Topoi (2007) 26:19–36
DOI 10.1007/s11245-006-9002-8

How our brains reason logically

Markus Knauff

Abstract The aim of this article is to strengthen links between cognitive brain research and formal logic. The work covers three fundamental sorts of logical inferences: reasoning in the propositional calculus, i.e. inferences with the conditional ''if...then'', reasoning in the predicate calculus, i.e. inferences based on quantifiers such as ''all'', ''some'', ''none'', and reasoning with n-place relations. Studies with brain-damaged patients and neuroimaging experiments indicate that such logical inferences are implemented in overlapping but different bilateral cortical networks, including parts of the fronto-temporal cortex, the posterior parietal cortex, and the visual cortices. I argue that these findings show that we do not use a single deterministic strategy for solving logical reasoning problems. This account resolves many disputes about how humans reason logically and why we sometimes deviate from the norms of formal logic.

Keywords Logical thinking · Reasoning · Brain · Mental models · Mental logic

1 Logical thinking and the ''independence of computational level'' hypothesis

> It was the second half of the winter term 1888–1889. I was a medical student at the psychiatric clinic of Professor Otto Binswanger in Jena. One day a patient who had been recently committed to the institution was brought in to the lecture hall. Binswanger introduced him to us: Professor Friedrich Nietzsche! [...] At first sight, he did not appear like a sick man. He was of medium build, and his expressive face was angular, but not derelict [...]. Sometime later I saw him again and then he appeared completely different. He was in a highly excitable state and his consciousness was obviously clouded. He was sitting there with wild-painful, fiery eyes and was watched by a guardsman.

Friedrich Nietzsche was one of the greatest philosophers of the nineteenth century. The report of his mental illness was published by a medical student—later Dr. S. Simchowitz—in the Frankfurter Zeitung of Sep./07/1900 (Simchowitz 1900, the above passage is my translation from the original German newspaper article).

Many scholars have subscribed to the rumor that Nietzsche's dementia was caused by a cerebral syphilis that he had contracted some 20 years earlier (Möbius 1902). The diagnosis now seems problematic, because no Wasserman test (an antibody test for syphilis) was yet available, no autopsy was performed, and clinical grounds alone argue against the diagnosis (Fishman 2002; Sax 2003; Schain 2001). However, for a cognitive neuroscientist Nietzsche's condition is still of interest because some mental abilities were strongly affected by his brain illness while others seemed to be completely intact. Many witnesses reported that Nietzsche was unable to formulate coherent thoughts or to think rationally. His friends and relatives were unable to distinguish the ideas of the genial thinker from those of a madman. On the other hand, many biographers report that his memory was almost intact and he could speak without slurring his words (Schain 2001).

M. Knauff (✉)
Experimental Cognitive Psychology Unit, University of Giessen, Otto-Behaghel-Strasse 10F, 35394 Giessen, Germany
e-mail: markus.knauff@psychol.uni-giessen.de

Today it is well known that brain infections, or damage caused by tumors, strokes, or brain traumata, can have severe effects on some cognitive functions while others remain entirely untouched. We now know that there is a degree of modularity in the brain's overall organization (Goel 2005). For instance, we have identified brain regions specifically involved in the processing of visual information, others in the processing of spatial information, and yet more in speech comprehension and generation.

Recently a small number of psychological laboratories started to systematically investigate the connection between logical reasoning and the brain (overviews in: Goel 2005; Grafman and Goel 2002; Knauff 2006; Wharton and Grafman 1998). Before that most reasoning researchers were committed to the assumption that human reasoning should be studied in terms of computational processes. The neural implementation of these computations was considered irrelevant, because any computational function can be computed on any hardware (and thus also the brain) that is equivalent to a Turing machine (e.g. Fodor 1975; Newell 1980; Pylyshyn 1984). However, one reason for the advancing neurocognitive movement in reasoning research is the realization that universal realizability is not unqualifiedly true and thus appears to be unjustifiable as a basic assumption (Goel 2005). A second reason for the current interest of reasoning researchers is that localization in the brain can help us to understand the format of mental representations. As already mentioned there are highly specific brain regions dedicated to particular representational formats. If a reasoning process is associated with brain areas which are known to respond to verbal information, then this might support one class of reasoning theories. If it is associated with brain areas that are typically involved in the computation of information in a visual or spatial format then this speaks in favor of other theories of reasoning. It is this commitment to the testing of cognitive hypotheses that distinguishes the cognitive neuroscience of reasoning from pure brain research.

The present paper is a condensed version of an article that will appear in the Journal *Synthese* (Knauff 2007). In this long version I discuss the explanatory gap between the logical behavior we can observe and the brain activity we can measure. I argue that the cognitive neuroscience of logical reasoning also needs computational models that precisely determine what algorithms are run in the cerebral cortex. The present paper is more focused on experimental findings and brings together what is known about mental logical reasoning and the underlying brain processes. Its intention is to provide an overview of the relevant findings and to create a link between the cognitive neuroscience of reasoning and formal logic that is easily accessible for an interdisciplinary readership.

The paper covers three fundamental sorts of logical inferences. Cognitive psychologists refer to the first as *conditional reasoning*. In such inferences the statements of the problem consists of an ''if A then B'' construct that posits B to be true if A is true. The two logically valid inferences are the Modus Ponens (if p then q; p; q, MP) and the Modus Tollens (if p then q; not-q; not-p, MT). For a logician the normative model of such inferences is the propositional calculus. The second type of inference, cognitive scientists call *syllogistic reasoning*. Here the problem consists of quantified statements such as ''All A are B'', ''Some A are B'', ''No A are B'', and ''Some A are not B''. Logically speaking, the normative model for such inferences are the Aristotelian syllogisms, which are a subset of the (first order) predicate calculus. The third group of inferences treated in this paper is probably the most frequently used in daily life (and in the psychological lab). It is based on n-place relations. At least two relational terms A r_1 B and B r_2 C are given as premises and the goal is to find a conclusion A r_3 C that follows from the premises. These relations can represent spatial (e.g., left of), temporal (e.g., earlier than), or more abstract information (e.g., is akin to). In the first part of the paper, I report empirical findings on all three sorts of inferences. The findings show how we deal with such problems, what logical abilities break down after local brain injuries, and what areas of the cortex are involved if neuropsychologically healthy human beings think logically.

The second part of the paper is concerned with one of the oldest questions related to logical reasoning. What is the role of imagination in (logical) thinking? People with no education in the cognitive sciences but also many cognitive researchers often believe that our ability to reason logically relies on ''seeing with the inner eye'' (e.g. Kosslyn 1994). So do we think logically by visualizing ''mental pictures'' in the ''mind's eye'' and ''look'' at these pictures to find new, not explicitly given information? I will use the subset of relational inferences to describe the research in our lab on this question.

In the third part of the paper I formulate some general ideas on the link between formal and mental logical reasoning and the role of cognitive neuroscience in reasoning research.

2 Logical thinking from the classical and a neurocognitive perspective

When we use the term ''logical thinking'' in daily life, we mean almost all kinds of thoughts, ranging from very simple inferences up to the complex development of scientific theories. To avoid terminological confusion, in contemporary psychology we prefer the term ''deductive reasoning'' instead of ''logical reasoning'', although formally speaking both expressions mean exactly the same: an inference in which one or more propositions are true given that other propositions are true. The propositions that are taken for granted are called *premises*. The propositions that are deducted from the premises are referred to as *conclusions*. The participants of an experiment can draw the conclusions in three different ways. I will refer to them as ‘‘inference verification’’ and two sorts of ‘‘active inference’’ (Knauff et al. 1995). To make the difference explicit we can introduce the notation $\{\varphi1, \varphi2\} > \varphi3$, to denote the fact that the conclusion follows from the premises. Then the three paradigms may be written as follows (Knauff et al. 1995):

(1) inference verification: does $\{\varphi1, \varphi2\} > \varphi3$ hold?
(2a) active general inference: find all $\varphi3$ such that $\{\varphi1, \varphi2\} > \varphi3$
(2b) active particular inference: find some $\varphi3$ such that $\{\varphi1, \varphi2\} > \varphi3$

The conclusion must be generated by a human reasoner, or a statement referred to as ‘‘conclusion’’ is presented and the individual has to decide if it logically follows from the premises. Thus, in the psychology of reasoning a ‘‘conclusion’’ can be logically invalid and the response of a human reasoner is correct if he or she recognizes that it is. The words ‘‘true’’ and ‘‘false’’ and ‘‘valid’’ and ‘‘invalid’’ are strictly reserved for the logical evaluation of the statement, while the terms ‘‘correct’’ and ‘‘incorrect’’, ‘‘right’’ or ‘‘wrong’’ refer to the reasoners’ decisions.

Reasoning researchers typically distinguish between two types of correct and two types of incorrect decisions. If a presented conclusion is logically valid and the participant of the experiment says it is valid, this is counted as a correct response–a ‘‘hit’’. If the presented conclusion is logically invalid and the participants identifies it as invalid this is also a correct response—a ‘‘correct rejection’’. If the presented conclusion is logically valid and the participant says it is invalid then the response is considered incorrect—‘‘false alarm’’. If, finally, the presented conclusion is logically invalid but the participants of the experiment says it is valid this is also an incorrect response—a ‘‘miss’’. This important connection between logical validity and psychological correctness is summarized in Table 1.

Table 1 The connection between logical validity and psychological correctness

		Reasoners’ decision:	
		Valid	Invalid
Logical validity:	Valid	Hit	False alarm
	Invalid	Miss	Correct rejection

2.1 Errors, solution times, and response preferences in logical reasoning

Experiments in which the psychologist presents logical problems to test persons and then explores their responses remain an indispensable method of reasoning research. The experimental technique enables us to systematically observe behavior and then to draw inferences from the observed data about unobservable mental processes. The premises and the putative conclusion are presented on the screen and the participants are instructed to evaluate whether the conclusion follows *necessarily* or *logically* from the premises. The program records the reading times for premises and the response to the conclusion and its latency. In this way, the researcher seeks to answer one or more of the three traditional questions of reasoning research:

- What factors cause reasoning difficulty and lead people into errors?
- What are the cognitive mechanisms that enable us to reason logically (although we sometimes make mistakes)?
- How do content and background knowledge affect human logical reasoning?

Most researchers believe that humans have the competence to perform error-free deductive inferences. Errors do occur, however, because reasoning performance is limited by capacities of the cognitive system, misunderstanding of the premises, ambiguity of problems, and motivational factors (Johnson-Laird and Byrne 1991; Evans et al. 1993; Manktelow 1999). In the following I report the main findings on conditional, syllogistic, and relational reasoning, and then summarize the main ideas of the most influential cognitive theories of human reasoning.

2.1.1 Conditional reasoning

In the propositional calculus there are two logically valid and two logically invalid inferences. They are summarized in Table 2. The logically valid inferences are the modus ponens (MP) and the modus tollens (MT). However, many people also believe that they can draw a conclusion from the affirmation of consequent (AC) and the denial of antecedent (DA), both of which are logically false.

Table 3 summarizes the findings from a number of classical studies that explored how often people draw the valid inferences MP and MT and the invalid ones AC and DA from a set of conditional premises. In most of the studies the reasoners were very good in drawing the MP and they were also quite accurate in drawing the MT. However, in almost all of the studies around half of the participants also drew the two invalid conclusions DA and AC.

One of the most important experimental paradigms in reasoning research centres on the Wason-Selection-Task

Table 2 The four inference schemas of conditional reasoning

Inference schema	Logical validity	Example
MP, Modus Ponens (affirmation of antecedent) If p, then q p ——————— q	valid	If it rains, the street is wet. It rains. The street is wet.
AC, affirmation of consequent If p, then q q ——————— p	invalid	If it rains, the street is wet. The street is wet. It rains.
DA, denial of antecedent If p, then q ¬p ——————— ¬q	invalid	If it rains, the street is wet. It does not rain. The street is not wet.
MT, Modus Tollens (denial of consequent) If p, then q ¬q ——————— ¬p	valid	If it rains, the street is wet. The street is not wet. It does not rain.

MP and MT are logically valid; AC and DA are logical invalid

(WST), which was invented by the ingenious British psychologist Peter Wason (1966). In the task, the four cards shown in Fig. 1 are presented to the experimental participants and they are instructed to verify the conditional rule ‘‘*If there is a vowel on one side of the card, then there is an even number on the other side*’’. The individuals are allowed to turn over the cards in order to verify the rule. The visible letters and numbers on the cards correspond to the four possible propositions p, ¬p, q, and ¬q. Figure 2 shows the same problem with the four possible propositions as

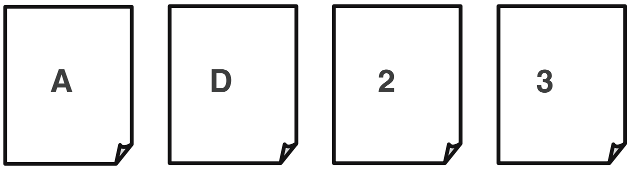

Fig. 1 The Wason selection task. The participants have to verify the rule ‘‘If there is a vowel on one side of the card, then there is an even number on the other side’’

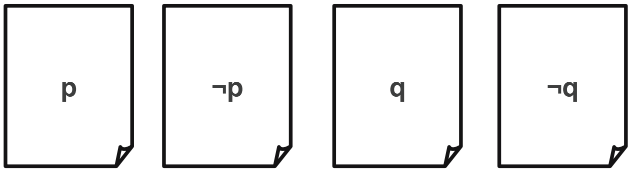

Fig. 2 The visible letters and numbers on the card correspond to the four possible propositions p, ¬p, q, and ¬q

Table 3 Relative frequency [in %] of how often people treat the four inference schemas MP, DA, AC and MT as logically valid

Experiment	*n*	MP	DA	AC	MT
Taplin (1971)	56	92	52	57	63
Evans (1977)	16	100	69	75	75
Marcus and Rips (1979)*	78	99	37	28	57
Kern et al. (1983)	72	89	28	27	41
Markovits (1990)	76	100	52	42	59

Results are partly adopted from Evans et al. (1993)

* Mean of three experiments

Table 4 Results from the study by Wason and Johnson-Laird (1972)

Card selected		#Persons (%)
A, 2	p, q	46
A	p	33
A, 2, 3	p, q, ¬q	7
A, 3	p, ¬q	4
Other combinations		10

placeholders. According to the propositional calculus of formal logic the only correct choices are p (according to the MP a q must be on the other side) and ¬q (according to the MT a ¬p must be on the other side). What human reasoners actually do is summarized in Table 4. In fact, more than half of the participants in a traditional study by Wason and Johnson-Laird (1972) turned over cards that do not help to evaluate the logical validity of the conditional rule. This experimental finding has been replicated in dozens of experiments (e.g. Griggs 1995; Feeney and Handley 2000).

There are lots of other studies that have explored conditional reasoning with disjunctions, conjunctions, negations, and counterfactual premises. Overviews of these findings can be found for instance in Manktelow (1999), Johnson-Laird and Byrne (1991).

2.1.2 Syllogistic reasoning

The second group of deductions explored in reasoning research are the syllogistic inferences. In such tasks, people are usually confronted with problems that consist of generalizations, expressed in natural language sentences such as ''All x have the feature...'', ''No x is a...'' etc. Reasoning with such expressions goes beyond the scope of the propositional calculus. The normative model here is the (first-order) predicate calculus in which the atomic sentences have a predicate with one or more terms, rather than being a single unit as in propositional logic. The new element of the predicate calculus not found in propositional logic is quantification. The universal quantor ∀ stands for ''for all'' and allows us to state that all elements in the scope of the quantifier have a certain property. The existential quantor ∃ stands for ''there is at least one'' and allows us to state that there is at least one element in the scope of the quantifier that has a certain property. Psychologists have explored only a subset of all inferences that can be made in predicate logic. These inferences are the famous Aristotelian syllogisms and in fact I would label them the ''*Aplysia*'' of reasoning research.[1] The syllogisms rely on the four ''modes'' presented in Table 5. The table also shows their formal definition in predicate logic notation.

When the four moods of Table 5 are combined in two premises, then from a formal point of view we get exactly 16 different inferences (Salmon 1983). However, the ordering of the terms in the premises (and the conclusion) are relevant in psychological research (Garnham and Oakhill 1994). Taking this into account, we can distinguish 64 different syllogisms (although there are up to 512 syllogisms if all combinations are taken into account).[2] Table 6 gives an overview.

It is probably not surprising that certain combinations of premises are very easy to deal with while others overstrain even logically highly skilled individuals. Some are so easy to solve mentally that even children under five can draw correct conclusions from them (Leevers and Harris 1999). A typical example of an easy problem is

All professors are beekeepers.
All beekeepers are athletes.

From these premises almost all individuals correctly deduce the conclusion

All professors are athletes.

Other syllogisms, however, are so difficult that even the majority of adults fail to draw a correct conclusion. A typical example of such a difficult problem is

All beekeepers are professors.
None of the athletes are beekeepers.

In psychological studies less than 10% (Johnson-Laird and Byrne 1991), are able to draw the logically valid conclusion

Some of the professors are not athletics.

In Table 7, the 64 syllogisms that can be used in the verification paradigm are presented. The table shows the results of a meta-analysis by Chater and Oaksford (1999), in which all experiments are taken into account that asked the participants to evaluate the logical validity of all 64 inferences.

2.1.3 Relational reasoning

The third class of deductive inferences uncovers some of the most important characteristics of human logical thinking. Consider, for example, the following problem:

---

[1] Aplysia is a marine snail about five inches long, which has been more extensively dissected than any other animal by biopsychologists and biologists to study the principles of learning and memory.

[2] A detailed discussion on how many syllogisms must be distinguished from a cognitive point of view can be found in Garnham and Oakhill (1994, Chapter 6). Currently, most research follows the suggestion to distinguish between 27 syllogisms (Johnson-Laird 1983).

Table 5 The four modes of a syllogism

	Mode	Notation in predicate calculus		
A	All A are B	$\forall x (A(x) \rightarrow B(x))$	Affirmative	Universal
I	Some A are B	$\exists x (A(x) \wedge B(x))$	Affirmative	Particular
E	No A are B	$\neg\exists x (A(x) \wedge B(x))$	Negative	Universal
O	Some of the A are not B	$\exists x (A(x) \wedge \neg B(x))$	Negative	Particular

Table 6 In the psychology of syllogistic reasoning it makes a difference if the premises are presented in a continuous (A–B, B–C) or discontinuous order (B–C, A–B), if the terms in the conclusion are presented in the A–C or C–A order, or if the problem is presented in one of the four traditional figures of a syllogism

Premise order	Term order in conclusion	Figure 1	Figure 2	Figure 3	Figure 4
A–B B–C	A–C	A–B B–C ——— A–C	B–A B–C ——— A–C	A–B C–B ——— A–C	B–A C–B ——— A–C
	C–A	A–B B–C ——— C–A	B–A B–C ——— C–A	A–B C–B ——— C–A	B–A C–B ——— C–A
B–C A–B	A–C	B–C A–B ——— A–C	B–C B–A ——— A–C	C–B A–B ——— A–C	C–B B–A ——— A–C
	C–A	B–C A–B ——— C–A	B–C B–A ——— C–A	C–B A–B ——— C–A	C–B B–A ——— C–A

If all these versions are combined with the four moods of a syllogism that results into 64 different problems

Ann is taller than Bert.
Bert is taller than Cath.
Does it follow that Ann is taller than Cath?

Such problems are normally called ‘‘linear syllogisms’’ or ‘‘three-term series problems’’ (Johnson-Laird 1972). Now consider the following problem:

The hammer is to the right of the pliers.
The screwdriver is to the left of the pliers.
The wrench is in front of the screwdriver.
The saw is in front of the pliers.
Is it correct that the wrench is to the left of the saw?

Here the premises describe the spatial relations between objects in two dimensions rather than in one dimension as in the previous example. Reasoners might have a few more difficulties with the second problem than with the first, but on the whole they will also have no trouble giving the right answer—yes, the wrench is to the left of the saw!

Cognitive psychologists have explored such relational reasoning for many years and identified many factors that cause reasoning difficulty. For instance, Ehrlich and Johnson-Laird (1982) gave participants the premises of a transitive inference in continuous ($A\ r_1\ B$, $B\ r_2\ C$, $C\ r_3\ D$), semi-continuous ($B\ r_2\ C$, $C\ r_3\ D$, $A\ r_1\ B$), and discontinuous ($C\ r_3\ D$, $A\ r_1\ B$, $B\ r_2\ C$) premise orders (the letter r stands for a certain relation). Participants had to infer the conclusion $A\ r_4\ D$ and the results showed that continuous order (37% error) is easier than discontinuous order (60% error), and there is no significant difference between continuous and semi-continuous (39% error) tasks. Similar results are reported, for instance, in Carreiras and Santamaría (1997) and in an experiment from our own group. In our study, there was no significant difference in the percentage of errors between continuous (39.7%) and semi-continuous (40.1%) premise orders, but both were significantly easier than the discontinuous order, which led to 50.0% errors on average. Moreover, the data on premise processing times showed that the discontinuous premise order reliably increases the processing time for the third premise, because information from all premises must be combined at this point (see Table 8, Exp. 1 from Knauff et al. 1998).

Table 7 Relative frequency of selected conclusions [in %, rounded on an integer] for all syllogisms after Chater and Oaksford (1999)

Syllogism and figure	Logically valid conclusion	Chosen conclusion				Syllogism and figure	Logically valid conclusion	Chosen conclusion			
		A	I	E	O			A	I	E	O
AA1	A(I)	90	5	0	0	IE1	N	1	1	22	16
AA2	N	58	8	1	1	IE2	N	0	0	39	30
AA3	I	57	29	0	0	IE3	N	0	1	30	33
AA4	I	75	16	1	1	IE4	N	0	1	28	44
AI1	I	0	92	3	3	EI1	O	0	5	15	66
AI2	N	0	57	3	11	EI2	O	1	1	21	52
AI3	I	1	89	1	3	EI3	O	0	6	15	48
AI4	N	0	71	0	1	EI4	O	0	2	32	27
IA1	N	0	72	0	6	IO1	N	3	4	1	30
IA2	N	13	49	3	12	IO2	N	1	5	4	37
IA3	I	2	85	1	4	IO3	N	0	9	1	29
IA4	I	0	91	1	1	IO4	N	0	5	1	44
AE1	N	0	3	59	6	OI1	N	4	6	0	35
AE2	E(O)	0	0	88	1	OI2	N	0	8	3	35
AE3	N	0	1	61	13	OI3	N	1	9	1	31
AE4	E(O)	0	3	87	2	OI4	N	3	8	2	29
EA1	E(O)	0	1	87	3	EE1	N	0	1	34	1
EA2	E(O)	0	0	89	3	EE2	N	3	3	14	3
EA3	O	0	0	64	22	EE3	N	0	0	18	3
EA4	O	1	3	61	8	EE4	N	0	3	31	1
AO1	N	1	6	1	57	EO1	N	1	8	8	23
AO2	O	0	6	3	67	EO2	N	0	13	7	11
AO3	N	0	10	0	66	EO3	N	0	0	9	28
AO4	N	0	5	3	72	EO4	N	0	5	8	12
OA1	N	0	3	3	68	OE1	N	1	0	14	5
OA2	N	0	11	5	56	OE2	N	0	8	11	16
OA3	O	0	15	3	69	OE3	O	0	5	12	18
OA4	N	1	3	6	27	OE4	N	0	19	9	14
II1	N	0	41	3	4	OO1	N	1	8	1	22
II2	N	1	42	3	3	OO2	N	0	16	5	10
II3	N	0	24	3	1	OO3	N	1	6	0	15
II4	N	0	42	0	1	OO4	N	1	4	1	25

The first column denotes the syllogism and its figure, the second denotes the logically valid conclusion, and columns 3–6 the conclusions selected by the participants. A = all, I = some, E = no, O = some ... not; N = no valid conclusion. The grey cells mark the most frequently chosen conclusions

Table 8 Premise processing times for the first, second, and third premises in the tasks with continuous, semi-continuous, and discontinuous premise order from Knauff et al. (1998)

Premise order	Premise 1	Premise 2	Premise 3
Continuous	13,0	11,2	10,9
Semi-continuous	13,6	11,0	14,4
Discontinuous	12,4	13,9	19,5

Perhaps the most exciting results in the domain of relational reasoning come from the difference between *determinate* problems, in which only a single spatial arrangement can be constructed from the premises, and *indeterminate* tasks that call for multiple spatial arrangements. The example with the tools above is a determinate problem. Only one spatial arrangement is consistent with the premises. This arrangement is

screwdriver	pliers	hammer
wrench	saw	

Now consider the following premises:

The pliers are to the right of the hammer.
The screwdriver is to the left of the pliers.
The wrench is in front of the screwdriver.
The saw is in front of the pliers.

Here two arrangements are consistent with the premises:

Arrangement 1:

hammer	screwdriver	pliers
	wrench	saw

Arrangement 2:

screwdriver	hammer	pliers
wrench		saw

The two types of problems have been introduced by Byrne and Johnson-Laird (1989) and have since resulted in dozens of further experiments. The fascination for the tasks is primarily based on the fact that both types of problems lead to the same conclusion (the wrench is to the left of the saw) but reasoners have more difficulties with the second problem than with the first one (e.g., Boudreau and Pigeau 2001; Carreiras and Santamaria 1997; Roberts 2000; Schaeken et al. 1998; Schaeken and Johnson-Laird 2000; Schaeken et al. 1996a, b; Vandierendonck and de Vooght 1997).

One very interesting feature of mental reasoning with spatial relations has been uncovered by the work of our group. In our experiments we used indeterminate problems that called for three, five, or nine possible arrangements. The results showed that, whenever a reasoning problem corresponds to multiple arrangements, reasoners prefer one of them and that individuals consistently prefer the same arrangement. Based on the fact that this arrangement seems to be the first one that is cognitively available, it follows that this will favor certain inferences before others. We tested this prediction in an experiment in which the participants did not generate but rather verified conclusions for the same reasoning problems as before. The results corroborated our predictions: relationships that conformed to the preferred arrangement were verified faster than other possible relationships, and they were also more often correctly verified than other possible relationships. Apparently, our experimental participants focused only on a subset of possible arrangements and ignored others that are also consistent with the premises. Not surprisingly, this led them to erroneous conclusions (Rauh et al. 2005).

3 Cognitive theories of human logical reasoning

In the special case of relational reasoning difficulties can be explained by the fact that some possible arrangements are favored over others. However, many people are not very good in logical reasoning in general. So how can the difficulties many people have with logic be explained? The psychological literature provides numerous potential explanations. For instance, it seems that cognitive accounts of conditional and syllogistic reasoning must incorporate theories that focus on the misunderstanding of premises (e.g. Begg 1987; Chapman and Chapman 1959; Revlis 1975), the surface structure of premises that biases reasoners towards certain conclusions (e.g. Begg and Denny 1969; Woodworth and Sells 1935), the ease to match a presented conclusion with a mental calculation (Wethrick and Gilhooly 1990), and the application of heuristics that save cognitive resources (overview in Evans 1989). All such ‘‘mini’’ theories can be labeled ‘‘error theories’’ of reasoning, as they try to explain reasoning difficulty and the cause of error. They provide answers to the first question of reasoning research as they are concerned with the errors that are observable in the *performance* of logical reasoners. They do not say much about the second question, what cognitive processes underlie our logical *competence*. So, which cognitive mechanisms enable us to reason logically? Here there is a long-standing debate between the two main ‘‘schools’’ of human thinking. Their aim is much more ambitious than that of the mini theories as they seek to explain the mental computations that enable us to reason logically. The two theories differ in the postulated underlying mental representations and the computational processes that work on these representations. In one theory, it is believed that people think deductively by applying mental rules, which are similar to rules in computer programs (Braine 1978, 1990; Braine and O’Brien 1998; Rips 1994; O’Brien 2004). In the other theory, deductive reasoning is conceived as a process in which the reasoner constructs, inspects, and manipulates mental models (Johnson-Laird 1983; Johnson-Laird and Byrne 1991; Johnson-Laird 2006). The rule-based theory is a syntactic theory of reasoning, as it is based on the form of the argument only, whereas the mental models theory is a semantic theory, because it is based on the meaning (the interpretation) of the premises.

The differences between the two theories of reasoning can be best explained in the domain of relational reasoning. For a proponent of the rule-based school of reasoning, the inference in the ‘‘taller than’’ problem yields a transitive conclusion, in which its validity is dependent on the missing premise:

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

This camp of reasoning theories is primarily represented by the work of Rips (1994) and Braine and O'Brien (1998). Generally speaking, these authors claim that reasoners rely on formal rules of inference akin to those of formal logic, and that inference is a process of proof in which the rules are applied to mental ''sentences''. The formal rules govern sentential connectives such as ''if'' and quantifiers such as ''any'', and they can account for relational inferences when they are supplemented with axioms governing transitivity. The rules are represented in specific areas of the human brain and the sequence of applied rules results in a mental proof or derivation that is seen as analogous to the proofs of formal logic (Rips 1994). Errors in this account do occur because the limited capacities of the cognitive system prevent us from applying all rules in a logically correct fashion, because the necessary rules are not available, or because conflicts between different rules are resolved in an inappropriate way. Many theories also explain errors by means of mistakes that appear when the sentential presentation of premises must be translated into a logical representation (Braine and O'Brien 1998).

Supporters of the model theory offer a completely different concept of how humans reason logically. They conceptualize reasoning as a process in which a mental structure—the mental model—is constructed that is consistent with the premises. A conclusion is true if it holds in all possible structures (models) compatible with the premises (Johnson-Laird 1983, 2001; Johnson-Laird and Byrne 1991). In fact, each spatial ''arrangement'' in the description of relational reasoning is one of such ''models''. In contrast to the rule-based approaches, here reasoning relies on a semantic interpretation of the premises, involving the construction of an integrated mental representation of the information given in the premises. These integrated representations are models in the strict logical sense (e.g. Hodges 1997). It is a mental representation that captures what is common to all the different ways in which the reasoner can interpret the premises. Cognitively speaking, they represent on a small scale how ''reality'' could be, according to the premises of a reasoning problem. The model theory distinguishes between three different mental operations. In the construction phase, reasoners construct the mental model that reflects the information from the premises. In the inspection phase, this model is inspected to find new information that is not explicitly given in the premises. In the variation phase, reasoners try to construct alternative models from the premises that refute the putative conclusion. If no such model is found, the putative conclusion is considered true (Johnson-Laird and Byrne 1991; Knauff et al. 1998).

4 Logical reasoning after brain damage

It is one thing to explore the functional dependencies between the logical characteristics of a reasoning problem and the (in)accuracy of mental reasoning. Another question is how logical reasoning is realized in our brains. Many people suffer (like Nietzsche) from brain injuries after infections, tumors, strokes, or accidents. One possible way in which to investigate how reasoning processes are neurally implemented is to ask these patients to help science to explore what reasoning abilities break down if parts of the neural ''hardware'' are destroyed. It is a chief advantage of the different reasoning theories that they lead to different neural hypotheses saying what brain damage should result in what logical deficit.

For the large group of error theories the predictions are of course very heterogeneous and neuroscientific methods can, if anything, only partially distinguish between all these ''mini theories'' (Knauff 2006). The rule-based and model-based theories, however, come up with clearly distinguishable assumptions concerning the neural correlates of reasoning. On the broadest level they make different predictions regarding the involvement of the two cortical hemispheres, relying on a simple—and only in part correct!—assumption of lateralization, or 'job sharing', between the ''abstract'' and ''language-related'' left hemisphere and the ''holistic'' and ''visuo-spatial'' right hemisphere (see Springer and Deutsch 1981).

The involvement of the two hemispheres has been a major concern of patient studies on logical reasoning. The idea of such studies is that damage to a specific brain area or hemisphere should result in defective information processing and thus in a cognitive disorder. Selective impairments of cognitive tasks following brain damage therefore are interpreted in terms of the loss of particular processing components (Plaut 1995). Special cases are the so-called double-dissociations in which a damage to region X produces the deficit x but not y, while damage to the region Y results in deficit y but not x. This method was introduced by Teuber (1955) to identify when lesions do have specific effects on distinct cognitive functions. The method usually used by researchers to measure the reasoning abilities of a brain damaged person is to confront patients with the various kinds of reasoning problems usually used with healthy people, and subsequently to compare performance between the patients and a healthy control group. In this way we can see, for instance, that damage to region or hemisphere X produces a deficit in solving syllogisms but not in relational reasoning, whereas damage to the region or hemisphere Y results in a deficit in relational inferences but not in syllogistic reasoning.

Early patient studies seem to emphasize the role of the left hemisphere for logical reasoning. For instance, Golding studied the performance of individuals that suffered from damage either to the left (left-hemispheric patients, LHP) or to the right cerebral hemisphere (right-hemispheric patients, RHP) on the WST (Golding 1981). She formulated an interesting (and complex) hypothesis concerning interference between visuo-spatial and verbal processes during reasoning. Here is the central quote from the paper:

> It was postulated that visual skills known to be lateralised to the right hemisphere inhibited the verbal skills of inference, thought to be lateralised to the left hemisphere, thus preventing insight into the problem (Golding 1981, p. 32).

To test this hypothesis Golding embedded the WST into a ‘‘perceptual classification’’ task in which the patients saw objects from a typical or an unconventional angle. Warrington and Taylor (1973) have shown that RHP have difficulty recognizing an object that has been photographed from above, whereas they have no problems if the object is shown from the side. Thus, Golding assumed that RHP should have a deficit in perceptual classification that results in a more accurate performance on the WST only when the materials are not viewed at an oblique angle. Table 9 summarizes Golding's main findings for the RHP and LHP. The controls showed the usual pattern with 55% for p q, 30% for p and the rest for all other combinations.

Golding's experiment is unusual, because a straightforward interpretation of lesions is that a lesion in an area that is responsible for a certain cognitive task should impede the performance on that task. However, Golding formulated an interference hypothesis, which predicted that right hemisphere brain lesions would *facilitate* insight into the logical problem. This hypothesis was upheld since patients who had a specific perceptual classification deficit solved the problem.

A more direct test of the left-hemisphere hypothesis of reasoning has been performed by Deglin and Kinsbourne (1996). They studied syllogistic reasoning with psychiatric patients who were recovering from transitory ictal suppression of one hemisphere by electroconvulsive therapy (ECT; this simulates a temporal lesion). The premises had a familiar or unfamiliar content and they were true or false. When the right hemisphere was suppressed, the participants tended to perform deductive inferences even when the factual answer was obviously false. While their left hemisphere was suppressed, the same participants used their prior knowledge and if the content was unfamiliar they completely refused to answer.

Table 9 Summary of findings from Golding (1981)

Card selected	RHP	LHP
p, q	40	15
p	0	50
p, q, ¬q	20	0
p, ¬q	30	5
Other combinations	10	20

Patient studies on relational reasoning have been reported by Caramazza et al. (1976) and Read (1981). Caramazza et al. (1976) presented relational premises such as ‘‘Mike is taller than George’’ to brain-damaged patients. After reading the statements they had to answer either a congruent (‘‘Who is taller?’’) or incongruent (‘‘Who is shorter?’’) question. The LHP showed impaired performance in all problems regardless of whether they were consistent or inconsistent. RHP, in contrast, showed impaired performance only in the incongruent problems. Read (1981) used two relational premises and asked patients who suffered from temporal-lobectomy to generate a conclusion from these statements. The LHP again performed worse than the RHP, but, interestingly, the RHP were more impaired with the incongruent conclusions.

5 Logical reasoning in the intact brain

Patient studies have been frequently interpreted in favor of a left hemispheric prevalence in logical reasoning (Goel et al. 1997, 1998). However, more sophisticated experimental methods show that this is probably just one side of the coin. A more precise method to explore the cortical substrates of cognition are brain imaging methods. Although there are several types of imaging, most of the research on logical reasoning has been performed by means of functional magnetic resonance imaging (fMRI). This technique works by measuring the local increase in oxygen delivery in the activated cerebral tissue while the subject is engaged with the task. The method takes advantage of the fact that this local increase in oxygen delivery is related to the cognitive processes that are involved in solving the problem. With this method the localization of brain activity can be very precise. To understand the brain imaging studies on deductive reasoning a few words about the main functions of the four lobes of the brain are needed. For more detailed anatomical and functional description of the human brain the interested reader is directed to the pertinent textbooks, for instance by Kandel et al. (2000).

Broadly speaking, the *occipital lobes* in the back of the brain process visual information. The occipital cortex can be divided into the primary visual cortex, also referred to as the striate cortex or functionally as V1, and to the visual association areas V2, V3, V4. The primary visual cortex receives visual input from the retina, while the association areas are responsible for the further processing of visual and spatial information. In the present context, it is essential that the visual cortices have been frequently related to visual mental imagery. For instance, patients who are blind in one side of the visual field are also unaware of objects on that side when imagining a visual scene. If the patient turned the mental image around so that they had to ‘‘look’’ at the image from the opposite direction, they reported objects on the other side and ignored those which they had previously reported ‘‘seeing’’ (Mellet et al. 1998). Consequently, one of the central research issues on imagery is whether the visual cortices are activated by visual mental imagery. Indeed, this assumption is supported by a series of studies by Kosslyn and his colleagues, who found increased blood flow in early visual areas during mental imagery of letters (Kosslyn et al. 1993) and objects in different sizes (Kosslyn et al. 1997).

The *temporal lobes*, which are located one on each side of the brain at about the level of the ears, are involved in the processing of auditory signals and hearing. They are also involved in high-level auditory processing including speech processing. Particularly, Wernicke’s area plays a key role in language understanding.

The *parietal lobes* are positioned above the temporal lobes. They are involved in diverse functions, but one of the most important jobs, in particular of the posterior (back) part, is to combine information from different sensory modalities to form a cognitive representation of space.

The central sulcus separates the parietal lobes from the frontal lobes, which lie in the front of the brain. The *frontal lobes* are involved in planning, problem solving, selective attention, and many other higher cognitive functions. The anterior (front) portion of the frontal lobe is called the prefrontal cortex (PFC). It is involved in executive processes in working memory and typically implicated when several pieces of information need to be monitored and manipulated (see Fig. 3).

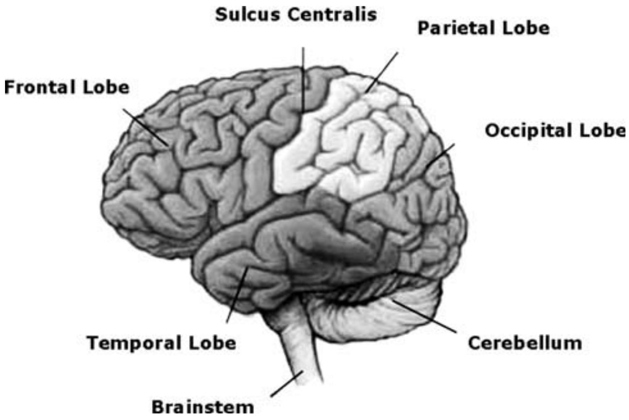

Fig. 3 The four lobes of the human brain from the side (lateral) view

6 Predictions of the reasoning theories and the activation of cortical regions

No researcher would deny that the frontal cortex plays a central role in logical reasoning, but rule-based and model-based reasoning theories make different predictions about the involvement of other brain areas. Since model-based reasoning theories suggest that reasoning is a visuo-spatial process, these theorists view the parietal and occipital cortices as essentials brain structure for logical reasoning. On the other hand, it is clear that rule-based theories must assume that language-related areas in the temporal cortices are active during reasoning, probably with a left hemispheric prevalence (Goel et al. 1997).

In the last decade, a few groups have conducted brain imaging studies on all three classes of deductive reasoning problems. Knauff et al. (2002) studied conditional reasoning problems by presenting premises such as ‘‘If the teacher is in love, then he likes pizza’’ to the participants. In half of the problems the second premise was ‘‘The teacher is in love’’ and the participants had to conclude (by MP) ‘‘The teacher likes pizza’’. In the other half of problems the second premise was ‘‘The teacher does not like pizza’’ and the participants had to conclude (by MT) ‘‘The teacher is not in love.’’ Interestingly, both sorts of problems activated a bilateral occipito–parietal–frontal network, including parts of the prefrontal cortex, the parietal cortex, and the visual association cortex.

Similar findings have been reported from a study on syllogistic reasoning. Goel et al. (2000) used problems with semantic content (e.g. ‘All apples are red; all red fruit are sweet; therefore all apples are sweet’) and without semantic content (e.g. ‘All A are B; all B are C; therefore all A are C). They found evidence for the engagement of both linguistic and visuo-spatial systems.

In a study of relational reasoning by Knauff et al. (2000) such problems activated similar brain areas as the conditional problems did. However, the activity in visual association areas was even higher than during conditional reasoning. The same findings are reported in Goel and Dolan (2001) who used problems with a spatial content. Their premises were either concrete (e.g. ‘‘The apples are in the barrel; the barrel is in the barn; therefore the apples are in the barn’’) or abstract (e.g. ‘‘A are in B; B is in C; therefore A is in C’’) and the authors reported that all problems activated a similar bilateral occipito–parietal network regardless of whether they were concrete or abstract. So the concreteness of the content of the problem is apparently not responsible for the involvement of visuo-spatial brain areas.

7 Background knowledge and belief biases

Many behavioral studies have shown that prior knowledge can significantly influence how efficiently a reasoning problem is solved. Technically speaking, the abstract (logical) truth value of an inference can be the same as the truth value of our prior knowledge, or the formal truth value can conflict with the truth value of the prior knowledge. In the latter case inferences result in more errors and take longer to evaluate or generate. If an inference generated by a person is biased towards the truth value of the prior knowledge or even overwritten by it, this is called ‘‘belief bias’’ (Evans 1989).

Some patient studies have therefore explored the effects of brain injuries on reasoning with concrete and abstract materials. Their findings agree with the brain imaging study by Goel et al. (2000) in which evidence for the engagement of both linguistic and spatial systems has been found. Reasoning with a semantic content activated a left-hemispheric temporal system, whereas problems without semantic content activated an occipito–parietal network distributed over both hemispheres. Goel and Dolan (2003) brought logic and belief into conflict and found evidence for the engagement of a left temporal lobe system during belief-based reasoning and a bilateral parietal lobe system during belief-neutral reasoning. Activation of right prefrontal cortex was found when the participants inhibited a response associated with belief-bias and correctly completed a logical task. When logical reasoning, in contrast, was overwritten by a belief-bias, there was engagement of ventral medial prefrontal cortex, a region implicated in affective processing.

8 Do we reasoning logically by using mental images?

How well a formal inference fits with our background knowledge is only one aspect of the content of a reasoning problem. Another aspect is how easy it is to visualize the matter of the problem. Cognitive psychologists and psychological laymen often think that we reason by using ‘‘mental pictures’’ in the ‘‘mind’s eye’’ and that we can ‘‘look’’ at these pictures to find new, not explicitly given information. Identifying logical inference with visual images might appear bizarre to a logician. In psychology, however, a very vigorous debate revolves exactly around this issue. In this part of the paper I want to describe my personal view on this problem and I will report ample evidence in support of this view. The results mainly focus on relational reasoning as these inferences, if any, have the strongest link to what we call a visual mental image. Of the many psychological definitions of visual mental imagery I prefer to define it as the inspection and manipulation of visual information that comes not from perception, but from memory, or from another non-visual external stimulus, such as the sentential premises of a logical problem.

A pioneering study on visual imagery in relational reasoning was carried out by De Soto et al. (1965), who argued that reasoners represent the entities of a relational reasoning problem as a mental image and then ‘‘read off’’ the conclusion by inspecting the image. Huttenlocher (1968) also argued that reasoners imagine an analogous physical arrangement of objects in order to cope with reasoning problems. Moreover, other authors report that reasoning is easier with problems that are easy to envisage than with problems that are hard to envisage (e.g. Shaver et al. 1975; Clement and Falmagne 1986). However, several studies have failed to detect any effect of imageability on reasoning. Johnson-Laird et al. (1989), for instance, examined reasoning with relations that differed in imageability—equal in height, in the same place as, and related to (in the sense of kinship)—and did not find any effect on reasoning accuracy. Newstead et al. (1986) reported a similar result, and Sternberg (1980) did not find any reliable correlation between scores on the imageability items of IQ-tests and reasoning ability. Overall, results from behavioral studies are thus far inconclusive.

In the past few years our own laboratory has been investigating the role of visual mental images in logical reasoning. Within all of these projects the aim has been to unify formal logic with experimental methods from cognitive psychology, the computational reconstruction of the obtained results, and the understanding of the underlying brain processes. Here I only report the findings from the experiments using behavioral methods and functional brain imaging. The computational work is discussed in the extended version of this paper (Knauff 2007).

To understand the following it is necessary to go back to the functional organization of the brain. As already mentioned, the partial and the occipital cortices are responsible for different cognitive functions. While the parietal cortex (or at least part of it) is responsible for the processing of spatial information from different modalities, the occipital lobes are modality-specific and responsible for the processing of visual information and have been frequently related to visual mental imagery (Kosslyn 1994). The parietal cortex, in contrast, has often been identified with more abstract spatially organized mental representations (Knauff 2006). They come very much closer to the logical meaning of a ‘‘model’’ as they capture what is true in *all* possible interpretations of the premises. A visual image must be entirely determined in all respects, and thus represents *one* interpretation of the premises, while a model in this sense can be underspecified in many respects as long as it represents the meaning of the premises. It is thus more abstract than an image.

8.1 Logical reasoning and visuo-spatial cortices

Some of the studies on logical reasoning reported above found little evidence that visual brain areas (in occipital cortex) are involved in reasoning (Goel et al. 1997, 1998). More recently, however, an increasing number of studies reported activity in primary and secondary visual areas accompanying engagement in logical reasoning problems. This, for instance, was the case in a study by Goel et al. (2000) in which the volunteers had to solve different kinds of relational inferences. Moreover, Knauff et al. (2000) studied relational and conditional inferences that were presented acoustically via headphones to the participants (to avoid a confounding of mental imagery and visual perception). In this study, both types of reasoning problems resulted in activity in a bilateral occipito–parietal–frontal network distributed over parts of the prefrontal cortex, the inferior and superior parietal cortex, the precuneus, and the visual association cortex. Similar results have been reported in Ruff et al. (2003). Here the brain activity of participants was scanned, and their visuo-spatial ability was also measured, with a well-known subset of tasks from an intelligence inventory. Interestingly, the brain activation was significantly modulated by the participants' visuo-spatial skill. The higher the participants' visuo-spatial skill, the better their reasoning performance, and the less activation was present in visual association areas during reasoning. This pattern agrees with recent findings on the effects of skill level on neuronal activity. Accordingly, the reasoning problems seemed to have placed less demand on the visuo-spatial processing resources of participants with high skill levels, so that less activity in the relevant cortical regions was required (Ruff et al. 2003).

A key disadvantage of the reported studies is that they were not designed to determine the exact role of visual images in reasoning and thus examined the brain activation during the whole reasoning process as a single block (e.g. Knauff et al. 2002) or just compared the neuronal processes during the conclusion of the reasoning problem with the presentation of irrelevant control sentences (e.g. Goel and Dolan 2001). In both paradigms it is impossible to determine whether the activity in occipital brain areas (indicating the employment of visual mental imagery) is associated with the processing of premises, the maintenance of problem information in working memory, or with the actual reasoning process. Reasoning-related processes during different stages of problem processing and other cognitive processes are inseparable. To overcome these disadvantages, our group carried out an fMRI study to disentangle the neuro-cognitive subprocesses underlying the different stages in the reasoning process, and at the same time to avoid potential conflicts in the previous studies on the neuronal basis of imagery and reasoning (Fangmeier et al. 2006). In the study, we scanned the brains of our volunteers while they solved relational reasoning problems. To avoid the need to read the premises and conclusions we replaced the sentences with graphical arrangements describing the spatial relations between three objects. Participants had to decide whether the conclusion logically (necessarily) followed from a pair of premises. The processing of the first premise, the second premise, and the conclusion, was time-locked so we could examine the brain activity elicited by different stages of the reasoning process.

The results of this study are illustrated in Fig. 4. The darker a region in the image, the more cortical activity was measured. As can be seen from the foci of activation, we identified three distinct patterns of neuronal activation associated with three stages of the reasoning process. During the presentation of the first premise, we found two large bilateral clusters of activation in the vision-related occipito-temporal cortex (see Fig. 4a). Then the participants needed to unify the second premise with the information from the first premise in order to construct an integrated representation of both premises. During this stage the two clusters in the occipito-temporal cortex and an additional cluster in the anterior prefrontal cortex were activated (see Fig. 4b). In the third stage participants had to inspect and manipulate this representation to draw a putative conclusion and to compare this conclusion with the displayed conclusion. Crucially, this stage activated spatial areas of posterior parietal cortex, whereas vision-related activity in occipital cortex completely disappeared (see Fig. 4c).

8.2 The visual-impedance effect

So far we were only concerned with reasoning problems that invoke the tendency to construct visual images. But what happens if the premises of a logical reasoning problem do not bias the reasoner to construct visual images? For example, they could straightforwardly lead to the spatial representation pertinent to reasoning without the phenomenal experience of an image. To answer these questions we designed a study which combined behavioral and neuroimaging methods. In this study, we systematically investigated the engagement of mental imagery and the related brain areas during reasoning (Knauff et al. 2003). We speculated that only premises that are easy to visualize spontaneously elicit visual images, while other premises do not push reasoners to construct visual images. Some premises are probably more difficult to visualize and, therefore, no visual images are pressed into service during reasoning. Is reasoning easier or more difficult with these relations and does it activate different brain areas? In Knauff and Johnson-Laird (2002; see also Knauff and May 2006) we empirically identified four sorts of relations: (1) visuo-spatial relations that are easy to envisage visually

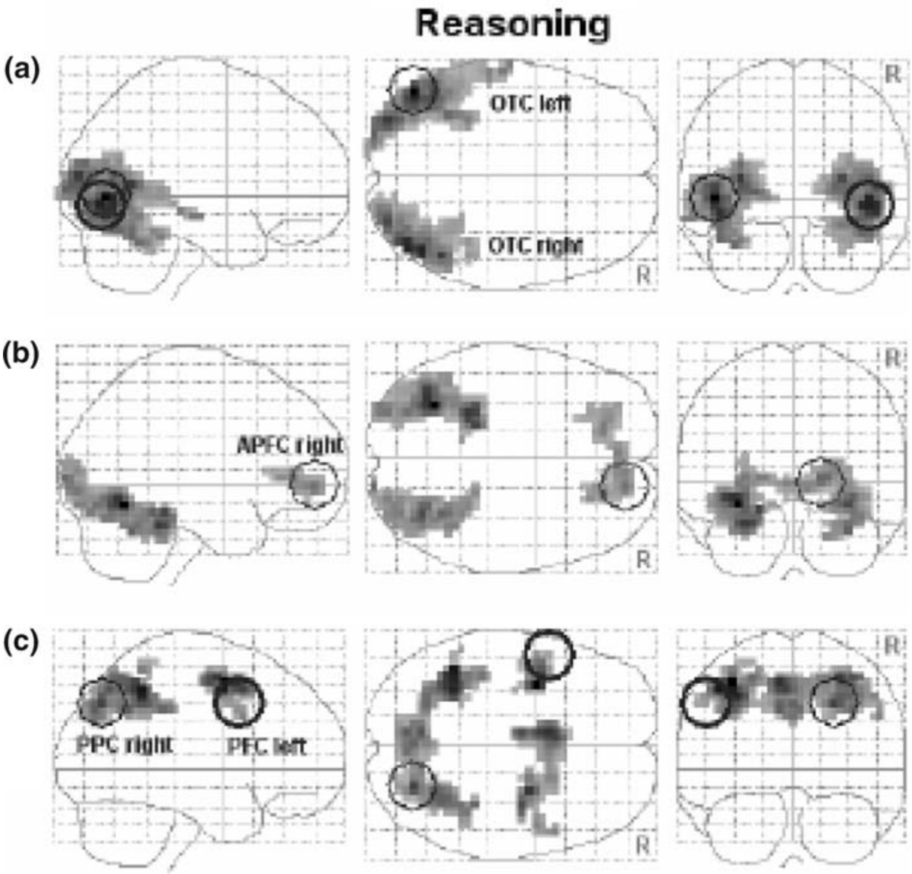

Fig. 4 Results from a study by Fangmeier et al. (2006). The brain is presented from three different perspectives. Left column: from the side as if vertically cut through at about the position of the eyes; middle column: horizontal as if horizontally cut through in parallel to the axis of the eyebrows; right column: transverse as if vertically cut through in parallel to the axis between the ears. The images represent differentially activated brain areas during the three steps of a relational inference. A darker region in the image indicates that more cortical activity was measured. Row (**a**) shows the activation during the premise processing phase, row (**b**) during the integration phase, and row (**c**) during the validation phase

and spatially, (2) visual relations that are easy to envisage visually but hard to envisage spatially, (3) spatial relations that are hard to envisage visually but easy to envisage spatially, and (4) control relations that are hard to envisage either visually or spatially. We started by conducting a series of behavioral experiments in which participants solved linear syllogisms with these relations (Knauff and Johnson-Laird 2002). Apparently, the orthodox imagery theory would predict an advantage of visual and probably visuo-spatial relations. Our prediction, however, was that relations that elicit visual images containing details that are irrelevant to an inference should impede the process of reasoning, because the information pertinent to reasoning must be retrieved from the image. In contrast, relations that directly yield a spatial model without the ‘‘detour’’ of a visual image should speed up the process of reasoning in comparison with relations that elicit images. Our findings supported these predictions: in three experiments we found relations that are easy to visualize impaired reasoning. Reasoners were significantly slower with these relations than with the other sorts of relations. In fact, the spatial relations were the quickest, while the visual relations were the slowest. We called this the *visual-impedance effect* (Knauff and Johnson-Laird 2002). We then performed an fMRI study using the same sorts of problems. As can be seen in Fig. 5, all types of reasoning problems again evoked activity in the parietal cortices. This activity seems to be a ‘‘default mode’’ of brain functioning during reasoning, perhaps because individuals have the facility to construct a spatial representation from all sorts of relations. Such a representation will be spatial in form for visuo-spatial and spatial relations, and, as long-standing evidence suggests, even relations such as ‘‘smarter’’ are also likely to elicit spatial representations (see, e.g., Johnson-Laird 1998; De Soto et al. 1965). However, only problems based on visual relations also activated areas of the visual cortices. Presumably, in the case of visual relations, reasoners cannot suppress a spontaneous visual image. Its construc-

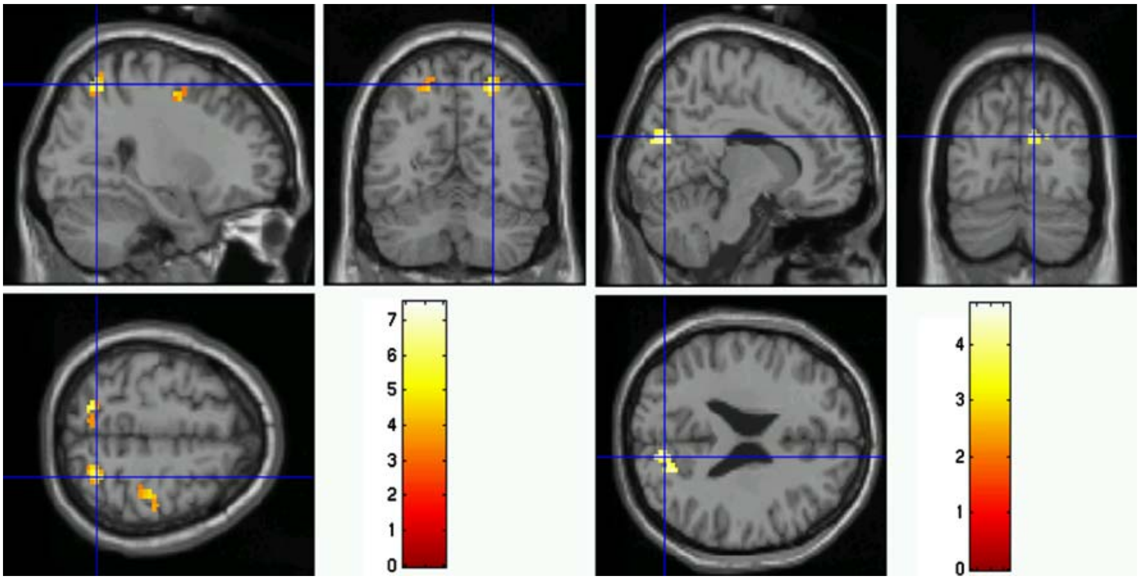

Fig. 5 Images representing differentially activated brain areas during reasoning. The location of the highlighted areas indicates that the spatial information from reasoning problems is mapped to areas of the brain responsible for the multimodal integration of space from perception and working memory. The three images on the right-hand side show the activity in the back of the brain suggesting that individuals naturally construct visual images, if the reasoning problem is easy to visualize (from: Knauff et al. 2003; see text for details)

tion calls for additional activity in visual cortices and retards the construction of a spatial representation that is essential for the inferential process.

9 Conclusions: is the debate between models and rules obsolete?

From the findings reported here we could get the impression that the distinction between model-based and rule-based theories of logical reasoning is obsolete. If we try to maintain the classical distinction, we must admit that we only very seldom use a single deterministic strategy for solving logical reasoning problems. Sometimes the way we reason is logically analogous to the proofs of formal logic, sometimes we think logically by using models in the strict logical sense, and sometimes we use visual mental images. Some colleagues draw the conclusion from these results that content is primarily responsible for the application of one or the other reasoning strategy. Goel (2003) for instance argues that a frontal-temporal system processes familiar, conceptually coherent material while a parietal system processes unfamiliar, nonconceptual, or incoherent material. The former engages language-related rule-based computations while the latter engages visuo-spatial model-based processing systems. There is much in this position with which to agree. In particular, it may be granted that different processes are involved in mental reasoning; also that the visuo-spatial subsystems are involved in reasoning. What is doubtable is the conclusion these authors draw from their findings. In particular, the conclusion that the frontal-temporal system is more ‘‘basic’’, and effortlessly engaged, while the parietal system is effortfully engaged only when the frontal-temporal route is blocked due to a lack of familiar content (Goel 2003). From the current point of view, the question of how mental logical reasoning is implemented in the human brain is a question of *formal* reasoning. If, as many findings suggest, the right hemisphere is involved in ‘‘abstract reasoning’’, then this hemisphere, if any, is the more ‘‘basic’’ for reasoning. It seems to be responsible for all operations that are compatible with what, according to most logicians, logic is about. The left hemisphere, in contrast, is ‘‘only’’ occupied with the processing of ‘‘content’’, that, according to most logicians, logic is not about. From this perspective the language-based system corresponds to more knowledge-based processing while the parietal model-based system corresponds to the ‘‘logical’’ system.

Another way to think about the models versus rules debate is to see it as based on confused distinctions. Of course, rules are needed to construct models, models can be in a language-based format (e.g. Polk and Newell 1995), and there are well-known emulations of the one kind of description in the other (Stenning and Oberlander 1995). More generally, Stenning and collaborators have shown the abstract equivalence of all the main psychological competence theories of human reasoning (Stenning and Oberlander 1995; Stenning 1998). In the long version of this paper I therefore try to provide a new framework to differentiate rule-based and model-based reasoning by introducing distinctions that have genuine content both empirically and theoretically (Knauff 2007). I propose that the true contrasts between different reasoning theories lie in two other details:

(1) whether the parietal (and occipital) cortex is involved in reasoning, and
(2) whether integrated representations of the premises are constructed and inspected.

In fact, most of the experiments have shown that the parietal cortices are involved in reasoning and that reasoners indeed construct integrated representations of the premises and inspect these representations in order to find new information that is not explicitly given in the premises. Both criteria agree with what we can call ‘‘model-based’’ reasoning in the literal sense and that has nothing to do with the question whether or not verbal or language-based processes are involved in the cognitive processes underlying human logical reasoning. The integrated spatial representations in the parietal cortex seem to be ‘‘models’’ in the strictest logical sense. The representation’s parts correspond to the parts of what it represents, and its structure corresponds to the structure of the reasoning problem. Such a new framework might make the models versus rules debate that has given a context for the present work obsolete.

Acknowledgements I am very thankful to Johan van Benthem, Wilfrid Hodges, Marian Counihan, and an anonymous reviewer for their very helpful und insightful comments and suggestions. I am also thankful to the German National Research Foundation (DFG) for supporting my work by a Heisenberg Award.

References

Begg I (1987) Some. Can J Psychol 41:62–73
Begg I, Denny J (1969) Empirical reconciliation of atmosphere and conversion interpretations of syllogistic reasoning. Quart J Exp Psychol 81:351–354
Boudreau G, Pigeau R (2001) The mental representation and processes of spatial deductive reasoning with diagrams and sentences. Int J Psychol 36:42–52
Braine MDS (1978) On the relation between the natural logic of reasoning and standard logic. Psychol Rev 85:1–21
Braine MDS (1990) The ‘‘natural logic’’ approach to reasoning. In: Overton WF (ed) Reasoning, necessity and logic: developmental perspectives. Erlbaum, Hillsdale, NJ
Braine MDS, O’Brien DP (eds) (1998) Mental logic. Lawrence Erlbaum Associates, Mahwah, NJ
Byrne RMJ, Johnson-Laird PN (1989) Spatial reasoning. J Memory Lang 28:564–575
Caramazza A, Gordon J, Zurif EB, DeLuca D (1976) Right-hemispheric damage and verbal problem solving behavior. Brain Lang 3:41–46
Carreiras M, Santamaría C (1997) Reasoning about relations: spatial and nonspatial problems. Thinking Reasoning 3:191–208
Chapman LJ, Chapman AP (1959) Atmosphere effect re-examined. J Exp Psychol 58:220–226
Chater N, Oaksford M (1999) The probability heuristic model of syllogistic reasoning. Cogn Psychol 38(2):191–258
Clement CA, Falmagne RJ (1986) Logical reasoning, world knowledge, and mental imagery: interconnections in cognitive processes. Memory Cogn 14:299–307
Deglin VL, Kinsbourne M (1996) Divergent thinking styles of the hemispheres: how syllogisms are solved during transitory hemisphere suppression. Brain Cogn 31:285–307
De Soto CB, London M, Handel MS (1965) Social reasoning and spatial paralogic. J Personal Soc Psychol 2:513–521
Ehrlich K, Johnson-Laird PN (1982) Spatial descriptions and referential continuity. J Verbal Learn Verbal Behav 21:296–306
Evans J St BT (1977) Linguistic factors in reasoning. Quart J Exp Psychol 29:297–306
Evans J St BT (1989) Bias in human reasoning. Lawrence Erlbaum, Hove, UK
Evans J St BT, Newstead SE, Byrne RMJ (1993) Human reasoning. Lawrence Erlbaum Associates, Hove, UK
Fangmeier T, Knauff M, Ruff CC, Sloutsky V (2006) fMRI evidence for a three-stage-model of relational reasoning. J Cogn Neurosci 18(3):320–334
Feeney A, Handley SJ (2000) The suppression of q card selections: evidence for deductive inference in Wason’s Selection Task. Quart J Exp Psychol 53A(4):1224–1242
Fishman RS (2002) Thus saw Zarathustra: the eyes and brain of Friedrich Nietzsche. In: Abstracts of the 7th Annual Meeting of the International Society for the History of the Neurosciences (ISHN). http://www.bri.ucla.edu/nha/ishn/abs2002.htm
Fodor JA (1975) The language of thought. Crowell, New York
Garnham A, Oakhill J (1994) Thinking and reasoning. Blackwell, Oxford, UK
Goel V (2005) Cognitive neuroscience of deductive reasoning. In: Holyoak K, Morrison R (eds) Cambridge handbook of thinking and reasoning. Cambridge University Press, Cambridge, pp 475–492
Goel V (2003) Evidence for dual neural pathways for syllogistic reasoning. Psychologica 32:301–309
Goel V, Büchel C, Frith C, Dolan RJ (2000) Dissociation of mechanisms underlying syllogistic reasoning. NeuroImage 12:504–514
Goel V, Dolan RJ (2001) Functional neuroanatomy of three-term relational reasoning. Neuropsychologia 39:901–909
Goel V, Dolan RJ (2003) Explaining modulation of reasoning by belief. Cognition 87:B11–B22
Goel V, Gold B, Kapur S, Houle S (1997) The seats of reason? An imaging study of deductive and inductive reasoning. NeuroReport 8:1305–1310
Goel V, Gold B, Kapur S, Houle S (1998) Neuroanatomical correlates of human reasoning. J Cogn Neurosci 10:293–302
Golding E (1981) The effect of unilateral brain lesion on reasoning. Cortex 17:31–40
Grafman J, Goel V (2002) Neural basis of reasoning. Encyclopedia of cognitive science. Macmillan, London
Griggs RA (1995) The effects of rule clarification, decision justification and selection instruction on Wason’s abstract selection task. In: Newstead SE, Evans J St BT (eds) Perspectives on thinking and reasoning. Essays in honour of Peter Wason. Lawrence Erlbaum Associates, Hove, UK
Hodges W (1997) A shorter model theory. Cambridge University Press, Cambridge, CA
Huttenlocher J (1968) Constructing spatial images: a strategy in reasoning. Psychol Rev 75:550–560
Johnson-Laird PN (1972) The three-term series problem. Cognition 1:57–82

Johnson-Laird PN (1983) Mental models. Cambridge University Press, Cambridge

Johnson-Laird PN (1998) Imagery, visualization, and thinking. In: Hochberg J (ed) Perception and cognition at century's end. Academic Press, San Diego, CA, pp 441–467

Johnson-Laird PN (2001) Mental models and deduction. Trends Cogn Sci 5:434–442

Johnson-Laird PN (2006) How we reason. Oxford University Press, Oxford

Johnson-Laird PN, Byrne RMJ (1991) Deduction. Erlbaum, Hove (UK)

Johnson-Laird PN, Byrne RMJ, Tabossi P (1989). Reasoning by model: the case of multiple quantification. Psychol Rev 96:658–673

Kandel ER, Schwartz JH, Jessell TM (2000) Principles of neural science. McGraw-Hill, New York

Kern LH, Mirels HL, Hinshaw VG (1983) Scientists' understanding of propositional logic: an experimental investigation. Soc Stud Sci 13:131–146

Knauff M (2006) A neuro-cognitive theory of relational reasoning with mental models and visual images. In: Held C, Knauff M, Vosgerau G (eds) Mental models and the mind: current developments in cognitive psychology, neuroscience, and philosophy of mind. Elsevier, Oxford (UK), pp 127–149

Knauff M (2007) Logical thinking and the brain. *Synthese* (scheduled for spring 2007)

Knauff M, Fangmeier T, Ruff CC, Johnson-Laird PN (2003) Reasoning, models, and images: behavioral measures and cortical activity. J Cogn Neurosci 4:559–573

Knauff M, Johnson-Laird PN (2002) Visual imagery can impede reasoning. Memory Cogn 30:363–371

Knauff M, Kassubek J, Mulack T, Greenlee MW (2000) Cortical activation evoked by visual mental imagery as measured by functional MRI. NeuroReport 11:3957–3962

Knauff M, May E (2006) Visual imagery, reasoning, and blindness. Quart J Exp Psychol 59:161–177

Knauff M, Mulack T, Kassubek J, Salih HR, Greenlee MW (2002) Spatial imagery in deductive reasoning: a functional MRI study. Cogn Brain Res 13:203–212

Knauff M, Rauh R, Schlieder C (1995) Preferred mental models in qualitative spatial reasoning: A cognitive assessment of Allen's calculus. In Proceedings of the Seventeenth Annual Conference of the Cognitive Science Society (pp 200–205). Erlbaum, Mahwah, NJ

Knauff M, Rauh R, Schlieder C, Strube G (1998) Mental models in spatial reasoning. In: Freksa C, Habel C, Wender KF (eds) Spatial cognition—an interdisciplinary approach to representing and processing spatial knowledge. Springer, Berlin, pp 267–291

Kosslyn SM (1994) Image and brain. MIT Press, Cambridge, MA

Kosslyn SM, Alpert NM, Thompson WL, Maljkovic V, Weise SB, Chabris CF, Hamilton SE, Rauch SL, Buonanno FS (1993) Visual mental imagery activates topographically organized visual cortex: PET investigations. J Cogn Neurosci 5:263–287

Kosslyn SM, Thompson WL, Alpert NM (1997) Neural systems shared by visual imagery and visual perception: a positron emission tomography study. NeuroImage 6:320–334

Leevers HJ, Harris PL (1999) Persisting effects of instruction on young children's syllogistic reasoning with incongruent and abstract premises. Think Reason 2:145–173

Manktelow KI (1999) Reasoning and thinking. Psychology Press, Hove (UK)

Marcus SL, Rips LJ (1979) Conditional reasoning. J Verbal Learn Verbal Behav 18:199–233

Markovits AS (1990) The other ''American Exceptionalism'': why is there no soccer in the United States? Int J Hist Sport 7(2):230–264

Mellet E, Petit L, Mazoyer B, Denis M, Tzourio N (1998) Reopening the mental imagery debate: lessons from functional anatomy. NeuroImage 8:129–139

Möbius PJ (1902) Über das Pathologische bei Nietzsche. Bergmann, Wiesbaden

Newell A (1980) Physical symbol systems. Cogn Sci 4:135–183

Newstead SE, Pollard P, Griggs RA (1986) Response bias in relational reasoning. Bull Psychon Soc 2:95–98

O'Brien DP (2004) Mental-logic theory: what it proposes, and reasons to take this proposal seriously. In: Leighton J, Sternberg RJ (eds) The nature of reasoning. Cambridge University Press, New York, pp 205–233

Plaut DC (1995) Double dissociation without modularity: evidence from connectionist neuropsychology. J Clin Exp Neuropsychol 17:291–321

Polk TA, Newell A (1995) Deduction as verbal reasoning. Psychol Rev 102:533–566

Pylyshyn ZW (1984) Computation and cognition: toward a foundation for cognitive science. MIT Press, Cambridge, MA

Rauh R, Hagen C, Knauff M, Kuß T, Schlieder C, Strube G (2005) From preferred to alternative mental models in spatial reasoning. Spatial Cogn Comput 5(2&3):239–269

Read DE (1981) Solving deductive-reasoning problems after unilateral temporal lobectomy. Brain Lang 12:116–127

Revlis R (1975) Two models of syllogistic inference: feature selection and conversion. J Verbal Learn Verbal Behav 14:180–195

Rips LJ (1994) The psychology of proof. MIT Press, Cambridge, MA

Roberts MJ (2000) Strategies in relational reasoning. Think Reason 6:1–26

Ruff CC, Knauff M, Fangmeier T, Spreer J (2003). Reasoning and working memory: common and distinct neuronal processes. Neuropsychologia 41:1241–1253

Salmon WC (1983) Logik. Reclam, Stuttgart

Sax L (2003) What was the cause of Nietzsche's dementia? J Med Biogr 11:47–54

Schaeken W, Girotto V, Johnson-Laird PN (1998) The effect of irrelevant premise on temporal and spatial reasoning. Kognitionswissenchaft 7:27–32

Schaeken W, Johnson-Laird PN (2000) Strategies in temporal reasoning. Think Reason 6:193–219

Schaeken W, Johnson Laird PN, d'Ydewalle G (1996a) Tense, aspect, and temporal reasoning. Think Reason 2:309–327

Schaeken W, Johnson Laird PN, d'Ydewalle G (1996b) Mental models and temporal reasoning. Cognition 60:205–234

Schain R (2001) The legend of Nietzsche's syphilis. Greenwood Press, Westport, CT

Shaver P, Pierson L, Lang S (1975) Converging evidence for the functional significance of imagery in problem solving. Cognition 3:359–375

Simchowitz S (1900) Der kranke Nietzsche. Frankfurter Zeitung, 7. August 1900

Springer SP, Deutsch G (1981). Left brain, right brain. W. H. Freeman and Company, San Francisco

Stenning K (1998) Aligning logical and psychological contributions to the understanding of human reasoning. Kognitionswissenschaft 7(1)

Stenning K, Oberlander K (1995) A cognitive theory of graphical and linguistic reasoning: logic and implementation. Cogn Sci 19:97–140

Sternberg RJ (1980) Representation and process in linear syllogistic reasoning. J Exp Psychol: Gen 109:119–159

Taplin GV (1971) Reasoning with conditional sentences. J Verbal Learn Verbal Behav 10:219–225

Teuber H-K (1955) Physiological psychology. Annu Rev Psychol 6:267–296

Vandierendonck A, De Vooght G (1997) Working memory constraints on linear reasoning with temporal and spatial contents. Quart J Exp Psychol 50A:803–820

Warrington EK, Taylor AM (1973) Contribution of the right parietal lobe to object recognition. Cortex 9:152–164

Wason PC (1966) Reasoning. In: Foss BM (ed) New horizons in psychology. Penguin, Harmondsworth, Middx

Wason P, Johnson-Laird P (1972) Psychology of reasoning: structure and content. Harvard University Press, Cambridge, MA

Wethrick N, Gilhooly K (1990) The figural effect in syllogistic reasoning. In: Gilhooly K, Keane MTG, Logie R, Erdos G (eds) Lines of thought: reflections on the psychology of thinking, vol. 1, Wiley, London, pp 99–108

Wharton C, Grafman J (1998) Deductive reasoning and the brain. Trends Cogn Sci 2:54–59

Woodworth RS, Sells SB (1935) An atmosphere effect in formal syllogistic reasoning. J Exp Psychol 18:451–460