Mental models and deduction

Philip N. Johnson-Laird

According to the mental-model theory of deductive reasoning, reasoners use the meanings of assertions together with general knowledge to construct mental models of the possibilities compatible with the premises. Each model represents what is true in a possibility. A conclusion is held to be valid if it holds in all the models of the premises. Recent evidence described here shows that the fewer models an inference calls for, the easier the inference is. Errors arise because reasoners fail to consider all possible models, and because models do not normally represent what is false, even though reasoners can construct counterexamples to refute invalid conclusions.

In a civil action, an expert witness conceded two points:

If a pollutant (TCE) had come from the river, it would be in the river bed.

TCE was not in the river bed.

The opposing lawyer claimed:

The pattern is consistent with the conclusion that no TCE came from the river.

The witness agreed (Ref. 1, pp. 361–362). But the lawyer had made a mistake: the evidence is not merely consistent with the conclusion. It yields a valid deduction of the conclusion, that is, if the evidence is true then the conclusion must be true too. The example shows that human reasoning is not always successful. Cognitive scientists need to understand the causes of such failures.

Reasoning lies at the core of human intelligence[2]. And it is central to science, society, and the solution of practical problems. It starts with premises, which can be statements, perceptions or beliefs. Ideally, it yields a valid conclusion that is not explicit in the premises. But the intervening processes are mysterious. Many theorists suppose that the mind constructs syntactic representations of the logical form of assertions and applies the rules of a formal logic to them[3,4]. There is another possibility: reasoners could rely instead on their grasp of meaning, their general knowledge, and principles akin to those for the semantics of a logic[5]. They construct mental models of the premises, which represent the situation, and draw conclusions from them[6,7]. This article outlines the theory and its recent developments.

The theory of mental models

Thinking depends on tacit processes that are guided by constraints: the thinker's goal, if any, and pertinent knowledge and beliefs. The idea that it depends on mental models goes back to the Scottish psychologist, Kenneth Craik, who suggested that perception constructs 'small-scale models' of reality that are used to anticipate events and to reason[8]. Mental models might originally have evolved as the ultimate output of perceptual processes. They can represent spatial relations[9], events and processes[10], and the operations of complex systems[11]. They can yield both inductive and deductive inferences. But what exactly is a mental model? The current theory makes three main assumptions, which distinguish models from syntactic representations of logical form, from semantic networks, and from other sorts of proposed mental representation. Possibilities lie at the heart of the mental-model theory, and the first assumption relates them to models:

(1) *Each mental model represents a possibility*

It captures what is common to the different ways in which the possibility might occur. Like a diagram, a model is iconic, that is, its parts correspond to the parts of what it represents, and its structure corresponds to the structure of the possibility. An exclusive disjunction, such as:

Either TCE is in the river or else it doesn't come from the river

allows one possibility or the other, but not both. It calls for two mental models to represent the two possibilities:

TCE-in-river

¬ *TCE-comes-from-river*

where '*TCE-in-river*' denotes a mental model of the possibility in which TCE is in the river, '¬' denotes negation, and so '¬ *TCE-comes-from-river*' denotes a model of the possibility in which TCE does not come from the river. Indeed, individuals describe these two possibilities when they are asked to state what is possible given the assertion. The diagram above uses words, but that does not imply that mental models are made up of words. They represent the relations between the TCE and the river. Models in general can represent relations among three-dimensional entities or abstract entities; they can be static or kinematic. They underlie visual images, although many components of models are not visualizable[12].

(2) *The principle of truth: mental models represent what is true according to the premises, but by default not what is false*

This principle applies at two levels. At one level, models represent only the possibilities that are true given a premise, as do the models of the disjunction above. At a lower level, however, a model represents a clause in the premises only when it is true in the possibility. For example, the first model of the disjunction: TCE-in-river, represents TCE in the river, but it does not represent explicitly that in this possibility it is *false* that TCE does not come from the river, that is, it comes from the river.

Philip N. Johnson-Laird
Dept of Psychology, Princeton University, Green Hall, Princeton, NJ 08544, USA.
e-mail: phil@princeton.edu

http://tics.trends.com 1364-6613/01/$ – see front matter PII: S1364-6613(00)01751-4

The principle of truth postulates that individuals by default do not represent what is false. But there are exceptions that overrule the principle. Individuals make 'mental footnotes' about the falsity of clauses, and if they retain the footnotes they can flesh out *mental* models into *fully explicit* models, which represent clauses even when they are false. The following fully explicit models, for example, represent the earlier exclusive disjunction:

TCE-in-river *TCE-comes-from-river*
¬ *TCE-in-river* ¬ *TCE-comes-from-river*

Nevertheless, the principle of truth is the norm. It makes for parsimonious representations, because reasoners do not have to bother with what is false. This parsimony, as we will see, comes at a price.

Mental models can represent discourse about real, hypothetical, or imaginary situations[13]. They can reside in long-term memory as a representation of knowledge[14,15]. And they can be used for deductive reasoning, according to the rational principle that a conclusion is valid if it holds in all the models of the premises. The third principle embodies this idea.

(3) *Deductive reasoning depends on mental models*

If a conclusion holds in all the models of the premises, that is, it has no counterexamples, it is necessary given the premises. If it holds in a proportion of models, its probability is equal to that proportion, granted that the models represent equiprobable alternatives. If it holds in at least one model, it is possible given the premises. And if it holds in none of the models, it is impossible given the premises. The theory therefore unifies deductive reasoning about necessity, probability and possibility.

Consider how a computer program implementing the theory copes with an inference from a disjunction and a categorical assertion:

Either TCE is in the river or else it doesn't come from the river

TCE does come from the river

The program parses the disjunction and uses the semantics for '*or else*' to construct models of the two possibilities:

TCE-in-river
¬ *TCE-comes-from-river*

It combines its model of the categorical premise with each of these models in turn. The result eliminates the second model, which cannot be true. Only the first model remains, and so it follows that:

TCE is in the river

This conclusion is valid because it holds in all the models – in this case, the single model – of the premises. The ability to draw valid conclusions is compatible with other theories of reasoning[16]. However, the mental-model theory yields some crucial predictions.

One model is better than many

The fewer the number of models needed for an inference, and the simpler they are, the easier the inference should be. It should take less time, and be less prone to error. This prediction is a consequence of the limitations of working memory[17]: multiple models can overload its processing capacity and lead to errors in which reasoners fail to consider some models of the premises. Halford and his colleagues have shown that the fewer the entities in a model of a relation, the easier inferences tended to be[18]. Similarly, the difficulty of problems requiring multiple models has been corroborated in studies of relational, sentential, and quantified reasoning[19,20]. Schaeken and his colleagues, for instance, have investigated problems based on temporal relations[21], as in the following example about everyday events. '*John takes his shower before he drinks his coffee*':

a before b
b before c
d while b
e while c
What is the relation between d and e?

The computer program implementing temporal reasoning constructs a model of the premises within a framework in which time is represented on the horizontal axis and contemporaneity on the vertical axis:

a *b* *c*
d *e*

People, on the other hand, might construct a kinematic model that unfolds in time. Granted that all the events are of a comparable length, it follows from either the static or the kinematic model that *d* occurs before *e*. A problem in which the second premise is modified to: *c* before *b*, calls for multiple models because *a* might occur before *c*, after *c*, or at the same time as *c*. Problems requiring one model elicit a greater number of correct responses than problems requiring multiple models, and a premise calling for one model takes less time to read than the corresponding premise yielding multiple models.

Similar effects occur in other domains[22,23]. Vandierendonck and his colleagues have corroborated them in temporal and spatial reasoning[24]. But differences between two and three models are often so small that it is unlikely that reasoners constructed all three models[25]. They are more likely to have built a single static model with one element represented as having two or more possible locations. When reasoning calls for spatial models, spatial tasks such as visual tracking interfere with performance, but do not impair other sorts of reasoning[26,27].

One model is indeed better than many for human reasoners. This principle predicts a crucial interaction in modal reasoning (see Box 1). When reasoners construct one model, the availability of its elements depends on their positions in the model[28]. When they have to construct multiple models, they are liable to fail to envisage a model[29], and so their conclusions correspond to only some of the models of the premises[30]. For example, given a syllogism of the form:

Box 1. Models and modal reasoning

Modal reasoning is about what is possible and what is necessary[a]. The mental-model theory predicts a crucial interaction. It should be quite easy to infer a possibility, because reasoners need to find only a single model of the premises in which the conclusion holds. It should be harder to infer that something is not possible, because reasoners need to check that it does not hold in any of the models of the premises. By contrast, difficulty should switch round in the case of necessity: it should be easier to infer that something is not necessary (one model suffices) than to infer that something is necessary (all models have to be checked). As an example, consider the following problem about a one-on-one basketball game (only two can play):

If Allan is in then Betsy is in.
If Carla is in then David is out.
Can Betsy be in the game?

The first premise allows as a possible game:

Allan vs Betsy

which is consistent with the second premise. Hence, it should be easy to respond 'yes' to the question. It should be harder to make the correct 'no' response to the following 'dual' of the problem:

If Allan is out then Betsy is out.
If Carla is out then David is in.
Can Betsy be in the game?

Reasoners need to consider the models of all three possible games to ensure that the response is correct.

Table I. The percentages of correct responses to modal inference problems

	Correct 'yes' responses (%)	Correct 'no' responses (%)
Questions about what is possible	91 (18.0)[a]	65 (22.3)
Questions about what is necessary	71 (25.6)	81 (22.7)

[a]Average latency of the response in seconds

However, when the two sets of premises are paired with the question:

Must Betsy be in the game?

the correct 'yes' response to the first set of premises should be harder than the correct 'no' response to the second set of premises. Table I shows the results of a study corroborating the interaction[b]. Evans and his colleagues have observed analogous results in modal reasoning based on premises containing quantifiers, such as *all* and *some*[c].

References

a Hughes, G.E. and Cresswell, M.J. (1996) *A New Introduction to Modal Logic*, Routledge
b Bell, V. and Johnson-Laird, P.N. (1998) A model theory of modal reasoning. *Cognit. Sci.* 22, 25–51
c Evans, J. St B.T. *et al.* (1999) Reasoning about necessity and possibility: a test of the mental model theory of deduction. *J. Exp. Psychol. Learn. Mem. Cognit.* 25, 1495–1513

None of the A is a B
All the B are C

they draw a conclusion, such as *None of the A is a C*, which is compatible with only one model of the premises[31]. They fail to realize that any *C*s that are not *B*s could be *A*s. Hence, they miss the valid conclusion interrelating *A*s and *C*s, which is: *Some of the C are not A*. They may have misinterpreted the second premise, taking it to mean in addition that all the *C*s are *B*s[32,33].

Several theorists have proposed that because working memory is limited, reasoners construct as few models as possible, and often just a single model. Ormerod and his colleagues have pioneered the study of immediate inferences from one sort of conditional to another, and from conditionals to disjunctions, and vice versa[34]. They argue that reasoners construct the minimal set of models needed to infer the conclusion[35]. Similarly, Sloutsky and his colleagues have observed that reasoners often base their conclusions on just a single model of the premises[36]. The meaning of assertions, however, might call for multiple models, and reasoners can spontaneously construct them. An important instance of this phenomenon occurs in reasoning from counterfactual conditionals (see Box 2).

Truth, meaning, and knowledge

Reasoners focus on what is true and neglect what is false. One consequence is the difficulty of the selection task (Box 3). Another consequence is the occurrence of systematic fallacies (Box 4). And because meaning is central to models, the content of inferences and background knowledge can modulate reasoning. They influence the interpretation of premises[37]. The following inference, for example, is valid in form:

Eva is in Rio or she's in Brazil;
She's not in Brazil.
Therefore, she's in Rio.

But no sensible person other than a logician is likely to draw this conclusion. It is impossible for Eva to be in Rio but not in Brazil, because Rio is in Brazil. By contrast, the following inference is easy[38]:

Eva is in Rio or else she's in Norway;
She's not in Norway.
Therefore, she's in Rio.

The content of premises and background knowledge can lead to the addition of information to models. A conditional such as:

If Viv entered the lift as it was going up, then Pat left it at the next floor

Box 2. Conditional reasoning

There are four main conditional inferences:

If A then B, A	∴ *B*. (modus ponens)
If A then B, not B	∴ *not A*. (modus tollens)
If A then B, B	∴ *A*. (affirming the consequent premise)
If A then B, not A	∴ *not B*. (denying the antecedent premise)

Affirming the consequent and denying the antecedent are valid if the conditional is interpreted as a biconditional: *if, and only if, A then B*. Otherwise, they are fallacies. The mental-model theory postulates that conditionals have one explicit and one implicit mental model:

A B
...

The implicit model (denoted by the ellipsis) has no explicit content, but it has a 'mental footnote' that it represents the possibilities in which the antecedent, *A*, is false. Modus ponens is easier than modus tollens (cf. the lawyer's mistake at the beginning of this article). The mental models allow modus ponens immediately, whereas the categorical premise for modus tollens, *Not-B*, eliminates the explicit model. Only the empty implicit model is left, and so it seems that nothing follows. But modus tollens is feasible if reasoners build fully explicit models of the conditional:

A B
¬*A B*
¬*A* ¬*B*

The categorical premise, *Not-B*, eliminates all but the last of these models, from which the conclusion, *Not-A*, follows necessarily. Girotto *et al.* corroborated a prediction from the mental-model theory: modus tollens is easier when the categorical premise is presented first rather than second[a]. Presented first, it provides an initial negative model: ¬ *B*, which prevents the construction of the explicit mental model of the conditional. Hence, reasoners are more likely to flesh out their models to include the case above in which the antecedent is false.

Byrne and her colleagues have developed the mental-model theory of counterfactual thinking[b,c]. The counterfactual meaning of a conditional of the form:

If A had happened then B would have happened

conveys both what is true and what is false. This meaning overrules the principle of truth, and calls for mental models of both the factual and the counterfactual possibilities:

Factual:	¬*A* ¬*B*
Counterfactual:	*A B*

The factual model is the one needed for modus tollens, and so, as predicted, the inference is easier to make from a counterfactual conditional than from an indicative one[d].

As Markovits and others have shown[e,f], the fallacies can be suppressed when reasoners bring to mind the possibility: ¬ *A B*, which undermines the necessity of *A* for *B*. The valid inferences can be suppressed when reasoners bring to mind the counterexample: *A* ¬ *B*, which undermines the sufficiency of *A* for *B* (Ref. g.) The availability of such possibilities, in turn, depends on knowledge[h]. Evans has investigated the effects of negation on conditional reasoning[i], and Schroyens and his colleagues have extended the mental-model theory to account for the phenomena[j,k]. Here, too, knowledge plays a role in establishing implicit mismatches between entities; for example, between '*snake*' and '*not a mammal*'[l].

One intriguing developmental trend is that young children treat conditionals as conjunctions (one fully explicit model), slightly older children treat them as biconditionals (two fully explicit models), and adolescents and adults can treat them as proper conditionals (three fully explicit models). Barrouillet and his colleagues have observed exactly this trend, and shown that it correlates with working memory capacity[m]. Young children can cope with two models[n], but they probably use the categorical premise as a retrieval cue[o].

References

a Girotto, V. *et al.* (1997) The effect of premise order in conditional reasoning: a test of the mental model theory. *Cognition* 63, 1–28
b Byrne, R.M.J. and McEleney, A. (2000) Counterfactual thinking about actions and failures to act. *J. Exp. Psychol. Learn. Mem. Cognit.* 26,1318–1331
c Byrne, R.M.J. *et al.* (2000) The temporality effect in counterfactual thinking about what might have been. *Mem. Cognit.* 28, 264–281
d Byrne, R.M.J. and Tasso, A. (1999) Deductive reasoning with factual, possible, and counterfactual conditionals. *Mem. Cognit.*, 27, 726–740
e Markovits, H. *et al.* (1998) The development of conditional reasoning and the structure of semantic memory. *Child Dev.* 64, 742–755
f Vadeboncoeur, I. and Markovits, H. (1999) The effect of instructions and information retrieval on accepting the premises in a conditional reasoning task. *Think. Reason.* 5, 97–113
g Byrne, R.M.J. *et al.* (1999) Counterexamples and the suppression of inferences. *J. Mem. Lang.* 40, 347–373
h Manktelow, K.I. and Fairley, N. (2000) Superordinate principles in reasoning with causal and deontic conditionals. *Think. Reason.* 6, 41–65
i Evans, J. St B.T. (1999) The role of negation in conditional inference. *Q. J. Exp. Psychol.*, 52A, 739–769
j Schroyens, W. *et al.* (2000) Heuristic and analytic processes in propositional reasoning with negative conditionals. *J. Exp. Psychol. Learn., Mem. and Cognit.* 26, 1713–1734
k Schroyens, W. *et al.* (2001) The processing of negations in conditional reasoning: a meta-analytic case study in mental model and/or mental logic theory. *Think. Reason.* 7, 121–172
l Schroyens, W. *et al.* (2000) Conditional reasoning with negations: implicit and explicit affirmation or denial and the role of contrast classes. *Think. Reason.* 6, 221–251
m Barrouillet, P. and Lecas, J-F. (1999) Mental models in conditional reasoning and working memory. *Think. Reason.* 5, 289–302
n Barrouillet, P. *et al.* (2000) Conditional reasoning by mental models: chronometric and developmental evidence. *Cognition* 75, 237–266
o Markovits, H. (2000) A mental model analysis of young children's conditional reasoning with meaningful premises. *Think. Reason.* 6, 335–347

Box 3. The selection task

This well-known task was devised by Wason more than 30 years ago[a]. The experimenter places four cards on a table, which have a letter or number visible to the participant:

A B 2 3

The participants know that each card has a number on one side and a letter on the other side. They have to choose which cards to turn over to determine whether the following rule is true or false about the four cards:

If a card has an 'A' on one side, then it has a '2' on the other side.

To make the correct selections, they need to overrule the principle of truth in order to envisage the counterexample to the conditional:

A ¬ 2

and then to choose the corresponding cards: A and 3. Most people fail. They think instead of the salient possibility in which the conditional is true:

A 2

and they choose the 'A' card, and sometimes the innocuous '2' card too. Few realize the need to choose the '3' card, even though an 'A' on its other side falsifies the rule.

A change in the content of the task can yield a striking improvement in performance, particularly when participants have to select potential violations of a 'deontic' conditional, that is, a conditional about what is obligatory, such as:

If a person is drinking beer then he must be over the age of 18.

Certain conditionals tend to yield a biconditional interpretation, such as:

If you tidy your room then you can go out to play.

Performance is then susceptible to the participants' point of view[b]. Those with the parent's concern that the child might cheat tend to select the cards corresponding to the counterexample:

¬ tidy played

Those with the child's concern that the parent might renege on the deal tend to select the cards corresponding to the counterexample:

tidied ¬ play

Those with a neutral point of view tend to select all four cards.

Theorists have debated the causes of these effects. Some postulate that reasoners use schemas for reasoning about deontic matters[c]. Others postulate that they use pragmatic knowledge to select what is relevant[d]. Evolutionary psychologists postulate that they rely on an innate module for reasoning about cheaters[e]. Still others, notably Oaksford and Chater, argue for a probabilistic approach in which it is rational not to select the 3 card in the abstract task above[f,g]. Yet, according to the mental-model theory, people *are* reasoning[h,i], but they construct models of what is true, not what is false, especially if they lack cognitive ability[j]. Hence, any manipulation that helps them to consider counterexamples, and to match them to the corresponding cards, should improve performance[k]. This explanation is compatible with the effects of probability[l]. It predicts effects of point of view with biconditionals that are not deontic[m], and Sloman *et al.* have corroborated their occurrence[n].

References

a Wason, P.C. (1966) Reasoning. In Foss, B.M. (Ed.) *New Horizons in Psychology* (Vol. 1), pp. 135–151, Penguin

b Manktelow, K.I. and Over, D.E. (1995) Deontic reasoning. In *Perspectives on Thinking and Reasoning: Essays in Honour of Peter Wason* (Newstead, S.E. and Evans, J. St B.T., eds), pp. 91–114, Erlbaum

c Holyoak, K.J. and Cheng, P.W. (1995) Pragmatic reasoning with a point of view. *Think. Reason.* 1, 289–313

d Sperber, D. *et al.* (1995) Relevance theory explains the selection task. *Cognition* 52, 3–39

e Fiddick, L. *et al.* (2000) No interpretation without representation: the role of domain-specific representations and inferences in the Wason selection task. *Cognition* 77, 1–79

f Oaksford, M. and Chater, N. (1998) A revised rational analysis of the selection task: exceptions and sequential sampling. In *Rational Models of Cognition* (Oaksford, M. and Chater, N., eds), pp. 372–398, Oxford University Press

g Oaksford, M. and Chater, N. (2001) The probabilistic approach to human reasoning. *Trends Cognit. Sci.* 5, 349–357

h Green, D.W. (1997) Hypothetical thinking in the selection task: amplifying a model-based approach. *Curr. Psychol. Cognit.* 16, 93–102

i Feeney, A. and Handley, S.J. (2000) The suppression of q card selections: evidence for deductive inference in Wason's Selection Task. *Q. J. Exp. Psychol.* 53A, 1224–1242

j Stanovich, K.E. and West, R.F. (1998) Cognitive ability and variation in selection task performance. *Think. Reason.* 4, 193–230

k Liberman, N. and Klar, Y. (1996) Hypothesis testing in Wason's selection task: social exchange cheating detection or task understanding. *Cognition* 58, 127-156

l Green, D.W. *et al.* (1997) Probability and choice in the selection task. *Think. Reason.* 3, 209–235

m Johnson-Laird, P.N. and Byrne, R.M.J. (1996) A model point of view: a comment on Holyoak and Cheng. *Think. Reason.* 1, 339–350

n Almor, A. and Sloman, S.A. (2000) Reasoning versus memory in the Wason selection task: non-deontic perspective on perspective effects. *Mem. Cognit.* 28, 1060–1070

establishes a spatial and temporal relation between the events referred to in the two clauses. In logic, connectives such as disjunctions and conditionals have idealized meanings that are 'truth functional', that is, the truth or falsity of a sentence they form depends solely on the truth or falsity of the clauses they interconnect[5]. The preceding examples, however, show that natural language is not truth functional.

Knowledge and beliefs also influence the *process* of reasoning. Individuals, for example, search harder for counterexamples to conclusions that violate their knowledge. This search is compatible with a robust phenomenon: knowledge has a bigger effect on invalid inferences than on valid inferences[19,39–43]. It also has a crucial role in deductions about probabilities (see Box 5).

Strategies in reasoning

An important recent discovery is that when individuals carry out a series of inferences, they develop strategies for coping with them. Deduction itself can be a strategy[44], and Western cultures might resort to it more than East Asian cultures[45]. However, deduction in turn elicits a variety of strategies[46]. An earlier version of the mental-model theory implied that reasoners start reasoning with the most informative premise but this claim is not always true[47]. Reasoners' strategies determine which premise they take into account first[48]. Consider a problem of the form:

Box 4. Illusory inferences

Readers are invited to solve the problems in Fig. I. For problem 1, the program implementing the mental-model theory predicts that individuals consider the true possibilities for each premise. For the first premise, they should consider the following mental models:

king
ace
king *ace*

Two of the models show that an ace is possible. Hence, individuals should respond, 'yes'. In fact, this response is wrong. It is impossible for an ace to be in the hand, because both of the first two premises would then be true, contrary to the rubric that only one of them is true. Problem 1 is an 'illusion of possibility': reasoners infer wrongly that a card is possible. A similar problem to which reasoners should respond 'no' and thereby commit an 'illusion of impossibility' can be created by replacing the two occurrences of '*there is an ace*' in Problem 1 with, '*there is not an ace*'. One experiment examined the two sorts of illusion and comparable control problems[a]. The participants succumbed to the illusions but did well with the control problems (Fig. II), and the illusions of possibility were more telling than those of impossibility (for an explanation of this difference, see Box 1).

The rubric '*one of these assertions is true and one of them is false*' is equivalent to an exclusive disjunction between two assertions: *A or else B, but not both*. This usage leads to still more compelling illusions that seduce novices and experts alike. Consider Problem 2 (Fig. I). Nearly everyone infers that there is a king in the hand[b]. The present author also succumbed in testing the program implementing the mental-model theory. Yet, it is a fallacy granted a disjunction, exclusive or inclusive, between the two conditionals. The disjunction implies that one or other of the two conditionals could be false. Suppose, for instance, that the first conditional is false. Then there could be a jack but *not* a king, a judgment with which most individuals concur. And so the conclusion that there is a king is invalid: it could be false even if the premises are true.

Many experts have fallen for illusory inferences, and then proposed ingenious explanations for their errors. For example, the premises are so complex that they confuse people. But reasoners are highly confident in their conclusions, and the control problems are equally complex. Other putative explanations concern the interpretation of conditionals[c]. However, the illusions occur with disjunctions too, and their interpretation is not controversial. The illusions corroborate the principle of truth. They occur in a variety of domains[d,e]. Some procedures alleviate them[a,f–i], but no-one has discovered a perfect antidote.

Problem 1

Only one of the following premises is true about a particular hand of cards:

There is a king in the hand or there is an ace, or both.

There is a queen in the hand or there is an ace, or both.

There is a jack in the hand or there is a 10, or both.

Is it possible that there is an ace in the hand?

Problem 2

Suppose you know the following about a particular hand of cards:

If there is a jack in the hand then there is a king in the hand, or else if there isn't a jack in the hand then there is a king in the hand.

There is a jack in the hand.

What, if anything, follows?

Fig. I. Two reasoning problems. Answer both of the problems and write down their answers; then see text.

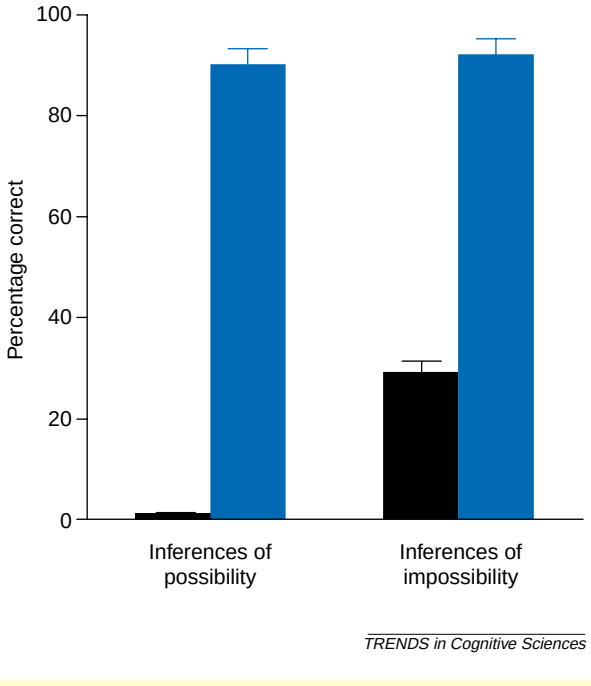

Fig. II. The percentages of correct responses to four sorts of inferences: illusions of possibility and illusions of impossibility (black bars) and their respective control inferences (blue bars). Subjects succumbed to the illusions, particularly those of possibility, but did well with the control problems.

References

a Goldvarg, Y. and Johnson-Laird, P.N. (2000) Illusions in modal reasoning. *Mem. Cognit.* 28, 282–294

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

c Rips, L.J. (1997) Goals for a theory of deduction. Reply to Johnson-Laird. *Minds Machines* 7, 409–424

d Yang, Y. and Johnson-Laird, P.N. (2000) Illusory inferences in quantified reasoning: how to make the impossible seem possible, and vice versa. *Mem. Cognit.*, 28, 452–465

e Johnson-Laird, P.N. *et al.* (2000) Illusions in reasoning about consistency. *Science* 288, 531–532

f Tabossi, P. *et al.* (1999) Mental models in deductive, modal, and probabilistic reasoning. In *Mental Models in Discourse Processing and Reasoning* (Habel, C. and Rickheit, G., eds), pp. 229–331, Elsevier

g Barrouillet, P. and Lecas, J-F. (2000) Illusory inferences from a disjunction of conditionals: a new mental models account. *Cognition* 76, 3–9

h Santamaria, C. and Johnson-Laird, P.N. (2000) An antidote to illusory inferences. *Think. Reason.* 6, 313–333

i Yang, Y. and Johnson-Laird, P.N. (2000) How to eliminate illusions in quantified reasoning. *Mem. Cognit.* 28, 1050–1059

Box 5. Mental models and deductions about probabilities

Mental-model theory explains how people deduce the probability of an event from their knowledge of the different ways in which the event can occur, that is, so-called 'extensional' reasoning about probabilities[a]. Individuals assume that each model is equiprobable unless they have knowledge to the contrary. They infer the probability of an event from the proportion of models in which the event occurs. Consider this problem:

There is a box in which there is at least a red marble, or else a green marble and a blue marble, but not all three marbles.

What is the probability that there is both a red and a blue marble in the box?

The premises elicit mental models of two possibilities:

Red

Green Blue

As these models predict, reasoners tend to infer a probability of zero for red and blue[b]. This response is an illusion, because when there is a red marble, there are three distinct ways in which it can be false that there is both a green marble and a blue marble:

Red Green ¬Blue
Red ¬Green Blue
Red ¬Green ¬Blue

Granted equiprobability with the possibility in which there is no red marble,

¬Red Green Blue

it can be seen that the probability of red and blue is 1/4.

Deductions about conditional probabilities can be difficult[c]. Consider the following problem:

The chances that Pat has the disease are 4 out of 10.

If she has the disease, then the chances are 3 out of 4 that she has the symptom.

If she does not have the disease, then the chances are 2 out of 6 that she has the symptom.

Pat has the symptom. So, what are the chances that she has the disease?

One way to infer the answer is to use Bayes's theorem from the probability calculus. However, naive individuals can build either equiprobable models of the premises, or ones that are tagged with the appropriate chances out of 10:

disease	*symptom*	3
disease	*¬ symptom*	1
¬ disease	*symptom*	2
¬ disease	*¬ symptom*	4

The conditional probability can be computed from the appropriate subset relation: there are 3 chances that Pat has the disease out of the 5 chances that she has the symptom. Evolutionary psychologists argue that these problems are soluable provided that they concern natural samples of frequencies[d,e]. However, studies show that individuals can fail with frequencies[f] and succeed with the chances of unique events[g]. Problems are soluble if it is easy to find the appropriate subset of models and to calculate the ratio.

References

a Tversky, A. and Kahneman, D. (1983) Extensional versus intuitive reasoning: the conjunction fallacy in probability judgment. *Psychol. Rev.* 90, 293–315

b Johnson-Laird, P.N. *et al.* (1999) Naive probability: a mental model theory of extensional reasoning. *Psychol. Rev.* 106, 62–88

c Nickerson, R.S. (1996) Ambiguities and unstated assumptions in probabilistic reasoning. *Psychol. Bull.* 120, 410–433

d Gigerenzer, G. and Hoffrage, U. (1995) How to improve Bayesian reasoning without instruction: frequency formats. *Psychol. Rev.* 102, 684–704

e Cosmides, L. and Tooby, J. (1996) Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgment under uncertainty. *Cognition* 58, 1–73

f Evans, J. St B.T. *et al.* (2000) Frequency versus probability formats in statistical word problems. *Cognition* 197–213

g Girotto, V. and Gonzalez, M. (2001) Solving probabilistic and statistical problems: a matter of question form and information structure. *Cognition* 78, 247–276

A if and only if B
Either B or else C, but not both
C if and only if D
Does it follow that if not A then D?

Some individuals spontaneously develop a strategy based on suppositions. When they think aloud, they say, for instance:

'Suppose not A. It follows from the first premise that not B. It then follows from the second premise that C. The third premise then implies D. So, yes, the conclusion does follow.'

Each of these inferential steps can be carried out using models. Another strategy is to make an inference from a pair of premises, and then to make another from its conclusion and the third premise. Still another strategy is to draw a horizontal line across the middle of the page, and to write down the possibilities compatible with the premises (Fig. 1). These individuals work through the premises in whatever order they are stated, and even take into account irrelevant premises. When individuals are taught to use this strategy in a systematic way, their reasoning is both faster and more accurate (V. Bell, unpublished studies in the author's laboratory).

Reasoners develop diverse strategies for relational reasoning[49], suppositional reasoning[50–52], and reasoning with quantifiers[53]. Some strategies have surprised researchers. So far, however, they all reflect a reliance on meaning and models, although individuals who have learned logic could make a strategic use of formal rules.

Strategies can resolve a puzzle. The mental-model theory predicts that inferences based on a conjunction (one model) should be easier than those based on a

Fig. 1. A reasoner's diagram representing the two possibilities (A and B, or C and D) compatible with the premises (see text for discussion).

disjunction (multiple models). However, Rips[3] found that an inference of the following sort:

A and B
If A then C
If B then C
Therefore, C

was no easier to evaluate than:

A or B
If A then C
If B then C
Therefore, C

However, García-Madruga and his colleagues have corroborated the model theory's prediction when reasoners drew their own conclusions from these premises, as opposed to evaluating the given conclusions[54,55]. And when the premises were presented one at a time on a computer screen, the results corroborated the mental-model theory even in the evaluation task. Reasoners' strategies are likely to differ from one task to another.

An inference is valid if its conclusion holds in all the models of the premises, or if no model of the premises is a counterexample to the conclusion. Reasoners do not search for counterexamples routinely[56,57]. In a counterexample, the premises are true but the conclusion is false, and so such a model violates the principle of truth. Moreover, reasoners can often determine that an inference is valid by constructing all the models of the premises. Otherwise, however, a feasible strategy is to search for counterexamples. Given the following sort of problem:

More than half of the people at this conference speak French.
More than half of the people at this conference speak English.
Does it follow that more than half of the people at this conference speak both French and English?

reasoners spontaneously drew diagrams of counterexamples to the putative conclusion[58] (Fig. 2).

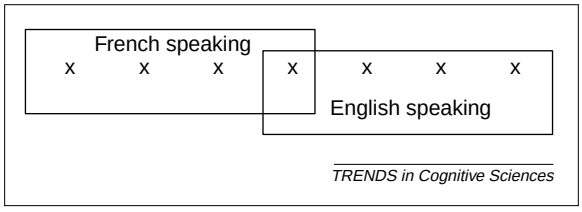

Fig. 2. A typical diagram of a counterexample to a problem (see text) drawn by a subject. In a group of 7 people (each x represents an individual), more than half of them speak French, more than half of them speak English, but it is not the case that more than half speak both languages.

Acknowledgements
Preparation of this article was supported by a grant from the National Science Foundation (Grant 0076287) to study strategies in reasoning. It was made possible by the community of reasoning researchers. There are too many individuals to name, but the author thanks them all.

Questions for future research

- What underlies the vast differences in reasoning ability from one individual to another? The processing capacity of working memory accounts for some, but not all, of these differences.
- Reasoning takes into account relevant general knowledge. What process triggers its recovery from long-term memory?
- What is the origin of the different strategies for reasoning, and how do reasoners develop them?
- Logic means never having to be sorry about a valid conclusion. In daily life, however, individuals withdraw a valid conclusion when it conflicts with subsequent facts. What are the mental processes underlying the resolution of such conflicts?
- What regions of the brain underlie deductive reasoning?

Conclusions

Deductive reasoning is under intense investigation[59]. The field is fast moving and controversial. This article has reviewed just one theory: that reasoning depends on imagining the possibilities compatible with the premises, and drawing conclusions from these mental models. The theory makes five main predictions, which have been corroborated experimentally:

(1) One model is better than many. That is, the fewer models needed for an inference, and the simpler they are, the easier the inference.

(2) Reasoners sometimes fail to consider all models in multiple-model problems. They then draw conclusions that are possible rather than necessary.

(3) When falsity matters, fallacies occur. One result is errors in the selection task; another is illusory inferences.

(4) Content and background knowledge modulate the interpretation of assertions, and so no connectives are truth functional. They also modulate the process of reasoning.

(5) With experience, reasoners develop tailor-made strategies for particular sorts of problem. To refute invalid conclusions, they can search for counterexamples.

By contrast, theories of deduction based on formal logic make no use of possibilities, and postulate a single deterministic strategy based on valid rules of inference[3,4]. They do not appear to account for any of the five preceding phenomena. Yet formal rules and mental models are not in principle incompatible. As reasoners develop, they can learn to construct formal rules for themselves in certain idealized domains, an essential step for the development of logic. The mental-model theory itself seems likely to continue to develop; it contains many lacunae, and is far from complete. It could even be overturned by the discovery of, say, sets of inferences in which multiple-model problems are easier than one-model problems. If it is refuted by systematic counterexamples, then it will at least account for its own demise.

References

1 Harr, J. (1995) *A Civil Action*, Random House
2 Stanovich, K.E. (1999) *Who is Rational? Studies of Individual Differences in Reasoning*, Erlbaum
3 Rips, L.J. (1994) *The Psychology of Proof*, MIT Press
4 Braine, M.D.S. and O'Brien, D.P., eds (1998) *Mental Logic*, Erlbaum
5 Jeffrey, R. (1981) *Formal Logic: Its Scope and Limits* (2nd Edn), McGraw-Hill
6 Johnson-Laird, P.N. and Byrne R.M.J. (1991) *Deduction*, Erlbaum
7 Bara, B.G. *et al.* (2000) In favour of a unified model of deductive reasoning. In *Mental Models in Reasoning* (García Madruga, J.A. *et al.*, eds), pp. 69–81, Madrid Universidad Nacional de Educación a Distancia
8 Craik, K. (1943) *The Nature of Explanation*, Cambridge University Press
9 Glasgow, J. (1994) Array representations for model-based spatial reasoning. In *Proc. 16th Annu. Conf. Cognit. Sci. Soc.* (Ram, A and Eiselt, K., eds), pp. 375–380, Erlbaum
10 Hegarty, M. (1992) Mental animation: Inferring motion from static displays of mechanical systems. *J. Exp. Psychol. Learn. Mem. Cognit.* 18, 1084–1102
11 Moray, N. (1999) Mental models in theory and practice. In *Attention and Performance XVII: Cognitive Regulation of Performance: Interaction of Theory and Application* (Gopher, D. and Koriat, A., eds), pp. 223–258, MIT Press
12 Johnson-Laird, P.N. (1998) Imagery, visualization, and thinking. In *Perception and Cognition at Century's End* (Hochberg, J., ed.), pp. 441–467, Academic Press
13 Garnham, A. and Oakhill, J.V. (1996) The mental models theory of language comprehension. In *Models of Understanding Text*, (Britton, B.K. and Graesser, A.C., eds), pp. 313–339, Erlbaum
14 Gentner, D. and Stevens, A.L., eds (1983) *Mental Models*, Erlbaum
15 Johnson-Laird, P.N. (1983) *Mental Models: Towards a Cognitive Science of Language, Inference, and Consciousness*, Cambridge University Press / Harvard University Press
16 Stenning, K. and Yule, P. (1997) Image and language in human reasoning: a syllogistic illustration. *Cognit. Psychol.* 34, 109–159
17 Baddeley, A. (1986) *Working Memory*, Oxford University Press
18 Halford, G.S. *et al.* (1998) Processing capacity defined by relational complexity: implications for comparative, developmental, and cognitive psychology. *Brain Behav. Sci.* 21, 803–864
19 Evans, J. St B.T. *et al.* (1993) *Human Reasoning: The Psychology of Deduction*, Erlbaum
20 Johnson-Laird, P.N. (1999) Deductive reasoning. *Annu. Rev. Psychol.* 50, 109–135
21 Schaeken, W.S. *et al.* (1996) Mental models and temporal reasoning. *Cognition* 60, 205–234
22 Carreiras, M. and Santamaría, C. (1997) Reasoning about relations: spatial and nonspatial problems. *Think. Reason.* 3, 191–208
23 Knauff, M. *et al.* (1998) Mental models in spatial reasoning. In *Spatial Cognition: An Interdisciplinary Approach to Representing and Processing Spatial Knowledge* (Freksa, C. *et al.*, eds) pp. 267-291, Springer-Verlag
24 Vandierendonck, A. and De Vooght, G. (1997) Working memory constraints on linear reasoning with spatial and temporal contents. *Q. J. Exp. Psychol.* 50A, 803–820
25 Vandierendonck, A. *et al.* (2000) Model construction and elaboration in spatial linear syllogisms. In *Deductive Reasoning and Strategies* (Schaeken, W. *et al.*, eds), pp. 191–207, Erlbaum
26 Klauer, C. *et al.* (1997) Working memory involvement in propositional and spatial reasoning. *Think. Reason.* 3, 9–47
27 Knauff, M. *et al.* Spatial reasoning: no need for visual information. In *Spatial Information Theory* (Montello, D. *et al.*, eds), Springer-Verlag (in press)
28 Schroyens, W. *et al.* (1999) Error and bias in meta-propositional reasoning: a case of the mental model theory. *Think. Reason.* 5, 29–65
29 Sloutsky, V.M. and Johnson-Laird. P.N. (1999) Problem representations and illusions in reasoning. In *Proc. 21st Annu. Conf. Cognit. Sci. Soc.* (Hahn, M. and Stones, S., eds), pp. 701–705, Erlbaum
30 Espino, O. *et al.* (2000) Eye movements during syllogistic reasoning. In *Mental Models in Reasoning* (García Madruga, J.A. *et al.*, eds), pp. 179–188, Madrid Universidad Nacional de Educación a Distancia
31 Newstead, S.E. and Griggs, R.A. (1999) Premise misinterpretation and syllogistic reasoning. *Q. J. Exp. Psychol.* 52A, 1057–1075
32 Newstead, S.E. *et al.* (1999) Falsifying mental models: testing the predictions of theories of syllogistic reasoning. *Mem. Cognit.* 27, 344–354
33 Espino, O. *et al.* (2000) Activation of end terms in syllogistic reasoning. *Think. Reason.* 6, 67–89
34 Richardson, J. and Ormerod, T.C. (1997) Rephrasing between disjunctives and conditionals: mental models and the effects of thematic content. *Q. J. Exp. Psychol.* 50A, 358–385
35 Ormerod, T.C. (2000) Mechanisms and strategies for rephrasing. In *Deductive Reasoning and Strategies* (Schaeken, W. *et al.*, eds) pp. 131–151, Erlbaum
36 Sloutsky, V.M. and Goldvarg, Y. (1999) Effects of externalization on representation of indeterminate problems. In *Proc. 21st Annu. Conf. Cognit. Sci. Soc.*, (Hahn, M. and Stones, S., eds), pp. 695–700, Erlbaum
37 Newstead, S.E. *et al.* (1997) Conditional reasoning with realistic material. *Think. Reason.* 3, 49–76
38 Bouquet, P. and Warglien, M. (1999) Mental models and local models semantics: the problem of information integration. *Proc. Eur. Conf. Cognit. Sci.* (Bagnara, S., ed.), pp. 169–178, University of Siena
39 Quayle, J.D. and Ball, L.J. (1997) Subjective confidence and the belief bias effect in syllogistic reasoning. In *Proc. 19th Annu. Conf. Cognit. Sci. Soc.* (Shafto, M.G. and Langley, P., eds) pp. 626–631, Erlbaum
40 Santamaria, C. *et al.* (1998) Reasoning from double conditionals: the effects of logical structure and believability. *Think. Reason.* 4, 97–122
41 Cherubini, P. *et al.* (1999) Can any ostrich fly? Some new data on belief bias in syllogistic reasoning. *Cognition* 69, 179–218
42 Torrens, D. *et al.* (1999) Individual differences and the belief bias effect: mental models, logical necessity, and abstract reasoning. *Think. Reason.* 5, 1–28
43 Evans, J. St B.T. (2000) Thinking and believing. In *Mental Models in Reasoning* (García Madruga, J.A. *et al.*, eds), pp. 41–55, Madrid Universidad Nacional de Educación a Distancia
44 Evans, J. St B.T. (2000) What could and could not be a strategy in reasoning. In *Mental Models in Reasoning*, (García Madruga, J.A. *et al.*, eds), pp. 1–22, Madrid Universidad Nacional de Educacion a Distancia
45 Nisbett, R. *et al.* (2001) Culture and systems of thought: holistic versus analytic cognition. *Psychol. Rev.* 108, 291–310
46 Roberts, M.J. *et al.* (1997) Individual differences and strategy selection in reasoning. *Br. J. Psychol.* 88, 473–492
47 Dekeyser, M. *et al.* (2000) Preferred premise order in propositional reasoning: semantic informativeness and co-reference. In *Deductive Reasoning and Strategies* (Schaeken, W.S. *et al.*, eds), pp. 73–95, Erlbaum
48 Johnson-Laird, P.N. *et al.* (1999) Strategies and tactics in reasoning. In *Deductive Reasoning and Strategies*, (Schaeken, W.S. *et al.*, eds), pp. 209–240, Erlbaum
49 Roberts, M.J. (2000) Strategies in relational inference. *Think. Reason.* 6, 1–26
50 Byrne, R.M.J. and Handley, S.J. (1997) Reasoning strategies for suppositional deductions. *Cognition* 62, 1–49
51 Handley, S.J. and Evans, J. St B.T. (2000) Supposition and representation in human reasoning. *Think. Reason.* 6, 273–311
52 Dieussaert, K. *et al.* (2000) Strategies during complex conditional inferences. *Think. Reason.* 6, 125–160
53 Bucciarelli, M. and Johnson-Laird, P.N. (1999) Strategies in syllogistic reasoning. *Cognit. Sci.* 23, 247–303
54 García-Madruga, J.A.G. *et al.* (2000) Task, premise order, and strategies in Rips's conjunction–disjunction and conditional problems. In *Deductive Reasoning and Strategies*, (Schaeken, W.S. *et al.*, eds), pp. 49–71, Erlbaum
55 García-Madruga, J.A.G. *et al.* (2001) Are conjunctive inferences easier than disjunctive inferences? A comparison of rules and models. *Q. J. Exp. Psychol.* 54A, 613–632
56 English, L.D. (1998) Children's reasoning in solving relational problems of deduction. *Think. Reason.* 4, 249–281
57 Handley, S.J. *et al.* (2000) Individual differences and the search for counterexamples in syllogistic reasoning. In *Deductive Reasoning and Strategies*, (Schaeken, W.S. *et al.*, eds), pp. 241–265, Erlbaum
58 Neth, H. and Johnson-Laird, P.N. (1999) The search for counterexamples in human reasoning. *Proc. 21st Annu. Conf. Cognit. Sci. Soc.* (Hahn, M. and Stones, S., eds), p. 806, Erlbaum
59 A web-page on mental models is maintained by Ruth Byrne and her colleagues at www.tcd.ie/Psychology/Ruth_Byrne/mental_models/