select cards that could *confirm* the rule, that is, the p and q cards. However, according to falsification the choice of the q card is irrational and is an example of so-called confirmation bias (Evans and Lynch 1973; Wason and Johnson-Laird 1972). The rejection of confirmation as a rational strategy follows directly from the falsificationist perspective.

We have argued that the usual standard of "correctness" in the selection task follows from Popper's hypothetico-deductive view of science. Rejecting the falsificationist picture would eliminate the role of logic and hence deduction in the selection task. In the preceding section on epistemology, we have already seen that the hypothetico-deductive view faces considerable difficulties as a theory of scientific reasoning. This suggests that psychologists should explore alternative views of scientific inference that may provide different normative accounts of experiment choice and hence might lead to a different "correct" answer in the selection task. Perhaps the dictates of an alternative theory might more closely model human performance and hence be consistent with the possibility of human rationality.

We (Oaksford and Chater 1994a) adopt this approach, adapting the Bayesian approach to philosophy of science (Earman 1992; Horwich 1982; Howson and Urbach 1989), rather than the hypothetico-deductive view, to provide a rational analysis (Anderson 1990, 1991) of the selection task. We view the selection task in probabilistic terms, as a problem of Bayesian optimal data selection (Good 1966; Lindley 1956; MacKay 1992). Suppose that you are interested in the hypothesis that eating tripe makes people feel sick. Should known tripe eaters or tripe avoiders be asked whether they feel sick? Should people known to be or not to be sick be asked whether they have eaten tripe? This case is analogous to the selection task. Logically, you can write the hypothesis as a conditional sentence: if you eat tripe (p), then you feel sick (q). The groups of people that you may investigate then correspond to the various visible card options, p, *not-p*, q, and *not-q*. In practice, who is available will influence decisions about which people you question. The selection task abstracts away from this factor by presenting one example of each potential source of data. In terms of our everyday example, it is like coming across four people, one known tripe eater, one known not to have eaten tripe, one known to feel sick, and one known not to feel sick. The task is to decide who to question about how they feel or what they have eaten.

We (Oaksford and Chater 1994a) suggest that hypothesis testers should choose experiments (select cards) to provide the greatest possible "expected information gain" in deciding between two hypotheses: (1) that the task rule, if p then q, is true, that is, *ps* are invariably associated with *qs*, and (2) that the occurrences, of *ps* and *qs* are independent. For each hypothesis, we (1994a) define a probability model that derives from the prior probability of each hypothesis (which for most purposes we assume to be equally likely, i.e., both = .5) and the probabilities of p and of q in the task rule. We define information gain as the difference between the uncertainty *before* receiving some data and the uncertainty *after* receiving that data where we measure uncertainty using Shannon-Wiener information. Thus, we define the information gain of data D as:

$$\text{Information before receiving } D\text{: } I(H_i) = -\sum_{i=1}^{n} P(H_i)\log_2 P(H_i)$$

$$\text{Information after receiving } D\text{: } I(H_i \mid D) = -\sum_{i=1}^{n} P(H_i \mid D)\log_2 P(H_i \mid D)$$

$$\text{Information gain: } I_g = I(H_i) - I(H_i \mid D)$$

We calculate the $P(H_i|D)$ terms using Bayes's theorem. Thus, information gain is the difference between the information contained in the *prior* probability of a hypothesis (H_i) and the information contained in the *posterior* probability of that hypothesis given some data D.[11]

When choosing which experiment to conduct (that is, which card to turn), the subjects do not know what that data will be (that is, what will be on the back of the card). So they cannot calculate actual information gain. However, subjects can compute *expected* information gain. Expected information gain is calculated with respect to all possible data outcomes, for example, for the p card: q and *not-q*, and both hypotheses.

We (1994a) calculated the expected information gain of each card assuming that the properties described in p and q are rare. J. Klayman and Y. Ha (1987) make a similar assumption in accounting for related data on Wason's (1960) 2-4-6 task. The order in expected information gain is:

$$E(Ig[p]) > E(Ig[q]) > E(Ig[not\text{-}q]) > E(Ig[not\text{-}p])$$

This corresponds to the observed frequency of card selections in Wason's task: $p > q >$ *not-q* $>$ *not-p* and thus explains the predominance of p and q card selections as a rational inductive strategy. We (1994a) also show how our model generalizes to all the main patterns of results in the selection task (for discussions of this account see Almor and Sloman 1996; Evans and Over 1996; and Laming 1996 and for a response see Oaksford and Chater 1996). Specifically, it accounts for the nonindependence of card selections (Pollard 1985), the negations paradigm (e.g., Evans and Lynch 1973), the therapy experiments (e.g., Wason 1969), the reduced array selection task (Johnson-Laird and Wason 1970), work on so-called fictional outcomes (Kirby 1994), and deontic versions of the selection task (e.g., Cheng and Holyoak 1985), including perspective and rule-type manipulations (e.g., Cosmides 1989; Gigerenzer and Hug 1992), the manipulation of probabilities and utilities in deontic tasks (Kirby 1994), and effects of relevance (Oaksford and Chater 1995a; Sperber, Cara, and Girotto 1995).

We noted earlier that the philosophy of science that underlies the "deductive" conception of the selection task now has few adherents. The consensus is that scientific theories do not deductively imply predictions and hence that the general problem of choosing which experiment to perform (or, analogously, which card to turn in the selection task) cannot be reconstructed deductively. Further, our (Oaksford and Chater 1994a) probabilistic and hence nonmonotonic account provides a better model of human performance on the selection task. According to this model, people do not use deduction when solving the selection task—rather, they use a nonmonotonic inferential strategy. Crucially, this model preserves human rationality: the optimal data selection model shows that under certain minimal assumptions (the rarity assumption: real-world properties only apply to small subsets of all the objects in the world) people's behavior on the selection task conforms to the principles of a normative theory.

CONDITIONAL INFERENCE TASKS The selection task is perhaps the most celebrated "deductive" reasoning task. However, the conditional inference task is perhaps the task that seems most unequivocally to engage deductive reasoning processes. For example, Rips (1994) uses example (3) as the paradigm example of deductive inference in introducing his mental logic theory of reasoning. Therefore, if human reasoning is not deductive even in this task, then it seems unlikely that other areas of human reasoning will be well explained in deductive terms. For this reason, the conditional reasoning task is a particularly crucial test case for theories of reasoning that employ deductive logic as a computational-level theory.

In the standard conditional inference task, subjects see a conditional rule, *if p then q*, and an additional premise (*p*, *q*, *not-p*, or *not-q*) and are asked whether a given conclusion (again, *p*, *q*, *not-p*, or *not-q*) follows. Consider the simplest form of the task, where the premises are *p* and *if p then q* and subjects decide whether *q* follows. This appears to be an example of the paradigmatic deductive inference of *modus ponens*. Rips's (1994) central example of deductive inference appears to have this form:

(3) If Calvin deposits 50 cents, he'll get a Coke.
Calvin deposits 50 cents.
Therefore, Calvin will get a Coke.

As we have argued earlier, interpreting this natural-language argument involves applying a standard logical analysis, which presupposes that it should be viewed in deductive terms. But in the light of our discussion of epistemology and AI, this inference seems to be a typical example of default inference and not an instance of the deductive reasoning at all, despite Rips. Calvin won't get the Coke if the machine is broken, if the Cokes have run out, if the power is turned off, and so on. That is, additional premises can overturn the conclusion, which monotonic deductive inference does not allow. Thus, although the task is *intended* as a test of deductive reasoning, the subject may be more likely to *interpret* the reasoning materials so that it involves nonmonotonic, nondeductive reasoning.

The question for the psychology of reasoning, then, is which account of how people interpret and reason with the materials in the task provides the best fit with reasoning performance. It turns out that the experimental data support the claim that people treat such inferences as defeasible rather than

deductive. Work on conditional inference indicates that subjects interpret conditional sentences as default rules (Holyoak and Spellman 1993) even in laboratory tasks (Oaksford, Chater, and Stenning 1990). Byrne (1989) and Cummins, T. Lubart, O. Alksnis, and R. Rist (1991) have shown that background information derived from stored world knowledge can affect inferential performance (see also Markovits 1984, 1985). Specifically, they showed that *additional antecedents* influence the inferences conditional statements allow. For example:

(4) If you turn the key the car starts.
(5) *Additional Antecedent*: You are out of gas.

Example (4) could be used to predict that the car will start if you turn the key. This is an inference by *modus ponens*. However, including information about an additional antecedent, example (5), *defeats* this inference (Byrne 1989). Moreover, confidence reduces in this inference for rules that possess many alternative antecedents even when this information is only implicit (Cummins et al. 1991). Additional antecedents also affect inferences by *modus tollens*. If the car does not start, you can infer that you didn't turn the key, unless you are out of gas. Explicitly providing information about alternative antecedents undermines the use of *modus tollens* (Byrne 1989) and reduces confidence in rules that possess many alternative antecedents even when this information is only implicit (Cummins et al. 1991). This result was very striking and unexpected within the context of the psychology of reasoning. However, from the point of view of epistemology and AI it is just what we would expect. Human inferences about Coke machines, like those about the rest of the external world, are defeasible.

In sum, the experimental data seem to show that people treat conditionals in laboratory reasoning tasks as default rules. So it seems that even the commonsense inferences that some reasoning researchers regard as paradigmatic examples of deduction, like example (3), are not examples of deductive inference at all. If defeasibility infects even such paradigmatic cases of deductive reasoning, then it threatens to leave the advocate of deductive reasoning with no commonsense reasoning at all to explain.

How can theories in the psychology of reasoning that place deduction at center stage in human cognition attempt to account for the ubiquity of defeasibility? A popular view is that defeasibility is illusory. The idea is that encountering a fresh premise defeats one of the existing premises and it is this that now allows the rejection of the conclusion.

For example, G. Politzer and M. D. S. Braine (1991)[12] and D. P. O'Brien (1993) argue that the observed effects on reasoning of additional premises do not show that the major premise is defeasible (thereby invoking some nonmonotonic inference regime) but simply show that it is false, according to the standard, nondefeasible, interpretation of the conditional. However, if this is how people interpret conditionals, then the only conditionals that people should believe are true will be those that never admit of counterexamples. Because any commonsense conditional, including example (4), admits exceptions, then all such conditionals will be false. Clearly, people do not reject such conditionals out of hand but freely assert them, argue about whether they are true, and use them to guide their behavior. This makes perfect sense if people interpret conditionals as default rules; it makes no sense at all if they interpret these conditionals logically.

A related line of argument is that people assume that the conditional premise is true until additional premises force them to question it. Before encountering such premises, they reason by *modus ponens* but retract this inference when additional information casts doubt on this premise (e.g., Garnham 1993; Johnson-Laird and Byrne 1991; Politzer and Braine 1991; Stevenson and Over 1995; see Chater 1993; Chater and Oaksford 1993; and Oaksford 1993 for discussion). The idea is that subjects start off assuming the truth of the conditional *if you turn the key the car starts*. If told that *you turn the key*, they infer by *modus ponens* that *the car starts*. However, on encountering the additional premise that *you are out of gas*, subjects question the conditional—this is because general knowledge tells them that the conditional does not apply to cars that are out of gas.

The proposal is that people treat conditional rules as rigid, rather than defeasible, and that people use them deductively. General knowledge acts as a "shield" for cases of apparent defeasible reasoning; general knowledge specifies that the rule does not apply in certain contexts (e.g., when the car is out of gas). The superficial plausibility of this story evaporates once we consider *when* general knowledge operates in the process of inference (see Chater and Oaksford 1993 and Oaksford and Chater 1991, 1995b, for detailed discussion of these issues). Suppose that it operates post hoc; that is, *after* you are presented with a particular problem: you apply *modus ponens* as usual; your general knowledge then determines whether to accept the conclusion; if not, you reject the conclusion, shielding the conditional inference from counterexample by deciding that it was not applicable in that instance. This story is viable only by virtue of being completely vacuous. The inferential processes are exactly those that would occur if the conditional were a default rule, except that every time you find a counterexample you put it aside by assuming that the inference did not apply. In a similar vein, one might maintain that all birds fly and cope with any apparent counterexamples by insisting post hoc that these are not birds. Such a strategy is clearly absurd: it only holds the counterexamples at bay by removing any empirical content from the generalization—it would be held true whatever the state of the world.

Suppose, however, that people adopt a strategy that prespecifies general knowledge of the conditions under which the rule applies. This means that the rule that the subject believes to be true is not that *if you turn the key, the car starts* but that *if you turn the key and there is gas in the car, the car starts*. This rule clearly does not apply to the case where there is no gas in the car. However, in the light of our discussion of epistemology and AI, it should be clear that prespecifying the conditions under which a default rule is true is an endless and impossible task. However many additional conditions you add, further uneliminated counterexamples are always available.[13]

What of apparent demonstrations of logically reasoning tasks? (Braine 1978; Braine and O'Brien 1991; Henle 1962; Lea et al. 1990; O'Brien, Braine, and Yang 1994; Pollitzer and Braine 1991; Rips 1983, 1994). If subjects show correct logical inference on some tasks, then surely we cannot dismiss logic as having a role in human thought? All such demonstrations, however, involve particular inference rules, such as *modus ponens* or *and*-elimination (Braine, Reiser, and Rumain 1984). Such demonstrations bear on the question of whether people are logical but are in themselves far from conclusive. The point of providing a logic is to specify a whole system of inference determined by the logical terms of a language. Logicality can only be determined by conformity to all the patterns of inference licensed by a logic. Isolated conformity to one or another rule does not mean that behavior is logical, because other explanations are equally as plausible and usually more parsimonious than attributing people with the inferential power of deductive logic. For example, isolated logical performance is explicable by procedural accounts that do not attempt to provide any rational basis for cognition. Consequently, in the face of the widespread illogicality observed in reasoning tasks, that behavior conforms to one or two logical rules provides little evidence that people are logical.

Supporters of mental logic get around people's apparent illogicality by arguing that people do not possess various rules or that some inferences are more complex than others. Such attempts to square logic with people's behavior on laboratory tasks cannot, however, resolve the central problem that scientific and commonsense inference is invariably nonlogical. Hence, theories based on logic and derived to explain behavior in laboratory tasks cannot generalize to real inference in the everyday world (Oaksford and Chater 1993, 1995b). Further, as we have argued, even in the laboratory it is clear that conditionals are regarded not as rigid logical rules but as default rules (Cummins et al. 1991; Holyoak and Spellman 1993). Ultimately, however, within the psychology of reasoning, the issue will only be decided by which theories provide the best fits to the empirical data. In this respect nonlogical, probabilistic accounts seem to do better than logical accounts. For example, as we indicated earlier, our account of the selection task (Oaksford and Chater 1994a) provides good fits to most of the existing data.

In short, psychologists of reasoning must conclude, along with philosophers and workers in AI, that human conditional reasoning (outside mathematical domains) is defeasible and not deductive.

Deduction at the Algorithmic Level?

We have argued that deduction cannot provide a computational-level theory of commonsense inference, as has been widely assumed (e.g., Johnson-Laird and Byrne 1991; Macnamara 1986; Rips 1994). But this leaves open the possibility that deduction may be important at the algorithmic, but not the computational, level. We now argue that this position, too, is not viable.

First, note that the principal reason that psychologists (Inhelder and Piaget 1958; Rips 1994) have postulated deduction at the algorithmic level is because it explains how people can reason deductively. But if people cannot reason deductively, then there is no need to postulate a deductive machinery (e.g., a mental logic) to explain reasoning performance.

Second, it remains conceivable in principle that logical procedures at the algorithmic level are used to implement nonlogical reasoning, in the same way as a computer program to perform arithmetic, manipulate lists, or carry out probabilistic calculations might be written in the logic programming language PROLOG. But it seems unlikely that the cognitive system uses algorithms based on logic, if only because it is so computationally expensive to convert procedures into steps of logical inference. Indeed, in practice, programs written in PROLOG run slowly, and moreover, practical programming requires extensive use of nonlogical tricks to reduce the complexity of the program (such as PROLOG's "cut," Clocksin and Mellish 1984). More generally, the "purer" the logical programming style used (i.e., the more nonlogical tricks are avoided), the more rapidly computational complexity grows. Indeed, results from computational complexity theory show that computational systems based solely on logic must be computationally intractable, even when the logical language is just the propositional calculus (Garey and Johnson 1979; for implications for the psychology of reasoning, see Oaksford and Chater 1991, 1993, 1995b and Chater and Oaksford 1993). Moreover, these computational complexity results apply to serial and parallel computational systems and hence must apply to the cognitive system (see Oaksford and Chater 1991).

In sum, then, it seems unlikely that deduction has a significant role in reasoning at either the computational or the algorithmic level.

6. Conclusions

We began by considering the view that deductive reasoning is central to human cognition. However, we have found that the scope of deductive reasoning is remarkably small: it includes neither scientific reasoning, commonsense reasoning, nor important paradigmatic laboratory "deductive" reasoning tasks. It is nondeductive, uncertain reasoning that appears to be cognitively ubiquitous. What does this imply for the psychology of reasoning?

The mental logic account of human reasoning (Braine 1978; Rips 1983, 1994), which advocates deduction at both computational and algorithmic levels of explanation, is directly under threat from this line of argument, for it seems that the deductive inferences for which the mental logic purports to account are actually extremely rare. According to the arguments we have presented, the vast majority of the examples of reasoning that mental logicians reconstruct deductively are not, in fact, deductive inferences at all.

Equally under threat is the mental model view of human inference (Johnson-Laird 1983; Johnson-Laird and Byrne 1991), for which deduction is part of the computational level of explanation. At a general level, this is because the apparatus of mental model theory simply provides a specific way

of drawing deductive, logical inferences: ". . . the [mental] model theory is in no way incompatible with logic: it merely gives up the formal approach (rules of inference) for a semantic approach (search for counter-examples)" (Johnson-Laird and Byrne 1991, p. 212). Since the mental model view relies on the search for counterexamples, it cannot extend beyond monotonic deductive inference. In nonmonotonic inference, it is, by definition, possible for the premises to be true but the conclusions false: that is, counterexamples will always be possible. Any approach based on search for counterexamples will reject all nonmonotonic inferences as invalid (Chater and Oaksford 1993).

If human reasoning is not deductive, how can it be modeled? Epistemology may provide valuable qualitative constraints on patterns of uncertain human reasoning, which have been subject to particularly intensive study in the philosophy of science. Furthermore, each of the three approaches to uncertain reasoning in AI provides possible psychological mechanisms. The logicist wants to somehow extend logical methods to handle uncertain reasoning, the probabilistic account uses probability theory and related formalisms as a starting point, and the procedural approach abandons the search for a rational underpinning for thought and proposes heuristics for particular types of uncertain reasoning.

To what extent are these approaches being taken up within the psychology of reasoning? As we have noted, theorists who argue for mental logics have been surprisingly unwilling to propose nonmonotonic logics as psychological mechanisms but have attempted to maintain that nonmonotonic logics are (at least in a large range of cases) unnecessary (Johnson-Laird 1986; Johnson-Laird and Byrne 1991; Rips 1994). The proceduralist approach has also been relatively little investigated, although J. St. B. T. Evans (1982, 1983, 1984, 1989) can be viewed as taking up this approach, in that he focuses on processing-based explanations of reasoning data. Nonetheless, Evans is not a full-blown proceduralist. He views his procedures only as *supplements* to accounts of deductive competence, to explain systematic errors and biases in people's deductive reasoning (1991). J. H. Holland, K. J. Holyoak, R. E. Nisbett, and P. R. Thagard (1986) provide a more thoroughgoing proceduralist account of default reasoning, which, however, they have not directly applied to reasoning research. They view knowledge as organized into hierarchies of default rules in a production system. Although such systems seem directly applicable to commonsense human reasoning, no detailed modeling of experimental reasoning data has so far been carried out. A number of researchers have taken up the probabilistic approach (Cheng and Novick 1992; Fischhoff and Beyth-Marom 1983; Gigerenzer and Hug 1992; Kirby 1994; Klayman and Ha 1987; Manktelow and Over 1991), including our work on the selection task described earlier (Oaksford and Chater 1994a). We view this approach as the most promising because it successfully explains some of the most problematic data on human reasoning as conforming to the principles of a normative theory; that is, it preserves the rationality of people's reasoning and thus explains why cognition is successful. Nonetheless, only further research within each of these three approaches will provide an answer to the question of which is the most appropriate framework for studying human reasoning.

NOTES

1. They argue that logic must be supplemented with additional principles that constrain which logical inferences people actually draw (Johnson-Laird and Byrne 1991).

2. There is ambiguity over whether the algorithmic level postulated by mental models is really a kind of logical proof theory, based on semantic principles (like truth tables or semantic tableaux). For example, this viewpoint seems implicit in Johnson-Laird and Byrne (1991, p. 212): "The [mental] model theory is in no way incompatible with logic: it merely gives up the formal (rules of inference) for a semantic approach (search for counterexamples)."

3. However, advocates of mental logic, for example, do explicitly argue that deductive reasoning has an important role in many nondeductive reasoning tasks, such as induction and decision making (e.g., Rips 1994).

4. Although J. H. Holland, K. J. Holyoak, R. E. Nisbett, and P. R. Thagard's (1986) assumption that knowledge is organized as hierarchies of default rules does implicitly involve a general rejection of the idea that much of cognition is deductive in character.

5. This view is explicitly taken in formal semantics, where the goal of logical analysis is typically to capture the inferences that people draw from a statement (e.g., Cresswell 1985; Kamp and Reyle 1993; Montague 1974). The set of these entailments is closely related to or even identical with its meaning.

6. Of course, many scientific theories make only probabilistic predictions, and these predictions are not readily modeled in deductive terms. It is worth noting that many predictions of commonsense thinking often appear to have this character: I predict that if I am late to the bus stop, I will miss the bus; but this prediction is only probabilistic, since the bus is late, say, 5 percent of the time. Although probabilistic cases were of considerable concern to Popper (1959) and also to Hempel (1965), who dubbed them "inductive-statistical" inference, we leave this complication aside here.

7. Indeed, even the deductive character of mathematics is under threat from this point of view (Lakatos 1976).

8. The open character of the everyday world is often dealt with by fiat in AI—by simply making a "closed-world" assumption (Clark 1978) that the current contents of the system's database include all relevant factors. Even within the closed world, however, researchers must assume that inference is nonmonotonic, and hence nondeductive, which is all we are concerned with in this chapter. As it happens, the open character of the commonsense world means that many inferences that follow from the closed-world assumption are clearly invalid from outside the perspective of that closed world.

9. J. St. B. T. Evans and D. Over (1996) suggest that the selection task is a "decision making task," rather than a deductive reasoning task—a decision must be made concerning which cards to choose. But it appears that, at this level of abstraction, almost any psychological task involves "decision making," in the trivial sense that the subject must decide how to respond. It is therefore not clear whether Evans and Over's viewpoint amounts to accepting that reasoning in the task involves uncertain inductive inference, as we argue.

10. As few as 4 percent of subjects ... to the deductive view, the bulk of the experimental data must be explained in terms of performance error.

11. In response to Evans and Over's (1996) observation that information gains may sometimes be negative and that this is intuitively unattractive, we (Oaksford and Chater 1996) develop an alternative motivation for their account, based on expected Kullback-Liebler distance. This revised account is mathematically identical to the current account but is somewhat more complex to derive, so we have retained our original account here. We also note here that the equations here have the opposite sign to those used in Oaksford and Chater (1994a): that is, information gains are all positive rather than negative.

12. See Oaksford and Chater 1995b for a formal demonstration of the fallacy in Politzer and Braine's (1991) argument.

13. Such shielding of rules would seem to preclude any role for counterexamples to refute rules, and hence inferences like reductio ad absurdum would no longer be possible. We don't believe this to be a problem because inferences like reductio can apply but now become more a matter of degree. The situation is analogous to I. Lakatos's (1970) account of the "protective belt" that surrounds a theory, protecting it from simple falsification by counterexample. Although the theory will survive many counterexamples by invoking auxiliary assumptions in its protective belt, there is a point at which this is no longer viable and at which counterexamples penetrate to falsify the theory.

REFERENCES

Almor, A., and S. A. Sloman (1996). Is deontic reasoning special? *Psychological Review* 103:374–380.

Anderson, J. R. (1983). *The Architecture of Cognition*. Cambridge: Harvard University Press.

Anderson, J. R. (1990). *The Adaptive Character of Thought*. Hillsdale, NJ: Erlbaum.

Anderson, J. R. (1991). Is human cognition adaptive? *Behavioral and Brain Sciences* 14:471–517.

Baron, J. (1994). *Thinking and Deciding*. Cambridge: Cambridge University Press.

Barsalou, L. W. (1987). The instability of graded structures: Implications for the nature of concepts. In U. Neisser (ed.), *Concepts and Conceptual Development*, pp. 101–140. Cambridge: Cambridge University Press.

Barwise, J., and R. Cooper (1981). Generalized quantifiers and natural languages. *Linguistics and Philosophy* 4:159–219.

Barwise, J., and J. Perry (1983). *Situation and Attitudes*. Cambridge: MIT Press.

Boolos, G. S., and R. C. Jeffrey (1980). *Computability and Logic*. Cambridge: Cambridge University Press.

Brachman, R. J., and H. Levesque (eds.) (1985). *Readings in Knowledge Representation*. Los Altos, CA: Morgan Kaufmann.

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

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

Braine, M. D. S., B. J. Reiser, and B. Rumain (1984). Some empirical justification for a theory of natural propositional logic. In G. Bower (ed.), *The Psychology of Learning and Motivation: Advances in Research and Theory*, vol. 18, pp. 313–370. New York: Academic Press.

Burnyeat, M. (1983). *The Skeptical Tradition*. Berkeley: University of California Press.
Byrne, R. M. J. (1989). Suppressing valid inferences with conditionals. *Cognition* 31:1–21.
Carey, S. (1988). Conceptual differences between children and adults. *Mind and Language* 3:167–181.
Charniak, E., and D. McDermott (1985). *An Introduction to Artificial Intelligence*. Reading, MA: Addison-Wesley.
Chater, N. (1993). Mental models and non-monotonic reasoning. *Behavioral and Brain Sciences* 16:340–341.
Chater, N., and M. Oaksford (1993). Logicism, mental models and everyday reasoning. *Mind and Language* 8:72–89.
Cheng, P. W., and K. J. Holyoak (1985). Pragmatic reasoning schemas. *Cognitive Psychology* 17:391–416.
Cheng, P. W., and L. R. Novick (1992). Covariation in natural causal induction. *Psychological Review* 99:365–382.
Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge: MIT Press.
Clark, H. H. (1977). Bridging. In P. N. Johnson-Laird and P. C. Wason (eds.), *Thinking: Readings in Cognitive Science*, pp. 411–420. Cambridge: Cambridge University Press.
Clark, K. L. (1978). Negation as failure. In H. Gallaire and J. Minker (eds.), *Logic and Data Bases*, pp. 293–322. New York: Plenum Press.
Clocksin, W. F., and C. S. Mellish (1984). *Programming in PROLOG*. Berlin: Springer-Verlag.
Collins, A. M., and E. F. Loftus (1975). A spreading activation theory of semantic processing. *Psychological Review* 82:407–428.
Collins, A. M., and M. R. Quillian (1969). Retrieval from semantic memory. *Journal of Verbal Learning and Verbal Behavior* 8:240–247.
Coolidge, J. L. (1940). *A History of Geometrical Methods*. Oxford: Clarendon Press.
Cosmides, L. (1989). The logic of social exchange: Has natural selection shaped how humans reason? Studies with the Wason selection task. *Cognition* 31: 187–276.
Craik, F. I. M., and R. S. Lockhart (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior* 11:671–684.
Craik, F. I. M., and E. Tulving (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General* 104:268–294.
Cresswell, M. J. (1985). *Structured Meanings*. Cambridge: MIT Press.
Cummins, D. D., T. Lubart, O. Alksnis, and R. Rist (1991). Conditional reasoning and causation. *Memory and Cognition* 19:274–282.
Curry, H. B. (1956). *An Introduction to Mathematical Logic*. Amsterdam: Van Nostrand.
Davidson, D. (1984). *Inquiries into Truth and Interpretation*. Oxford: Clarendon Press.
Dempster, A. P. (1967). Upper and lower probabilities induced by a multivalued mapping. *Annals of Mathematical Statistics* 38:325–339.
Duda, R. O., and P. E. Hart (1973). *Pattern Classification and Scene Analysis*. New York: Wiley.
Duhem, P. ([1914] 1954). *The Aim and Structure of Physical Theory*. Princeton: Princeton University Press.
Earman, J. (1992). *Bayes or Bust?* Cambridge: MIT Press.
Evans, J. St. B. T. (1982). *The Psychology of Deductive Reasoning*. London: Routledge and Kegan Paul.
Evans, J. St. B. T. (1983). Linguistic determinants of bias in conditional reasoning. *Quarterly Journal of Experimental Psychology* 35A:635–644.
Evans, J. St. B. T. (1984). Heuristic and analytic processes in reasoning. *British Journal of Psychology* 75:451–468.
Evans, J. St. B. T. (1989). *Bias in Human Reasoning: Causes and Consequences*. Hillsdale, NJ: Erlbaum.
Evans, J. St. B. T. (1991). Theories of human reasoning: The fragmented state of the art. *Theory and Psychology* 1:83–106.
Evans, J. St. B. T., and J. S. Lynch (1973). Matching bias in the selection task. *British Journal of Psychology* 64:391–397.
Evans, J. St. B. T., S. E. Newstead, and R. M. J. Byrne (1993). *Human Reasoning: The Psychology of Deduction*. Hillsdale, NJ: Erlbaum.
Evans, J. St. B. T., and D. Over (1996). Rationality in the selection task: Epistemic utility vs. uncertainty reduction. *Psychological Review* 103:356–363.
Fischhoff, B., and R. Beyth-Marom (1983). Hypothesis evaluation from a Bayesian perspective. *Psychological Review* 90:239–260.
Fodor, J. A. (1983). *Modularity of Mind: An Essay on Faculty Psychology*. Cambridge: MIT Press.
Fodor, J. A., and Z. W. Pylyshyn (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition* 28:3–71.
Garey, M., and M. Johnson (1979). *Computers and Tractability*. San Francisco: W. H. Freeman.
Garnham, A. (1993). Is logicist cognitive science possible? *Mind and Language* 8:49–71.
Garrod, S., and A. J. Sanford (1977). Interpreting anaphoric relations: The integration of semantic information while reading. *Journal of Verbal Learning and Verbal Behavior* 16:77–90.
Gigerenzer, G., and K. Hug (1992). Domain-specific reasoning: Social contracts, cheating and perspective change. *Cognition* 43:127–171.
Ginsberg, M. L. (ed.), (1987). *Readings in Nonmonotonic Reasoning*. Los Altos, CA: Morgan Kaufmann.
Girard, J.-Y. (1989). *Proofs and Types*. Cambridge: Cambridge University Press.
Gluck, M. A., and G. H. Bower (1988). Evaluating an adaptive network model of human learning. *Memory and Language* 27:166–195.
Glymour, C. (1980). *Theory and Evidence*. Princeton: Princeton University Press.
Goldman, A. I. (1986). *Epistemology and Cognition*. Cambridge: Harvard University Press.
Good, I. J. (1966). A derivation of the probabilistic explication of information. *Journal of the Royal Statistical Society, Series B*, 28:578–581.
Goodman, N. (1951). *The Structure of Appearance*. Cambridge: Harvard University Press.
Goodman, N. ([1954] 1983). *Fact, Fiction, and Forecast*, 4th edition. Cambridge: Harvard University Press.
Gorman, M. E., and M. E. Gorman (1984). A comparison of disconfirmatory, confirmatory and control strategies on Wason's 2-4-6 task. *Quarterly Journal of Experimental Psychology* 36A:629–648.
Haack, S. (1978). *Philosophy of Logics*. Cambridge: Cambridge University Press.
Harman, G. (1965). The inference to the best explanation. *Philosophical Review* 74:88–95.

Harman, G. (1986). *Change in View: Principles of Reasoning.* Cambridge: MIT Press Bradford.

Hayes, P. (1979). The logic of frames. In D. Metzing (ed.), *Frame Conceptions in Text Understanding*, pp. 46–61. Berlin: Walter de Gruyter.

Hempel, C. (1965). *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*. New York: Free Press.

Henle, M. (1962). The relation between logic and thinking. *Psychological Review* 69:366–378.

Hertz, J., A. Krogh, and R. G. Palmer (1991). *Introduction to the Theory of Neural Computation.* Menlo Park, CA: Addison-Wesley.

Hodges, W. (1977). *Logic*. Harmondsworth, UK: Penguin.

Holland, J. H., K. J. Holyoak, R. E. Nisbett, and P. R. Thagard (1986). *Induction: Processes of Inference, Learning and Discovery*. Cambridge: MIT Press.

Holyoak, K. J., and B. A. Spellman (1993). Thinking. *Annual Review of Psychology* 44:265–315.

Horn, L. R. (1989). *A Natural History of Negation*. Chicago: University of Chicago Press.

Horwich, P. (1982). *Probability and Evidence*. Cambridge: Cambridge University Press.

Howson, C., and P. Urbach (1989). *Scientific Reasoning: The Bayesian Approach*. La Salle, IL: Open Court.

Inhelder, B., and J. Piaget (1958). *The Growth of Logical Reasoning*. New York: Basic Books.

James, S. (1975). *Francis Bacon and the Style of Science*. Chicago: University of Chicago Press.

Jaspars, J. M. F., M. R. C. Hewstone, and F. D. Fincham (1983). Attribution theory and research: The state of the art. In J. M. F. Jaspars, F. D. Fincham, and M. R. C. Hewstone (eds.), *Attribution Theory: Essays and Experiments*, pp. 3–36. Orlando, FL: Academic Press.

Johnson-Laird, P. N. (1983). *Mental Models*. Cambridge: Cambridge University Press.

Johnson-Laird, P. N. (1986). Reasoning without logic. In T. Myers, K. Brown, and B. McGonigle (eds.), *Reasoning and Discourse Processes*, pp. 13–50. London: Academic Press.

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

Johnson-Laird, P. N., and P. C. Wason (1970). Insight into a logical relation. *Quarterly Journal of Experimental Psychology* 22:49–61.

Josephson, J. R., and S. G. Josephson (1994). *Abductive Inference: Computation, Philosophy, Technology*. Cambridge: Cambridge University Press.

Kahneman, D., P. Slovic, and A. Tversky (eds.) (1982). *Judgment under Uncertainty: Heuristics and Biases*. Cambridge: Cambridge University Press.

Kamp, H., and U. Reyle (1993). *From Discourse to Logic: Introduction to Model Theoretic Semantics for Natural Language, Formal Logic and Discourse Representation Theory*. Kluwer: Dordrecht.

Karmiloff-Smith, A. (1988). The child is a theoretician, not an inductivist. *Mind and Language* 3:183–196.

Katz, J. J. (1990). *The Metaphysics of Meaning*. Cambridge: MIT Press.

Kelley, H. H. (1967). Attribution theory in social psychology. In D. Levine (ed.), *Nebraska Symposium on Motivation*, vol. 15, pp. 192–238. Lincoln: University of Nebraska Press.

Kintsch, W., and T. A. Van Dijk (1978). Toward a model of text comprehension and production. *Psychological Review* 85:363–394.

Kirby, K. N. (1994). Probabilities and utilities of fictional outcomes in Wason's four-card selection task. *Cognition* 51:1–28.

Klahr, D., and K. Dunbar (1988). Dual space search during scientific reasoning. *Cognitive Science* 12:1–48.

Klayman, J., and Y. Ha (1987). Confirmation, disconfirmation and information in hypothesis testing. *Psychological Review* 94:211–228.

Kornblith, H. (ed.) (1994). *Naturalizing Epistemology*. Cambridge: MIT Press.

Kuhn, T. S. (1962). *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.

Lakatos, I. (1970). Falsification and the methodology of scientific research programmes. In I. Lakatos and A. Musgrave (eds.), *Criticism and the Growth of Knowledge*, pp. 91–196. Cambridge: Cambridge University Press.

Lakatos, I. (1976). *Proofs and Refutations: The Logic of Mathematical Discovery*. Cambridge: Cambridge University Press.

Lakatos, I. (1977a). *Philosophical Papers*, vol. 1: *The Methodology of Scientific Research Programmes*. Cambridge: Cambridge University Press.

Lakatos, I. (1977b). *Philosophical Papers*, vol. 2: *Mathematics, Science and Epistemology*. Cambridge: Cambridge University Press.

Laming, D. (1996). On the analysis of irrational data selection: A critique of Oaksford and Chater (1994). *Psychological Review* 103:364–373.

Lea, R. B., D. P. O'Brien, S. M. Fisch, I. A. Noveck, and M. D. S. Braine (1990). Predicting propositional logic inferences in text comprehension. *Journal of Memory and Language* 29:361–387.

Lehrer, K. (1990). *Theory of Knowledge*. Boulder: Westview.

Lemmon, E. J. (1965). *Beginning Logic*. London: Nelson.

Levinson, S. (1983). *Pragmatics*. Cambridge: Cambridge University Press.

Lindley, D. V. (1956). On a measure of the information provided by an experiment. *Annals of Mathematical Statistics* 27:986–1005.

MacKay, D. J. C. (1992). Information-based objective functions for active data selection. *Neural Computation* 4:590–604.

Macnamara, J. (1986). *A Border Dispute: The Place of Logic in Psychology*. Cambridge: MIT Press.

Macnamara, J., and G. E. Reyes (eds.) (1994). *The Logical Foundations of Cognition*. Oxford: Oxford University Press.

Manktelow, K. I., and D. E. Over (1991). Social roles and utilities in reasoning with deontic conditionals. *Cognition* 39:85–105.

Markovits, H. (1984). Awareness of the "possible" as a mediator of formal thinking in conditional reasoning problems. *British Journal of Psychology* 75:367–376.

Markovits, H. (1985). Incorrect conditional reasoning among adults: Competence or performance. *British Journal of Psychology* 76:241–247.

Marr, D. (1982). *Vision*. San Francisco: W. H. Freeman.

McCarthy, J. M. (1980). Circumscription: A form of nonmonotonic reasoning. *Artificial Intelligence* 13:27–39.

McDermott, D. (1982). Non-monotonic logic II: Non-monotonic modal theories. *Journal of the Association for Computing Machinery* 29(1): 33–57.

McDermott, D. (1987). A critique of pure reason. *Computational Intelligence* 3:151–160.

McDermott, D., and J. Doyle (1980). Non-monotonic logic I. *Artificial Intelligence* 13:41–72.

Michalski, R., J. G. Carbonell, and T. Mitchell (1983). *Machine Learning: An Artificial Intelligence Approach*. Palo Alto, CA: Tioga.
Minsky, M. (1977). Frame system theory. In P. N. Johnson-Laird and P. C. Wason (eds.), *Thinking: Readings in Cognitive Science*, pp. 355–376. Cambridge: Cambridge University Press.
Montague, R. (1974). *Formal Philosophy: Selected Papers of Richard Montague*. R. H. Thomason (ed.). New Haven: Yale University Press.
Mynatt, C. R., M. E. Doherty, and R. D. Tweney (1977). Confirmation bias in a simulated research environment: An experimental study of scientific inference. *Quarterly Journal of Experimental Psychology* 29:85–95.
Newell, A. (1990). *Unified Theories of Cognition*. Cambridge: Harvard University Press.
Oaksford, M. (1993). Mental models and the tractability of everyday reasoning. *Behavioral and Brain Sciences* 16:360–361.
Oaksford, M., and N. Chater (1991). Against logicist cognitive science. *Mind and Language* 6:1–38.
Oaksford, M., and N. Chater (1993). Reasoning theories and bounded rationality. In K. I. Manktelow and D. E. Over (eds.), *Rationality*, pp. 31–60. London: Routledge.
Oaksford, M., and N. Chater (1994a). A rational analysis of the selection task as optimal data selection. *Psychological Review* 101:608–631.
Oaksford, M., and N. Chater (1994b). Another look at eliminative and enumerative behaviour in a conceptual task. *European Journal of Cognitive Psychology* 6:149–169.
Oaksford, M., and N. Chater (1995a). Information gain explains relevance which explains the selection task. *Cognition* 57:97–108.
Oaksford, M., and N. Chater (1995b). Theories of reasoning and the computational explanation of everyday inference. *Thinking and Reasoning* 1:121–152.
Oaksford, M., and N. Chater (1996). Rational explanation of the selection task. *Psychological Review* 103:381–391.
Oaksford, M., and N. Chater (1998). *Rationality in an Uncertain World: Essays on the Cognitive Science of Human Reasoning*. Hove, East Sussex: Psychology Press.
Oaksford, M., N. Chater, and K. Stenning (1990). Connectionism, classical cognitive science and experimental psychology. *AI and Society* 4:73–90.
O'Brien, D. P. (1993). Mental logic and human irrationality. In K. I. Manktelow and D. E. Over (eds.), *Rationality*, pp. 110–135. London: Routledge.
O'Brien, D. P., M. D. S. Braine, and Y. Yang (1994). Propositional reasoning by mental models? Simple to refute in principle and in practice. *Psychological Review* 101:711–724.
O'Brien, E. J., D. M. Shank, J. L. Myers, and K. Rayner (1988). Elaborative inference during reading: Do they occur on line? *Journal of Experimental Psychology: Learning, Memory and Cognition* 14:410–420.
Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems*. San Mateo, CA: Morgan Kaufmann.
Politzer, G., and M. D. S. Braine (1991). Responses to inconsistent premises cannot count as suppression of valid inferences. *Cognition* 38:103–108.
Pollard, P. (1985). Non-independence of selections on the Wason selection task. *Bulletin of the Psychonomic Society* 23:317–320.
Pollock, J. L. (1986). *Contemporary Theories of Knowledge*. Totowa: Rowman Littlefield.
Popper, K. R. ([1935] 1959). *The Logic of Scientific Discovery*. London: Hutchinson.

Putnam, H. (1974). The "corroboration" of theories. In P. A. Schilpp (ed.), *The Philosophy of Karl Popper*, vol. 1, pp. 221–240. La Salle, IL: Open Court.
Putnam, H. ([1962] 1975). The analytic and the synthetic. In H. Putnam (ed.), *Philosophical Papers*, vol. 2: *Mind, Language and Reality*, pp. 33–69. Cambridge: Cambridge University Press.
Putnam, H., and P. Benacerraf (eds.) (1983). *Philosophy of Mathematics: Selected Readings*, 2nd ed. Cambridge: Cambridge University Press.
Quine, W. V. O. (1953). Two dogmas of empiricism. In *From a Logical Point of View*, pp. 20–46. Cambridge: Harvard University Press.
Quine, W. V. O. (1960). *Word and Object*. Cambridge: MIT Press.
Quine, W. V. O. (1969). Epistemology naturalized. In *Ontological Relativity and Other Essays*, pp. 69–90. New York: Columbia University Press.
Quine, W. V. O. (1990). *Pursuit of Truth*. Cambridge: MIT Press.
Reiter, R. (1980). A logic for default reasoning. *Artificial Intelligence* 13:81–132.
Reiter, R. ([1978] 1985). One reasoning by default. In R. J. Brachman and H. Levesque (eds.), *Readings in Knowledge Representation*, pp. 401–410. Los Altos, CA: Morgan Kaufmann.
Rips, L. J. (1983). Cognitive processes in propositional reasoning. *Psychological Review* 90:38–71.
Rips, L. J. (1990). Reasoning. *Annual Review of Psychology* 41:321–353.
Rips, L. J. (1994). *The Psychology of Proof: Deductive Reasoning in Human Thinking*. Cambridge: MIT Press.
Russell, B. (1919). *Introduction to Mathematical Philosophy*. New York: Macmillan.
Russell, B. (1946). *History of Western Philosophy*. London: Macmillan.
Schank, R. C., and R. P. Abelson (1977). *Scripts, Plans, Goals, and Understanding*. Hillsdale, NJ: Erlbaum.
Schiffer, S. (1987). *Remnants of Meaning*. Cambridge: MIT Press.
Scott, D., and C. Strachey (1971). *Toward a Mathematical Semantics for Computer Languages*. Oxford: Oxford University Press.
Scruton, R. (1982). *Kant*. Oxford: Oxford University Press.
Shafer, G. (1976). *A Mathematical Theory of Evidence*. Princeton: Princeton University Press.
Smedslund, J. (1970). On the circular relation between logic and understanding. *Scandinavian Journal of Psychology* 11:217–219.
Sperber, D., F. Cara, and V. Girotto (1995). Relevance explains the selection task. *Cognition* 57:31–95.
Sperber, D., and D. Wilson (1986). *Relevance: Communication and Cognition*. Oxford: Basil Blackwell.
Stein, E. (1996). *Without Good Reason: The Rationality Debate in Philosophy and Cognitive Science*. Oxford: Oxford University Press.
Stevenson, R. J., and D. E. Over (1995). Deduction from uncertain premises. *Quarterly Journal of Experimental Psychology* 48:613–643.
Stich, S. (1985). Could man be an irrational animal? *Synthese* 64:115–135.
Stich, S. (1990). *The Fragmentation of Reason*. Cambridge: MIT Press.
Thagard, P. (1988). *Computational Philosophy of Science*. Cambridge: MIT Press Bradford.
Toulmin, S. (1961). *Foresight and Understanding*. London: Hutchinson.
Tversky, A., and D. Kahneman (1974). Judgment under Uncertainty: Heuristics and biases. *Science* 125:1124–1131.
Tversky, A., and D. Kahneman (1986). Rational choice and the framing of decisions. *Journal of Business* 59:251–278.