DOI: http://dx.doi.org/10.5334/pb-50-3-4-193
Psychologica Belgica
2010, 50-3&4, 193-221.

AGAINST LOGICAL FORM

P.N. JOHNSON-LAIRD
Princeton University, New Jersey, USA

This research was supported in part by a grant from the National Science Foundation SES 0844851 to study deductive and probabilistic reasoning. The author is grateful to Geoffrey Goodwin, Sangeet Khemlani, Niklas Kunze, and Max Lotstein, for their critical readings of an earlier draft, and to Monica Bucciarelli, Ruth Byrne, Vittorio Girotto, Paolo Legrenzi, Juan Madruga, Walter Schaeken, and Walter Schroyens, for help and encouragement over the years. The paper is dedicated to André Vandierendonck, a stalwart investigator of mental models.

Abstract

An old view in logic going back to Aristotle is that an inference is valid in virtue of its logical form. Many psychologists have adopted the same point of view about human reasoning: the first step is to recover the logical form of an inference, and the second step is to apply rules of inference that match these forms in order to prove that the conclusion follows from the premises. The present paper argues against this idea. The logical form of an inference transcends the grammatical forms of the sentences used to express it, because logical form also depends on context. Context is not readily expressed in additional premises. And the recovery of logical form leads ineluctably to the need for infinitely many axioms to capture the logical properties of relations. An alternative theory is that reasoning depends on mental models, and this theory obviates the need to recover logical form.

Introduction

Reasoning in daily life depends on the ability to grasp that a set of propositions implies a conclusion. Nearly everyone grasps, for example, that these premises:

The government subsidized the bank or the bank went broke.

In fact, the bank didn't go broke.

imply that the government subsidized the bank. That is, naïve individuals realize that if the premises are true then the conclusion must be true too.

P.N. Johnson-Laird, Professor P.N. Johnson-Laird, Department of Psychology, Princeton University, New Jersey 08540, USA, E-mail: phil@princeton.edu

The term *naïve* here refers to individuals who have not explicitly mastered logic; it does not impugn their intelligence. Plenty of intelligent people have never learned logic, and yet they reason well (Stanovich, 1999). However, the claim that naïve individuals can make deductions is controversial, because some logicians and some psychologists argue to the contrary (e.g., Oaksford & Chater, 2007). These arguments, however, make it much harder to understand how human beings were able to devise logic and mathematics if they were incapable of deductive reasoning beforehand.

In the mid-1970's when experimental psychologists began to propose theories about the mental processes of reasoning, they converged on a fundamental idea. Its intellectual god-father was Jean Piaget, and the idea was that naïve individuals construct an unconscious logical calculus that enables them to reason. Piaget used this idea to try to explain how infants developed into adults who could master logic and mathematics (e.g., Inhelder & Piaget, 1958). According to more recent proponents of this idea, this calculus contains formal rules of inference, such as:

A or B.
Not B.
Therefore, A.

where the values of the variables *A* and *B* can be any propositions whatsoever. The first step in an inference is accordingly to establish the *logical form* of the premises, and the second step is to match these forms to corresponding formal rules of inference that allow an inference to be made. Further rules may then be applicable until the chain of inferences yields a proof of the required conclusion. For the inference above about the bank, both of these steps are straightforward. The logical form of the premises matches the preceding formal rule. It delivers the conclusion, *A*, to which the reasoning system restores the rightful contents:

The government subsidized the bank

The formal hypothesis about the mental processes of reasoning is longstanding, plausible, influential. Nearly all cognitive scientists, including philosophers (e.g., Pollock, 1989), linguists (e.g., Sperber and Wilson, 1986), artificial intelligencers (e.g., Cherniak, 1986), and students of automated theorem-proving (e.g., Bledsoe, 1977; Wos, 1988) concurred with it at first. Likewise, psychologists saw their job as to find out which sort of logic, and which sort of formalization, human reasoners relied on. They made many proposals about these matters (e.g., Osherson, 1974-6; Johnson-Laird, 1975; Braine, 1978; Rips, 1983, 1994; Macnamara, 1986; Braine & O'Brien, 1998).

The story of psychological studies of reasoning since that happy time can be construed as a move away from this idea of *mental logic*. André Vandierendonck has played an important part in this story. He showed how reasoning about spatial and temporal relations depends, not on formal rules, but on mental models (Vandierendonck & De Vooght, 1996), which can be annotated with symbols to represent indeterminacies (e.g., Vandierendonck, Dierckx, & De Vooght, 2004), and which are affected both by the constraints of working memory (e.g., Vandierendonck & De Vooght, 1997; Duyck, Vandierendonck, & De Vooght, 2003) and by reasoners' strategies (e.g., Dierckx, Vandierendonck, & Pandelaere, 2003). And so my aim in this chapter – as a way of thanking André both for his research and for his kindness to me – is to consider a major weakness in the hypothesis of mental logic, one that has nagged away at me for years. It is the principle that the first step in using mental logic is the recovery of the *logical form* of the premises.

My argument is that logical form, and therefore formal rules of inference, are unlikely to play any significant role in the mental processes underlying reasoning in daily life. There is no reason why they should do so. Logic captures the implications among sentences, usually expressed in a formalized language. In contrast, reasoning is the mental process of drawing conclusions from sets of propositions, usually expressed in natural language. Hence, logic may tell us no more about reasoning than bookkeeping tells us about why people spend their money. Indeed, few logicians these days argue that logic should be a basis for psychological theory (see Harman, 1986). The present article is not a critique of logic. What it does criticize, however, is the doctrine that logic provides the basis for human reasoning. It does not argue that everyday propositions lack a logical form, but merely that the task of recovering it is extraordinarily difficult and probably unnecessary. If this criticism is correct, does it follow that psychologists should dispense with logic? Not at all. There is much that students of reasoning can learn from logic, including which implications among sentences are valid. The mistake is to import logic directly into psychological theory, and to assume that the mental processes of everyday reasoning extract the logical form of premises, and use it to reason. My aim is not wholly negative, however. I want to sharpen up the demarcation of human reasoning, and to defend an alternative theory of reasoning based on mental models.

Perhaps the most immediate sign of the difficulty of a formal approach to reasoning is that arguments in daily life are not laid out like formal proofs. A typical example is Mr. Micawber's famous advice (in Charles Dickens's novel, *David Copperfield*):

> "Annual income twenty pounds, annual expenditure nineteen pounds nineteen and six, result happiness. Annual income twenty pounds, annual expenditure twenty pounds ought and six, result misery."

What is the logical form of this claim? Normally, two assertions have the logical form of a conjunction, but that may be wrong here. But, even their disjunction isn't correct, because it doesn't recognize that the two claims about expenditure are inconsistent. One solution has this logical form:

> Income £20 & (the result is happiness if and only if expenditure is less than or equal to £20).

No more parsimonious way exists to capture the possibilities that Micawber had in mind, but, unlike his advice, this formulation covers the case of equal income and expenditure, and it implies that the worst that can happen is that you won't be happy rather than that you'll be miserable. The difficulty of recovering such a logical form has led many theorists to argue that logic is irrelevant to inferences in daily life (see e.g. Toulmin, 1958, and for an empirical test of his approach, see Green & McCloy, 2003).

One reason for the difficulty of logical analysis is that everyday reasoning depends on the meanings of propositions, whereas logic does not. Another is that reasoning depends on knowledge of context, whereas logic does not. And yet another difference is that reasoning depends on general knowledge and beliefs, whereas logic does not. In short, everyday reasoning depends on the meanings of words, general knowledge, and beliefs. The consequent difficulty of making logical analyses is borne out by one simple fact. No computer program exists for recovering the logical form of everyday inferences.

The present paper sets the stage for its arguments with a description of how logical form works in a simple branch of logic. It then describes how logical form would have to work with inferences in everyday life. It outlines an alternative to mental logic in the theory of mental models, and it illustrates a prediction of this theory: the content of assertions and the contexts in which they occur can modulate the interpretation of logical terms in the language, such as *if* and *or*. It describes how this phenomenon creates great difficulties for the analysis of logical form. And it shows that such an analysis leads to the need for infinitely many axioms to capture the logical properties of simple relations, such as *on the right of*. Many of these problems, however, do not arise if reasoning is based on mental models.

Logical form in sentential logic

The sentential calculus is a branch of logic that concerns implications depending on negation and on idealized versions of sentential connectives, such as *and* and *or*. Its modern formulations contain three components (see, e.g., Jeffrey, 1981):

1. a grammar, which specifies the well-formed sentences in the calculus and their logical forms;
2. a proof theory, which specifies a set of formal rules of inference for proving that conclusions follow from premises in virtue of their logical forms;
3. a model theory, which provides the semantics of the connectives and an independent method for determining the validity of inferences.

A version of the sentential calculus can be formalized using only negation and disjunction, *or*, because the other sentential connectives can be defined in terms of them, e.g.: *a and b* is equivalent to *not(not a or not b)*. The three components for this version of the calculus are summarized below.

1. The grammar has three rules specifying the set of sentences in the calculus:
   sentence = a, b, c, d, e, f, …
   sentence = *not* sentence
   sentence = (sentence *or* sentence)

Atomic propositions are those that contain neither negation nor connectives, and the first rule specifies that any variable, denoted by "a", "b", etc., with a value that is an atomic proposition, such as: *André is in Belgium*, is a sentence in the calculus. The second rule specifies that the negation of any sentence is in turn a sentence, e.g., *not b*. And the third rule specifies that the disjunction of any two sentences is in turn a sentence. The grammar is recursive because the second and third rules can be applied repeatedly to specify ever more complicated sentences, such as these two examples:

((c or not(d or e)) or ((d or e) or not f))
not((d or e) or not f)

Each sentence has only a single possible grammatical structure, which makes its logical form transparent.

2. A proof theory for the calculus consists of a set of formal rules of inference, which can be used to derive conclusions from premises. Various ways exist to formulate these rules. In one such system, the calculus includes this formal rule:

A or B; *not B*; therefore, *A*.

where *A* and *B* can have as values any sentences in the calculus, atomic or non-atomic. Hence, the rule applies to the pair of sentences above, and yields the conclusion:

(c or not(d or e)).

Proof theory determines which conclusions can be formally proved from a set of premises.

3. A model theory for the calculus specifies the meanings of its two logical terms, negation and disjunction, and provides an independent way to establish the validity of inferences. The model theory of the present calculus is specified in the two semantic rules:

   A sentence of the form *not A* is true if and only if *A* is false.
   A sentence of the form *(A or B)* is true if *A* is true, *B* is true, or both *A* and *B* are true, and false otherwise. This semantics corresponds to an *inclusive* disjunction, which is true when both its clauses are true.

These rules can be expressed in an equivalent way in the well-known format of truth tables, and they yield what is known as a "truth functional" account of the meaning of disjunction. A *valid* inference is defined in model theory as one in which the conclusion is true in every case in which the premises are true (Jeffrey, 1981, p. 1). Various methods, such as a truth table for a negation of the premises disjoined with the conclusion, can be used to establish whether or not an inference is valid.

Logicians have proved (in metalogic) that the proof theory of the sentential calculus captures all the inferences that are valid in its model theory, and vice versa. They have also proved that a finite decision procedure exists for determining whether or not any inference in the calculus is valid. This useful state of affairs does not apply to every logical calculus. For example, the more powerful logic known as the *first order predicate calculus* lacks a complete decision procedure. This calculus includes sentential logic, but also deals with inferences based on logical analogs of the quantifiers, *any* and *some*, with variables ranging over individuals in the universe of discourse. But, it is only semi-decidable, i.e., validity can be established in a finite number of steps, but no such procedure can exist to establish invalidity (see, e.g., Boolos & Jeffrey, 1989).

The essential step in formulating psychological theories of reasoning inspired by formal logic is to convert the proof theory of a logic, such as the one above, into a theory of deductive reasoning. This view still has proponents, though, as we will see, it is difficult to defend. But, I turn first to an alternative theory of human reasoning.

Mental models and the modulation of connectives

The theory of *mental* models makes the radical assumption that reasoning depends, not on logical form, but on mental models (Johnson-Laird, 1983, 2006; Johnson-Laird and Byrne, 1991). According to this theory, individuals use the meaning of all words – not just those of logical terms, the grammatical structure of sentences, and knowledge, to construct models of the possibilities to which propositions refer. They infer that a conclusion is valid if it holds in all these possibilities, i.e., there is no model of the premises in which the conclusion doesn't hold. But, the theory from its inception has distinguished between rapid intuitive inferences and slower deliberative inferences (see Ch. 6 of Johnson-Laird, 1983). It is accordingly what is now known as a "dual-process" theory (see, e.g., Stanovich, 1999; Evans, 2003; Verschueren, Schaeken, & d'Ydewalle, 2005). Not all such theories, however, specify how the two sorts of reasoning work together, or what the processes are on which they rely. Such an algorithm, however, is built into the model theory (e.g., Johnson-Laird & Byrne, 1991, Ch. 9). It presupposes that intuitive processes have no access to working memory, and so they cannot be recursive. As a corollary, intuitive processes construct only a single model of the premises. In contrast, deliberative processes use working memory, and so they can search recursively for alternative models.

The theory is analogous to the model theory of logic, but mental models are based on three fundamental assumptions, which distinguish them from other proposed sorts of mental representation. First, each mental model represents a possibility – strictly speaking, it represents what is common to a whole set of possibilities. Hence, there are models of the two outcomes of tossing a coin: one model represents that the coin came up heads, and the other model represents that it came up tails – even though there are many possible ways in which each outcome could occur. Second, mental models are iconic insofar as they can be. An iconic representation is one in which each part of the representation corresponds to each part of what it represents (for the notion of iconicity, see Volume 4 of Peirce, 1931–1958). Third, mental models are based on a principle of *truth*: they represent only those situations that are possible given a proposition, and each model of a possibility represents only what is true in that possibility according to the proposition (Johnson-Laird & Savary, 1999). An exclusive disjunction, such as:

Either there's a circle or else there isn't a triangle

refers to two possibilities, and excludes the case in which both its clauses are true. Hence, it has two *mental* models, denoted here on two separate lines:

o	
	not-Δ

where 'not' is a symbol representing negation. When the intellectual demands of a task are not so great, individuals can use the meaning of the proposition to flesh out their mental models into *fully explicit* models:

o	Δ
not-o	not-Δ

The principle of truth reduces the processing load on working memory, and, as Vandierendonck claimed, evidence shows that inferences that depend on multiple models are more difficult than those that depend on a single mental model (see Johnson-Laird & Byrne, 1991; Duyck, et al., 2003). But, the principle of truth also predicts a striking phenomenon. The failure to represent what is false leads to systematic fallacies from certain sorts of premise, and individuals succumb to these fallacies (see, e.g., Johnson-Laird, 2006). For example, a restaurant's menu states this condition about what you can have to eat:

You have the bread, or else you have the soup or else the salad.

Suppose that you have the bread. Can you have the soup *and* the salad? Nearly everyone infers that you can't. And the prediction of this response follows from the mental models of what you can eat according to the premise:

bread		
	soup	
		salad

But, the response is wrong, because *or else* means that one clause is true and the other clause is false. Granted that you have the bread, the second clause of the assertion: *you have the soup or else the salad*, is accordingly false. And one way in which it can be false is that you have both the soup and the salad. Hence, the correct response is that you can have both of them. Experiments have corroborated the occurrence of such illusions (Khemlani & Johnson-Laird, 2009). Theories that posit only correct formal rules of inference have no way of explaining the illusions. If they introduce incorrect rules to account for them, they run the risk that the rules are inconsistent – a disaster from a logical standpoint.

Logical form yields a concept of provability that concerns whole sets of inferences that have the same form. In contrast, as I mentioned earlier, the concept of validity in the model theory of a calculus is that an inference is valid if its conclusion is true in every case in which its premises are true (Jeffrey, 1981, p.1). The model theory adopts this concept of validity, and it is applicable to each inference in its own right. The existence of other invalid inferences – apparently of the same logical form – is irrelevant to the evaluation of the particular inference under consideration. Only a single question matters: is there any case in which the premises are true, but the conclusion false? If not, the particular inference is valid.

To reveal the problems of logical form, useful test cases are inferences based on conditional assertions, i.e., those of an *if-then* structure. Psychologists have reported many studies of reasoning from premises of this sort, but their experiments have tended to use abstract materials in order to prevent meaning and knowledge from affecting performance (see Evans, Newstead, & Byrne, 1993). An example of such an inference is:

> If there is a triangle then there isn't a circle.
> There is a circle.
> So, there isn't a triangle.

The inference is valid, because in every case in which its premises are true its conclusion is true too. The conclusion is also provable in psychological theories based on formal rules. The inference is an instance of a formal pattern of inference known as modus tollens:

> If A then B.
> Not B.
> Therefore, not A.

In the actual inference above *B* is the negative proposition that there isn't a circle, so *not B* is the affirmative proposition that there is a circle. Consider another example of the same apparent logical form:

> If she played a musical instrument then she didn't play a flute.
> She played a flute.
> So, she didn't play a musical instrument.

It is again provable using formal rules of inference. But, individuals are much less likely to make this inference, because they know that a flute is a musical instrument. Hence, if she played a flute, she played a musical instrument, and the conclusion above is false.

The model theory explains the difference between the two inferences as a consequence of the different meanings of conditionals, which are a result of interactions among a small set of simple components (Johnson-Laird and Byrne, 2002; Johnson-Laird, Byrne, & Girotto, 2008; Byrne & Johnson-Laird, 2009). One component yields a core meaning for conditionals, and another component is a mechanism for *modulation* that can transform the core meaning into an indefinite number of different meanings. The core meaning for a conditional typically occurs in conditionals that have no semantic relations between their clauses other than their co-occurrence in the same conditional. The conditional, *if there is a triangle then there is a circle,* has the core meaning, which refers to three different possibilities:

Δ	o
not-Δ	o
not-Δ	not-o

These possibilities correspond to those of material implication, a connective in the sentential calculus, which can be defined in terms of the inclusive disjunction: *there isn't a triangle or there is a circle* (Jeffrey, 1981, p. 61). Children develop the ability to list the possibilities for the core meaning (Barrouillet & Lecas, 1998, 1999; Barrouillet, Grosset, & Lecas, 2000). Younger children tend to list just the first possibility above; older children list the first and the third possibilities; and still older children, adolescents, and adults, list all three possibilities.

The principle of truth outlined earlier stipulates that mental models represent what is true, not what is false (Johnson-Laird & Savary, 1999). And the mental models of the core meaning consist of a single mental model that represents the first possibility above in which the if-clause and the then-clause are both true, and an implicit mental model – a place holder with no explicit content as shown by the ellipsis below – that represents the other possibilities in which the if-clause is false:

Δ o
. . .

These models explain the greater ease of conditional inferences based on a premise asserting the proposition in the if-clause (modus ponens) than of those based on a premise denying the proposition in the then-clause (modus tollens; see Evans et al., 1993, for the evidence).

The meaning of the clauses in conditionals and co-referential relations between them can modulate the core meaning of a conditional in a process of semantic *modulation*. Likewise, knowledge about the context and the topic of the conditional can modulate the core meaning in a process of *pragmatic* modulation. The effects of the two sorts of modulation are similar. One effect is to block the construction of models of possibilities; another effect is to add information, such as a temporal relation, to models of possibilities. Depending on which possibilities are blocked, conditionals have ten different classes of meaning (Johnson-Laird & Byrne, 2002). This phenomenon, in turn, explains the rejection of the earlier inference about the flute. The conditional, *if she played a musical instrument then she didn't play a flute*, does not refer to the three possibilities of the core meaning. Instead, it refers to just two possibilities about what she played:

musical instrument	not-flute
not-musical instrument	not-flute

The third possibility of the core meaning is:

not-musical instrument	flute

but individuals know that a flute is a musical instrument, and this knowledge blocks the construction of this possibility. The further premise, *she played a flute*, contradicts the two models of the conditional, and so the two premises together yield the null model, which represents contradictions, and no further conclusion can be drawn.

To recapitulate, individuals tend to accept the inference that there isn't a triangle, but to reject the inference that she didn't play a musical instrument. This difference is predicted by modulation, and its effects have been corroborated experimentally (Quelhas, Johnson-Laird, & Juhos, 2010). So, how is this difference to be explained in terms of mental logic? One possibility is to argue that the two conditionals differ in logical form. But, this argument raises the further question of how individuals can determine the logical form of:

If she played a musical instrument then she didn't play a flute.

A plausible answer is that they use their knowledge that a flute is a musical instrument to infer that the conditional refers to only two possibilities – the two possibilities displayed above. They can then describe these possibilities as:

> Either she played a musical instrument and didn't play a flute or she didn't play a musical instrument and didn't play a flute.

And, from this description, they then recover the logical form:

A or else not-A, and not-B.

One drawback to this solution is that the representation of possibilities is not part of any existing formal rule theories. Indeed, it is a step towards the model theory, and after its introduction no need exists for formal rules. Individuals can reason from the possibilities alone. If theorists do retain formal rules, then there is still a problem in blocking the inference to the conclusion that she didn't play a musical instrument. To grasp this difficulty, consider an alternative way to try to explain the difference between the two inferences.

A second defense of formal rules is that the inference about the musical instrument is an *enthymeme*, i.e., it is missing a premise, and general knowledge provides this premise: if she played a flute then she played a musical instrument. The complete premises for the inference are accordingly:

If she played a musical instrument then she didn't play a flute.
She played a flute.
If she played a flute then she played a musical instrument.
What follows?

From the second and third premises, formal rules imply that she played a musical instrument; and from this conclusion and the first premise, formal rules imply that she didn't play a flute. So, the premises yield a contradiction:

She played a flute and she didn't play a flute.

In the sentential calculus, however, any conclusion whatsoever follows from a contradiction. This consequence is blocked in psychological theories based on formal rules, because they do not allow reasoners to work forwards from a contradiction to a new conclusion. Nevertheless, *given* the putative conclusion that she didn't play a musical instrument, these theories yield its proof from the first and third premises (see, e.g., Rips, 1994, p. 116), and there is no obvious way to block the inference. Its blocking calls for the detection of the contradiction – no simple matter in formal rule theories (see Johnson-Laird, Girotto, & Legrenzi, 2004) – and some sort of injunction on inferences from contradictions.

Modulation also occurs with disjunctive assertions (as Tom Ormerod and the author have shown in unpublished experiments). Consider, for example, the inference:

Pat is in Brazil or Viv is in Paris
Viv is not in Paris.
Therefore, Pat is in Brazil.

The disjunctive premise has an inclusive interpretation in which both its clauses are true, and so the premise yields three mental models:

(Pat Brazil)	
	(Viv Paris)
(Pat Brazil)	(Viv Paris)

where the parentheses denote an iconic model of spatial inclusion, and so (Pat Brazil) represents Pat as in Brazil. The second premise that Viv is not in Paris eliminates the second and third models, and it follows from the remaining model that Pat is in Brazil.

The model theory postulates that general knowledge takes the form of fully explicit models of the known possibilities. The content of a current assertion triggers pertinent knowledge, which is then conjoined with the mental models of the assertion. For instance, consider the following disjunctive assertion in which "she" is co-referential with Pat:

Pat is in Rio or she is in Norway.

Plainly, the disjunction is exclusive. Individuals know that Norway and Brazil are different countries, and that Rio is in Brazil:

(Rio Brazil)	(Norway)

The conjunction of this model with those of the disjunction yields an exclusive interpretation corresponding to two possibilities:

((Pat Rio) Brazil)	(Norway)
(Rio Brazil)	(Pat Norway)

The possibility that Pat is in both Rio and Norway is blocked by the spatial separation of the two countries. The first inference above, in which the disjunction refers to three models, is therefore harder than following inference, in which the disjunction refers to only two possibilities:

Pat is in Norway or she is in Rio.
Pat is not in Rio.
Therefore, Pat is in Norway.

Knowledge can block possibilities in the interpretation of disjunctions. Consider this example:

Pat is in Rio or she is in Brazil.

Because individuals know that Rio is in Brazil, they should interpret the assertion as compatible with only two possibilities:

((Pat Rio) Brazil)
(Pat Rio Brazil)

Despite its disjunctive form, the assertion conveys definite information: Pat is in Brazil. All that is uncertain is whether or not she is in Rio. This interpretation should impede the inference:

Pat is in Rio or she is in Brazil.
Pat is not in Brazil.
Therefore, Pat is in Rio.

because if Pat is not in Brazil, she can't be in Rio. The model theory accordingly predicts that individuals should be inhibited from making this inference. Once again, the second premise contradicts the interpretation of the first premise, and so the treatment of the inference as an enthymeme, which is missing the premise: *If Pat is in Rio then she is in Brazil*, fails to explain why individuals should balk at the inference. In an experiment, we tested 100 Italian high school graduates, who had to choose which of two putative conclusions – an affirmative one, such as *Pat is in Rio*, and its negation – had to be true given the truth of the premises. The results corroborated the predicted effects of modulation. The helpful contents yielded more formally provable conclusions (81%) than the neutral contents (75%), which in turn yielded more formally provable conclusions than inhibitory contents (54%; Page's trend test, L (100) = 1278, z = 5.52, p < .00001).

Mental models can represent the temporal order of events (e.g., Santamaría and Espino, 2002; Byrne, 2005) and explicit temporal relations, such as, "John takes a shower before he drinks coffee" (Schaeken, Johnson-Laird, & d'Ydewalle, 1996). But, modulation can also add information about temporal relations to models. The addition of these sorts of relation can occur with premises that make no explicit assertions about temporal order, e.g.:

If Lisa received the money, then she paid Frederico.
If she paid Frederico, then he bought a new laptop.
Lisa received the money.
Did Lisa receive the money before Frederico bought a new laptop?

Participants in an experiment tended to respond, "yes" (Quelhas et al., 2010). Here is a contrasting problem with a similar grammatical form:

> If Tania gave Mauro a scooter, then he did well on the exams.
> If he did well on the exams, then he studied a lot.
> Tania gave Mauro a scooter.
> Did Tania give Mauro a scooter before he studied a lot?

In this case, the participants tended to respond, "no". They inferred that Mauro's studying preceded his performance on the exams, which in turn preceded Tania's gift of the scooter. These inferences show that modulation is not just an effect of the grammatical order of clauses in a conditional, but depends on knowledge about the typical sequences of events in everyday life. This knowledge is then embodied in mental models of the premises, and so reasoners infer the typical temporal sequence.

The conditional, "if she put a book on the shelf then it fell off", conveys both a temporal and a spatial relation: if she put the book on the shelf then *afterwards* it fell off and finished up *below* the shelf. The model theory allows that the core logical meanings of connectives can be modulated by the meanings of the clauses that they interconnect, their referents, and knowledge of the topic or context. There are indefinitely many distinct relations of this sort, and so there are indefinitely many interpretations of conditionals. We have modeled the process computationally using fully explicit models to represent knowledge in long-term memory.

A putative criticism of the model theory is that it postulates a "truth functional" account of sentential connectives, i.e., one in which the truth of assertions based on them depends solely on the truth values of the clauses that they connect. Evans & Over (2004) make this claim about conditionals. They defend instead their "suppositional" account of conditionals, i.e., the idea that conditionals elicit a supposition that their if-clause holds, and the then-clause is evaluated in this context. In fact, the model theory postulates that only the core meaning of a conditional corresponds to material implication in the sentential calculus, and that even this meaning concerns possibilities rather than truth values. And modulation calls for an interpretative process that cannot be "truth functional". For instance, a conditional such as: *If Lisa received the money then she paid Frederico*, is not true just because both of its clauses are true: the two events have to occur in the correct temporal order. Indeed, no sentential connectives in natural language, including conditionals, can be interpreted in a truth functional way (Johnson-Laird & Byrne, 2002, p. 673; Byrne & Johnson-Laird, 2009a, b; pace Evans & Over, 2009). The model theory does accommodate suppositions, albeit in a less constrained way than the suppositional theory, e.g., the evidence shows that individuals sometimes make a supposition of the then-clause of a conditional (Van der Henst, Yang, & Johnson-Laird, 2009).

How can the inferences about Lisa's payment and Tania's gift be treated in formal terms? The grammatical forms of the conditionals give no clues to their true logical forms, which yield opposite inferential consequences. From a formal standpoint, these consequences must depend on both a temporal logic and logical forms that make explicit temporal relations. Temporal logics exist, but there is a gap between them and temporal assertions in natural language, which distinguish between the time of an utterance, the time of an event it refers to, and a so-called "reference" time usually established in previous utterances (see, e.g., Miller & Johnson-Laird, 1976, Sec. 6.2). So, a formal analysis calls for a temporal logic that has yet to be devised.

The phenomena bear out the view that the complexity of sentential connectives, such as conditionals and disjunctions, is a result of the interaction among several components. One component yields a core interpretation, and other components use meaning, reference, and knowledge to modulate this interpretation. One effect of modulation is to block the construction of possibilities, and thereby to yield a variety of interpretations. This modulation in turn has predictable consequences for reasoning. The inferences that individuals deem to be valid depend on the possibilities to which the premises refer. As these possibilities change from one sort of interpretation to another so, too, does the pattern of inferences that individuals make. Another effect of modulation is to add information, such as a temporal or spatial relation, to models of events. Both these effects of modulation have been corroborated in experiments (Quelhas et al., 2010).

Logical form and reasoning about relations

Here is a simple inference that young children can make (Bryant and Trabasso, 1971):

> Ann is taller than Beth.
> Beth is taller than Cath.
> Therefore, Ann is taller than Cath.

A variety of theories account for such inferences, but unfortunately the data are not sufficiently discriminating to distinguish amongst them. According to theories based on formal rules, however, the inference is not provable, because it depends on more than the given premises. Once again, it is an enthymeme. It depends on a missing premise, which is an axiom to the effect that the relation, *is taller than*, yields transitive inferences:

> For any x, y, z, if x is taller than y, and y is taller than z, then x is taller than z.

Such axioms are known as *meaning postulates* because they capture the logical properties of the meanings of words (see Bar-Hillel, 1967). In the proof of the inference above, the value of *x* is set to Ann, the value of *y* is set to Beth, and the value of *z* is set to Cath. The rest of the proof is easy. The premises match the antecedent conjunction in the postulate of transitivity, and so the conclusion follows at once from a single formal rule: *If A then B; A*; therefore *B*, where the value of *A i*s the conjunction of the two premises of the inference, and *B* is the conclusion.

Now, consider this inference:

> Ann is a blood relative of Beth.
> Beth is a blood relative of Chris.
> What follows?

Many people draw the conclusion that Ann is a blood relative of Chris (Goodwin and Johnson-Laird, 2008). But, the inference is invalid, as a counterexample shows: Anne is Beth's mother (so, they are blood relatives), and Beth's father is Chris (so, they are blood relatives), yet her mother and father are *not* blood relatives. So, why do people draw the conclusion? The model theory predicts the error on the grounds that individuals tend to construct models of *typical* situations, and in this case a typical situation is that the three individuals are siblings, or lineal descendants. Indeed, an independent study of the diagrams that individuals drew to represent the premises corroborated this claim. And, of course, mental models of these typical situations yield the conclusion. The experiments also included genuine transitive relations, such as "is taller than", from which the participants drew transitive conclusions, and relations that are not transitive, such as "loves", from which the participants did not draw transitive conclusions. Hence, the results cannot be explained in terms of some general tendency to draw conclusions from premises laid out in the same way as the example above. The fallacies seem to occur because individuals overlook possibilities – the single most prevalent error in all sorts of thinking (Johnson-Laird, 2006) – and, in the present case, they fail to think of an alternative possibility that is a counterexample to their conclusion.

The model theory makes a further prediction. If individuals err because they think of simple and typical situations, then one remedy should be to cue them to think of alternatives that refute invalid transitive conclusions. They should then be more likely to refrain from pseudo-transitive inferences. For example, when they are told to bear in mind the consequences of marriage on kinship, they should be less likely to draw the invalid transitive conclusion from the premises above. Experiments established the existence of various pseudo-transitive relations, and corroborated both the invalid inferences from them, and the ameliorating effect of cues to models other than typical ones (Goodwin and Johnson-Laird, 2008).

How can formal rule theories explain these results? One way might be to postulate that pseudo-transitive relations are ambiguous: they have a sense that is transitive and a sense that is not transitive. Only the transitive sense invites a transitive inference (Politzer, 2004). This hypothesis is applicable, say, to a relation such as, *in front of*, that has both a sense concerning an individual's point of view, which is transitive, and a sense concerning the intrinsic parts of an object, which is not transitive. Hence, an assertion such as:

Steve is in front of John

can be true from the speaker's point of view, but false in the intrinsic sense if the two of them are standing back to back. A sensible test for ambiguity is accordingly that a sentence can be true in one sense, but false in another. This test shows that *blood relative* is not ambiguous. There are different sorts of blood relation, some of which are transitive (*sibling of*), and some of which are not (*parent of*). But, a proposition asserting a blood relation between two individuals is true if any of the different sorts of blood relation holds. Hence, the term *blood relative* refers to any sort of blood relation, and so it has a single meaning.

Another way to try to account for the pseudo-transitive fallacies in terms of formal rules is to suppose that reasoners make an assumption by default. Tsal (1977), for instance, observed that individuals often assumed by default that an unknown relation was transitive and symmetric: they knew nothing about the relation because it was denoted by a meaningless symbol. Hence, individuals might draw a pseudo-transitive conclusion from relations, such as *blood relative*, because they assume transitivity by default. The default assumption, however, is overruled by the linguistic cue to marriage. The problem with this account is that individuals do not make a default assumption that all relations are transitive. As the experiments showed, they made no such assumption for relations that are intransitive. So, until the account provides a mechanism to explain which relations are assumed to be transitive, and how context can block the assumption, this approach appears to be no more than a re-description of the results rather than an explanation of them.

The results create a dilemma for formal rule theories as they are currently formulated. On the one hand, individuals infer a transitive conclusion from a relation such as *blood relative*, and so they must have a meaning postulate that the relation is transitive. It follows that they should draw a transitive conclusion for the relation whenever they encounter premises that have a logical form that fits the antecedent condition of the meaning postulate. But, if so, how are they able to refrain from the inference in a context that reminds them that individuals can be related by marriage? On the other hand, if individuals do not draw the transitive conclusion in contexts that cue marriage, then they do not have the transitive meaning postulate for the relation. But, if so, how are they able to draw the transitive inference in the absence of this context? Consider a third example of a potentially transitive inference:

Cate is taller than Belle.
Belle *was* taller than Alice.
Who is tallest?

The change in tense no longer guarantees transitivity, and indeed many people no longer make the inference (see Goodwin and Johnson-Laird, 2008). The moral of this example is that no relational terms, not even "taller than" can be classified as transitive in all cases. Transitivity depends on the significance of the proposition as a whole, which in turn can depend on the meaning of the sentence and its context.

As a final example of a putatively transitive inference, consider these premises:

Philip is to the right of James.
James is to the right of Thomas.
What follows?

If the three individuals are seated down one side of a table, as they are in Leonardo's painting of the Last Supper, a relation such as *is to the right of* is transitive, and you can validly infer:

Philip is to the right of Thomas.

If, instead, the three individuals are seated round a small circular table, the relation is intransitive, and you might infer:

Philip is opposite to Thomas.

But, if the individuals were just three of those sitting round a very large circular table, next to one another, then you would make the transitive inference. Transitivity would extend over a certain number of individuals, but gradually break down over a larger number of individuals as they got further and further round the table. To capture these vagaries using formal rules calls in principle for an infinite number of meaning postulates for the relation of *is to the right of*. At one end of the continuum, the relation is intransitive (the small round table), next it is transitive over three individuals but not four (a slightly larger round table), and so on, up to a relation that is unboundedly transitive (a long rectangular table). It is the context of the premises that matters – the actual seating arrangement to which they refer. A single meaning of *is to the right of* determines the axis on which other individuals need to be if they are to satisfy the relation, and this meaning captures the vagaries of inference given access to a model of the seating arrangement. Transitivity emerges from models, but, as pseudo-transitivity shows, it sometimes emerges erroneously.

General Discussion

In summary, a crucial issue for theories of reasoning based on formal rules of inference is the recovery of logical form. In a logical calculus, logical form is transparent and specified by the grammar of the sentences in the calculus, and we saw earlier exactly how such a system works for a simple formalization of the sentential calculus. In natural language, however, inferences hinge on the propositions that sentences express. The underlying grammatical form of sentences in natural language is sometimes referred to as "logical form" (Chomsky, 1995), but it does not capture the logical form that has to match formal rules of inference. The present paper has raised three principal arguments against the use of logical form in this logical sense in the process of human reasoning. The first argument is that logical form transcends grammatical form, because it depends on both the meaning of sentences and the context in which they occur. For instance, the Secretary of State could assert, pointing to two areas on a map:

Bin Laden is here, or he is here in this area.

Unless you know which two areas the Secretary identified, you cannot tell whether her disjunction is exclusive, as it would be if she pointed to the Republic of Yemen and then to Afghanistan, or inclusive, as it would be if she pointed first to Razmak and then to the tribal regions of Pakistan. In the first case, given the further intelligence that Bin Laden is not in the second area (Afghanistan), you could validly infer that he is in the first (Yemen). In the second case, given the intelligence that Bin Laden is not the second area (the tribal regions of Pakistan), you could not validly infer that he is in the first area (Razmak), because Razmak is in the centre of the tribal regions.

Context can also overrule the grammatical form of a sentence. For example, a mother observing her child about to grab a forbidden cake can assert:

I'll smack you.

The force of this utterance is, not a categorical assertion that the mother will smack her child, but rather that *if,* and presumably *only if,* the child takes the cake, then the mother will smack her (Johnson-Laird, 1986). The correct logical form therefore corresponds to a biconditional. This logical form cannot be recovered from the sentence alone.

The second argument against logical form counters possible ways to try to deal with context. In one proposed method, the meaning of assertions is used to represent the possibilities to which the assertion refers. This method is a step in the direction of the model theory, and indeed the use of formal rules then becomes unnecessary. If they are retained, however, they suffer from the same problem of inconsistency as another way to try to cope with context. In this other way, the regions that the Secretary of State identifies are translated into additional premises, which supplement her assertions. Hence, the complete premises for the inference about Bin Laden's location are as follows:

Bin Laden is here, or he is here in this area.
Bin Laden is not here in this area.
Here he is in Razmak.
Here in this area he is in the tribal regions of Pakistan.
Razmak is in the tribal regions of Pakistan.

It follows that Bin Laden is here in Razmak. But, it also follows that he is not in the tribal regions, and therefore that he is not in Razmak. So, the premises are inconsistent. Knowledge alone can have a similar effect, as in the earlier example:

Pat is in Rio or she is in Brazil.
Pat is not in Brazil.
Therefore, Pat is in Rio.

When we add the premise that Rio is in Brazil, the complete premises are inconsistent. But, psychological theories based on formal rules have no obvious way to block the inference that Bin Laden is in Razmak, or the inference that Pat is in Rio.

The third argument against logical form concerns reasoning about relations. From a formal standpoint, it calls for meaning postulates that specify the logical properties of relations, such as their transitivity. These properties, however, are also highly dependent on meaning and context. A change of tense blocks the transitive inference in the case of:

Cate is taller than Belle.
Belle *was* taller than Alice.
Who is tallest?

And, as the model theory predicts, individuals tend to draw transitive conclusions from premises, such as:

Ann is a blood relative of Beth.
Beth is a blood relative of Chris.
What follows?

They think of a typical situation, such as three lineal descendants, and the resulting model yields the conclusion. But, a clue that marriage creates relations is enough to inhibit the inference. These phenomena are hard to explain in terms of a meaning postulate for *is a blood relative of*. Either reasoners have the postulate or not. With it, they should make the inference; without it, they should not make the inference. And so the postulate cannot explain the experimental results. In the case of the relation, *is to the right of*, the degree of transitivity over individuals ranges from none, to three only, to four only, and so on … up to any arbitrary number. Each degree of transitivity calls for its own meaning postulate, and so the uses of the relation as a whole call for infinitely many meaning postulates.

Consistent with the three arguments against logical form, the formal analysis of everyday inferences is extraordinarily difficult (see Keene, 1992). If theories take the recovery of logical form for granted (e.g., Rips, 1994), formal rules can account for many of the inferences in the experiments described in this paper. But, to take logical form for granted is to sweep an elephant under a carpet. The problem lies behind Bar-Hillel's (1969) famous remark that the scandal of twentieth century logic is its failure to grapple with everyday reasoning.

Other phenomena present severe challenges to formal rules. One is the occurrence of illusory inferences, i.e., the predictable and systematic fallacies described earlier in the paper (see, e.g., Johnson-Laird & Savary, 1999; Khemlani & Johnson-Laird, 2009; Goodwin & Johnson-Laird, 2010). Another is that logic allows infinitely many different conclusions to follow validly from any premises. For instance, these premises:

If the software is correct then there is a flaw in the chip.
The software is correct.

validly yield the following infinite series of conclusions:

The software is correct.
The software is correct and the software is correct.
The software is correct and the software is correct and the software is correct.

Of course such conclusions are preposterous. No sane individual – other than a logician – is likely to draw them. Yet they are all valid deductions in logic. Hence, logic alone cannot be a theory of deductive competence (pace Inhelder and Piaget, 1958). In fact, a frequent valid conclusion drawn from these premises is:

There is a flaw in the chip.

Given, say, the following two unrelated premises and asked what follows from them:

Spider phobia is not contagious.
The bus goes to Greenwich Village.

most people respond: "nothing." Yet, to repeat, logic permits infinitely many valid conclusions from any premises. So, the response that "nothing follows" is an error in logic. However, the conclusions that follow are neither interesting nor useful, e.g.:

Spider phobia is not contagious and the bus goes to Greenwich Village.

People are sensible enough not to draw just any valid conclusion, and sometimes to respond that nothing follows. Logic is at best an incomplete theory of the conclusions that individuals draw. It has nothing to say about which particular valid conclusions they draw, or about why they should ever respond that nothing follows. Theories based on formal rules recognize these difficulties (e.g., Rips, 1994), and so they couch their rules to make it impossible for individuals to prove nonsensical, though valid, conclusions. Unfortunately, if you were equipped only with these rules, you would fail to understand the force of the present argument. That is, you do in fact grasp that the nonsensical conclusions are valid, but you would be unable to do so if you had only the formal rules postulated in the theory. They do not allow you to make, or to understand, valid but nonsensical inferences.

In contrast, the model theory postulates that individuals draw parsimonious conclusions that maintain all the semantic information in the premises, and that do not add disjunctive alternatives over and above those to which the premises refer (Johnson-Laird & Byrne, 1991, Ch. 2). If no such conclusion can be drawn, naïve individuals respond sensibly that nothing follows. Yet, they also understand that an inference is valid given that there are no counterexamples to it: the conclusion holds in all the models of the premises. And so they can appreciate that a nonsensical inference is nevertheless valid if it has no counterexamples.

A further limitation on formal rules of inference is that their purview is narrow. They are restricted to valid deductions in most theories, and to inferences that can be captured in existing branches of logic. The approach is therefore not readily applicable to modal reasoning about what is possible and what is impossible, because these modalities in daily life diverge from those in all current modal logics (see, e.g., Karttunen, 1972; Byrne, 2005; Johnson-Laird, 2006). Likewise, the theories have little or nothing to say about how individuals discover that a set of assertions is inconsistent, and what they do to rectify inconsistencies. In contrast, the model theory allows for a process in which they withdraw a conclusion which conflicts with the facts – a form of so-called “nonmonotonic” reasoning (Brewka, Dix, & Konolige, 1997) – and then formulate an explanation that resolves the inconsistency (see, e.g., Johnson-Laird, et al., 2004).

Fortunately, you don’t need logical form or meaning postulates if you reason using models. You consider the meanings of the premises, you take context and knowledge into account, and then you imagine the possibilities compatible with this information, though you prefer to work with just a single mental model, which suffices for intuitions (Johnson-Laird, 1983, Ch. 6). If a conclusion holds in each of your models, you consider that the inference is valid. The transitivity of relations is an emergent property of building iconic models based on the meanings of relations, such as, *is taller than* and *is on the right of*. These models can represent many relations that are not asserted in the premises, and so your task is to find a useful relation and to formulate a conclusion embodying it. Experimental evidence corroborates this account, and a computer program implementing it shows that transitivity need not depend on meaning postulates, but can emerge from the meanings of relations.

Is it possible to refute the formal approach to human reasoning? Probably not, and therein lies a weakness. Experiments have shown that particular versions of the hypothesis are false, because the version predicts that one sort of inference should be easier than another – according to the number of steps in their proofs – but experiments yield the opposite findings (e.g., Byrne & Johnson-Laird, 1989; García-Madruga, Moreno, Carriedo, Gutiérrez, & Johnson-Laird, 2001). For example, formal rule theories treat inclusive disjunction as fundamental and define exclusive disjunction, in part, in terms of it, and so proofs based on exclusive disjunctions should be harder than those based on inclusive disjunctions (see, e.g., Rips, 1994). The model theory makes the opposite prediction, because an exclusive disjunction has two models whereas an inclusive disjunction has three models. The results corroborate the model theory (see, e.g., Bauer & Johnson-Laird, 1993). A new formal rule theory, however, could accommodate these results by changing the rules to make exclusive disjunction basic. In contrast, the model theory would be refuted once and for all if it could be shown that inferences that demand the construction of multiple mental models are easier than those in the same domain that demand the construction of only a single model.

On at least one account, both mental models and formal rules are wrong, because reasoning is not deductive but intrinsically probabilistic (e.g., Oaksford & Chater, 2007). One difficulty with this view, as I mentioned at the outset, is that it offers no obvious explanation of how human beings were able to devise logic and mathematics if they were incapable of deductive reasoning beforehand. Another difficulty is that individuals untrained in logic are able to assess whether or not a set of assertions is consistent (e.g., Johnson-Laird et al., 2004). The task is deductive: a common method in logic to prove that a conclusion follows from premises is to add its negation to the premises, and show that the resulting set of assertions is inconsistent (Jeffrey, 1981). But, probabilistic theories, as yet, have no way of assessing the consistency of assertions, e.g., the probability of a conjunction of a set of propositions is zero if they are inconsistent, and they may be inconsistent even though every proper subset of them is consistent. The solution may be to seek a way to integrate probabilistic considerations, which play a role in reasoning, with the model theory (see, e.g., Oaksford & Chater, 2010).

What this article has tried to show is that previous theories based on formal rules have finessed a very difficult problem: how reasoners recover the logical form of propositions. Unlike logic, the grammar of a sentence in natural language cannot alone yield the logical forms of propositions that the sentence can express. They can be determined only from the meaning of the sentence in context, which depends in part on knowledge, and even, as Stenning and Van Lambalgen (2008) suggest, on the particulars of an inferential task. These authors rightly argue that the recovery of logical form is itself dependent on reasoning. The claim is dangerous, of course, because this reasoning must depend on premises of some sort, and how is their logical form to be determined? There is the danger of a vicious circle, and we need to know how it is prevented. The problems of logical form are familiar to logicians. The late Jon Barwise (1989, p. 159), for instance, argued that everyday reasoning is not a formal process, and that the notion of logical form is unilluminating for natural language. He emphasized that content, not form, determines validity (p. 4). Hence, it is strange that psychological theories based on formal rules have yet to grapple with the recovery of logical form. No algorithm exists for its recovery; and no decisive evidence exists that it plays any role in human reasoning.

What are the open problems that the model theory has yet to solve? One major problem is the construction of models from perception, which can play a crucial role in the representation of the context of utterances. In computer programs implementing the model theory, knowledge is represented as fully explicit models, which can be conjoined with models based on verbal premises in the process of modulation (see, e.g., Johnson-Laird, et al., 2004). The assumption is plausible, but how such models are acquired is another major puzzle for which we have no solution as yet (but cf. Goodwin & Johnson-Laird, 2010).

Finally, readers may wonder how the theory of mental models relates to logic. It clearly resembles the model-theoretic procedures that logicians invoke (e.g., Jeffrey, 1981). The resemblance is closest for sentential reasoning. Yet, a clear divergence occurs between mental models and logical models. Because of the principle of truth, mental models are sometimes erroneous, whereas logical models are impeccable. The moral is clear: logic is neither a theory of what conclusions individuals tend to draw nor a theory of how they draw them. It tells us one important thing: which premises in a formalized language logically imply which conclusions. It is not dispensable, but it is not the direct route to a psychological theory.

References

Bar-Hillel, Y. (1967). Dictionaries and meaning rules. *Foundations of Language, 3*, 409-414.

Bar-Hillel, Y. (1969). Colloquium on the role of formal languages. *Foundations of Language, 5*, 256-284.

Barrouillet, P., & Lecas, J-F. (1998). How can mental models theory account for content effects in conditional reasoning? A developmental perspective. *Cognition, 67*, 209-253.

Barrouillet, P., & Lecas, J-F. (1999). Mental models in conditional reasoning and working memory. *Thinking & Reasoning, 5*, 289-302.

Barrouillet, P., Grosset, N., & Lecas, J. F. (2000). Conditional reasoning by mental models: chronometric and developmental evidence. *Cognition, 75*, 237-266.

Barwise, J. (1989). *The Situation in Logic*. Stanford, CA: Center for the Study of language and Information.

Bauer, M.I., & Johnson-Laird, P.N. (1993). How diagrams can improve reasoning. *Psychological Science, 4*, 372-378, 1993.

Bledsoe, W.W. (1977). Non-resolution theorem proving. *Artificial Intelligence, 9*, 1-35.

Boolos, G., & Jeffrey, R. (1989). *Computability and Logic*, 3rd Ed. Cambridge: Cambridge University Press.

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

Brewka, G., Dix, J., & Konolige, K. (1997). *Nonmonotonic Reasoning: An Overview.* Stanford, CA: CLSI Publications, Stanford University.

Bryant, P.E., & Trabasso, T. (1971). Transitive inferences and memory in young children. *Nature, 232*, 456-458.

Byrne, R.M.J. (2005). *The Rational Imagination: How People Create Alternatives to Reality*. Cambridge, MA: MIT.

Byrne, R.M.J., & Johnson-Laird, P.N. (1989). Spatial reasoning. *Journal of Memory and Language, 28*, 564-575.

Byrne, R.M.J., & Johnson-Laird, P.N. (2009a). 'If' and the problems of conditional reasoning. *Trends in Cognitive Sciences, 13*, 282-286.

Byrne, R.M.J., & Johnson-Laird, P.N. (2009b). Models redux: Response to Evans and Over. *Trends in Cognitive Sciences, 14*, 6.

Cherniak, C. (1986). *Minimal Rationality*. Cambridge, MA: MIT Press.

Chomsky, N. (1995). *The Minimalist Program*. Cambridge, MA: MIT Press.

Dierckx, V., Vandierendonck, A., & Pandelaere, M. (2003). Is model construction open to strategic decisions? An exploration in the field of linear reasoning. *Thinking & Reasoning, 9*, 97-131.

Duyck, W., Vandierendonck, A., & De Vooght, G. (2003). Conditional reasoning with a spatial content requires visuo-spatial working memory. *Thinking & Reasoning, 9*, 267-287.

Evans, J. St. B. T. (2003). In two minds: Dual process accounts of reasoning. *Trends in Cognitive Sciences, 7*, 454-459.

Evans, J.St.B.T., & Over, D.E. (2004). *If.* Oxford: Oxford University Press.

Evans, J.St.B.T., & Over, D.E. (2009). Conditional truth: comment on Byrne and Johnson-Laird. *Trends in Cognitive Sciences, 14*, 5.

Evans, J.St.B.T., Newstead, S.E., & Byrne, R.M.J. (1993). *Human Reasoning: The Psychology of Deduction*. Hillsdale, NJ: Lawrence Erlbaum Associates.

García-Madruga, J.A., Moreno, S., Carriedo, N., Gutiérrez, F., & Johnson-Laird, P.N. (2001). Are conjunctive inferences easier than disjunctive inferences? A comparison of rules and models. *Quarterly Journal of Experimental Psychology, 54A*, 613-632.

Goodwin, G.P., & Johnson-Laird, P.N. (2008). Transitive and pseudo-transitive inferences. *Cognition, 108*, 320-352.

Goodwin, G. P., & Johnson-Laird, P. N. (2010). Conceptual illusions. *Cognition, 114*, 253-265.

Green, D.W., & McCloy, R. (2003). Reaching a verdict. *Thinking & Reasoning, 9*, 307-333.

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

Inhelder, B., & Piaget, J. (1958). *The Growth of Logical Thinking from Childhood to Adolescence*. London: Routledge & Kegan Paul.

Jeffrey, R. (1981). *Formal Logic: Its Scope and Limits*. 2nd Ed. New York: McGraw-Hill.

Johnson-Laird, P.N. (1975). Models of deduction. In Falmagne, R. (Ed.). *Reasoning: Representation and Process* (pp. 7-54). Springdale, NJ: Erlbaum.

Johnson-Laird, P.N. (1983). *Mental Models: Towards a Cognitive Science of Language, Inference and Consciousness*. Cambridge: Cambridge University Press; Cambridge, MA: Harvard University Press.

Johnson-Laird, P.N. (1986). Conditionals and mental models. In Ferguson, C., Reilly, J., ter Meulen, A., & Traugott, E.C. (Eds.), *On Conditionals* (pp. 55-75). Cambridge: Cambridge University Press..

Johnson-Laird, P.N. (2006). *How We Reason*. New York: Oxford University Press.

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

Johnson-Laird, P.N., & Byrne, R.M.J. (2002). Conditionals: A theory of meaning, pragmatics, and inference. *Psychological Review, 109*, 646-678.

Johnson-Laird, P.N., Byrne, R.M.J., & Girotto, V. (2009). The mental model theory of conditionals. *Topoi, 28*, 75-80.

Johnson-Laird, P.N., Girotto, V., & Legrenzi, P. (2004). Reasoning from inconsistency to consistency. *Psychological Review, 111*, 640-661.

Johnson-Laird, P.N., & Savary, F. (1999). Illusory inferences: A novel class of erroneous deductions. *Cognition, 71*, 191-229.

Karttunen, L. (1972). Possible and must. In Kimball, J.P. (Ed.). *Syntax and Semantics* (pp. 1-20), Vol. 1. New York: Seminar Press..

Keene, G. B. (1992). *The Foundations of Rational Argument*. Lampeter, Wales: Edwin Mellen Press.

Khemlani, S., & Johnson-Laird, P.N. (2009). Disjunctive illusory inferences and how to eliminate them. *Memory & Cognition, 37*, 615-623.

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

Miller, G.A., & Johnson-Laird, P.N. (1976). *Language and Perception*. Cambridge, MA: Harvard University Press.

Oaksford, M., & Chater, N. (2007). *Bayesian Rationality: The Probabilistic Approach to Human Reasoning*. Oxford: Oxford University Press.

Oaksford, M., & Chater, N. (2010). Conditional inference and constraint satisfaction: Reconciling mental models and the probabilistic approach. In Oaksford, M. & Chater, N. (Eds.) *Cognition and Conditionals: Probability and Logic in Human Thinking* (pp. 309-333). Oxford: Oxford University Press..

Osherson, D.N. (1974-1976). *Logical Abilities in Children*, Vols. 1-4. Hillsdale, NJ: Lawrence Erlbaum Associates.

Peirce, C. S. (1931-1958). C. Hartshorne, P. Weiss & A. Burks (Eds.), *Collected papers of Charles Sanders Peirce*, 8 Vols. Cambridge, MA: Harvard University Press.

Politzer, G. (2004). Reasoning, judgement and pragmatics. In Noveck, I.A., & Sperber, D. (Eds.), *Experimental Pragmatics* (pp. 94-115). Hampshire, UK: Macmillan.

Pollock, J. (1989). *How to Build a Person: A Prolegomenon*. Cambridge, MA: MIT Bradford Books.

Quelhas, A.C., Johnson-Laird, P.N., & Juhos, C. (2010). The modulation of conditional assertions and its effects on reasoning. *Quarterly Journal of Experimental Psychology*, in press.

Rips, L.J. (1983). Cognitive processes in propositional reasoning. *Psychological Review, 90*, 38-71.

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

Santamaría, C., & Espino, O. (2002). Conditionals and directionality: On the meaning of *if* versus *only if*. *Quarterly Journal of Experimental Psychology*, *55A*, 41-57.

Schaeken, W.S., Johnson-Laird, P.N., & d'Ydewalle, G. (1996). Mental models and temporal reasoning. *Cognition, 60*, 205-234.

Sperber, D., & Wilson, D. (1986). *Relevance: Communication and Cognition*. Oxford: Basil Blackwell.

Stanovich, K.E. (1999). *Who is Rational? Studies of Individual Differences in Reasoning*. Mahwah, NJ: Erlbaum.

Stenning, K., & van Lambalgen, M.S. (2008). *Human Reasoning and Cognitive Science*. Cambridge, MA: MIT Press.

Tsal, Y. (1977). Symmetry and transitivity assumptions about a nonspecified logical relation. *Quarterly Journal of Experimental Psychology, 29*, 677-684.

Toulmin, S.E. (1958). *The Uses of Argument*. Cambridge: Cambridge University Press.

Van der Henst, J-B., Yang, Y., & Johnson-Laird, P.N. (2002). Strategies in sentential reasoning. *Cognitive Science, 26*, 425-468.

Vandierendonck, A., & De Vooght, G. (1996). Evidence for mental-model-based reasoning: A comparison of reasoning with time and space concepts. *Thinking & Reasoning, 2*, 249-272.

Vandierendonck, A., & De Vooght, G. (1997). Working memory constraints on linear reasoning with spatial and temporal contents. *Quarterly Journal of Experimental Psychology, 50A*, 803-820.

Vandierendonck, A., Dierckx, V., & De Vooght, G. (2004). Mental model construction in linear reasoning: Evidence for the construction of initial annotated models. *Quarterly Journal of Experimental Psychology, 57A*, 1369-1391.

Verschueren, N., Schaeken, W., & d'Ydewalle, G. (2005). A dual-process specification of causal conditional reasoning. *Thinking & Reasoning, 11*, 278-293.

Wos, L. (1988). *Automated Reasoning: 33 Basic Research Problems*. Englewood Cliffs, NJ: Prentice-Hall.