

How our brains reason logically

Markus Knauff

Center for Cognitive Science

University of Freiburg

Friedrichstr. 50

79098 Freiburg, Germany

Phone: ++49/(0)761/203-4951

FAX: ++49/(0)761/203-4938

Email: [knauff@cognition.iig.uni-freiburg.de](mailto:knauff@cognition.iig.uni-freiburg.de)

to appear in “Special Issue of TOPOI on Logic and Cognitive Science”

edited by Johan van Benthem, Helen Hodges, & Wilfrid Hodges

Last modification: 04/07/06

## Abstract

The intention of this article is to create a link between cognitive brain research and formal logic. As the paper is aimed at a highly interdisciplinary readership, I have cut down the physiological/anatomical and logical/technical treatments to a minimum. 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. 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. This account resolves many useless disputes about how humans reason logically and why we sometimes deviate from the norms of formal logic.

*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 was brought in to the lecture hall, who had been recently committed to the institution. Binswanger introduced him to us - Professor Friedrich Nietzsche! [...] At first sight, he did not appear like a sick man. He was of middle size, 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* from Sep./07/1900 (Simchowitz, 1900, the above passage is my translation from the original German newspaper article). Figure 1 shows the admission files from the psychiatric clinic in Jena and Figure 2 a photograph of Nietzsche sitting in the sanatorium. Some scholars have spread the rumor that Nietzsche's dementia was caused by a cerebral syphilis that he had contracted some 20 years ago (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 it is of more interest that some of Nietzsche's 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).

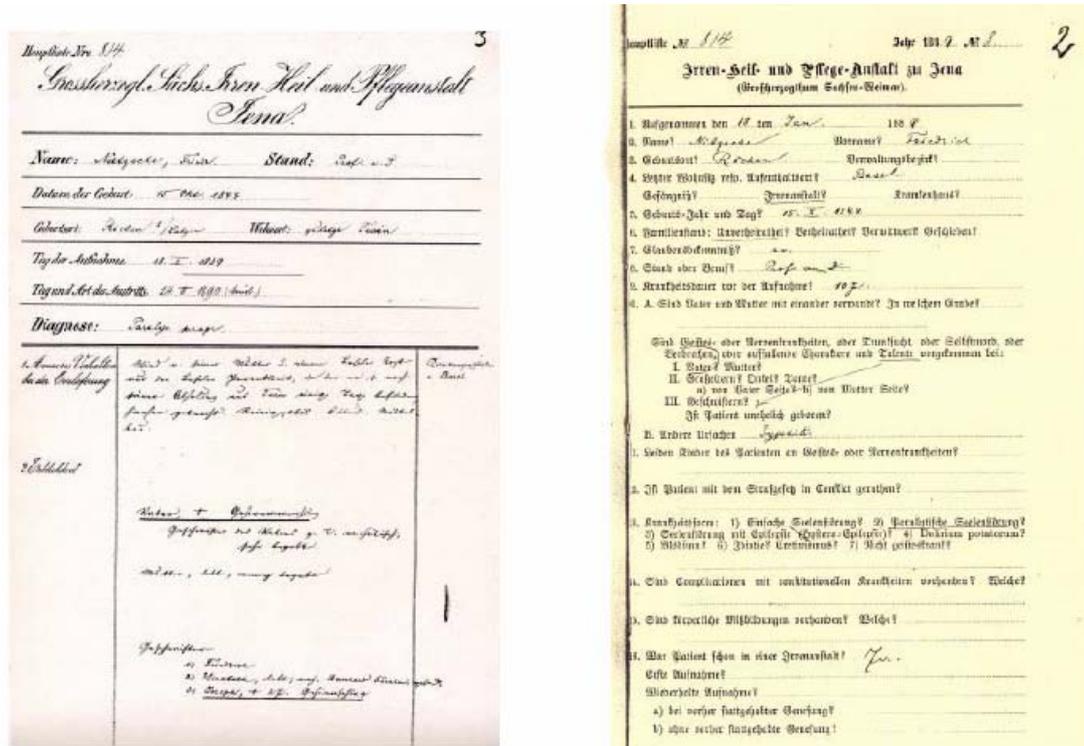


Figure 1: Left: Coversheet of the admission files from the psychiatric clinic in Jena. Right: filled-in patient admission form. © Picture: Dwars; Source: Klassik Stiftung Weimar, Goethe Schiller Archiv.



Figure 2: Nietzsche sitting in the “Irren- Heil- und Pflege-Anstalt zu Jena”. © Picture: Schellbach; Source: Klassik Stiftung Weimar, Goethe Schiller Archiv.

Today, about one century later, it is well known that brain infections, or damages from tumors, strokes, or brain traumata, can have severe effects on some cognitive functions while others remain entirely untouched. Historically viewed, the famous psychologist Karl Lashley (1929) was obviously wrong when he claimed that cognitive functions are distributed throughout the entire cortex rather than localized to one part of the brain. We now know that there is a degree of modularity in the brain's overall organization (Goel, 2005). For instance, we have identified some brain regions specifically involved in the processing of visual information, others in the processing of spatial information, and yet more in the processing of heard language and the generation of speech.

A few years ago a small number of psychological laboratories started to systematically investigate the connection between logical reasoning and the brain (overviews in: Goel, 2005; Grafman & Goel, 2002; Knauff, 2006; Wharton & Grafman, 1998). Before that most reasoning researchers were committed to the assumption that human reasoning should be studied in terms of computational processes. How these computations are biologically implemented in the human brain has been considered insufficient, because each computational function can be computed on each hardware (and thus also the brain) that is equivalent to a Turing machine (e.g. Fodor, 1975; Newell, 1980; Pylyshyn, 1984). The progressing neuro-cognitive movement in reasoning research is based on the fact that this "independence of computational level" hypothesis seems questionable to several researchers. First, the assumption that each function computable by a Turing Machine can be computed on all Turing-equivalent machines is not entirely true (Giunti 1997, Goel 1995). It is true that computational processes can be realized in many different systems, but it is not true that they can be realized in all Turing-equivalent machines. The assumption of universal realizability thus appears to be unjustifiable (Goel 2005). A second reason for the new 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 cognitive neuroscience of reasoning is not interested in what Uttal (2001) called the “new phrenology”, namely the pure localization of cognitive processes including all the reductionistic implications. Another way in which the field differs from pure brain research is that cognitive neuroscientists combine their exploration of the brain with behavioral methods in which we typically measure error rates, solution times, and response preferences. This is done because there is a many to many mapping between cortical regions and cognitive functions, and thus neuroanatomical data alone are too weak to formulate cognitive theories of human reasoning.

The aim of the paper is to create a link between the cognitive neuroscience of reasoning and formal logic. This 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 sylogistic reasoning. Here the problem consist 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 is 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 is consistent with 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). Logicians have studied such inferences since classical times, but the most ingenious work was done by the American logician C. S. Peirce (see the collected writings of Peirce in Hartshorne, Weiss, & Burks, 1931-1958). Peirce formally explored the way in which relations can be combined to form a single new relation and his approach still serves as the normative model to evaluate the formal correctness of relational inferences. 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, which logical abilities break down after local brain injuries, and which 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? Aristotle regarded imagination as a sort of mental faculty that plays indispensable roles in perception and thought (Aristotle, *De Anima*, III: 3, quoted after Wedin, 1988). In the *Critique of Pure Reason*, Kant distinguished between “reproductive” and “productive” imagination (*Einbildungskraft*; see Kant, 1787/1998, e.g. A 100-102, A 118-120). Reproductive imagination is essential in the apprehension of empirical phenomena as it supplements the incomplete input of the sensory systems; its materials are past perceptions and experiences. Productive imagination, in contrast, transcends the perceivable or what is empirically given and is the source of abstract and universal categories and thoughts. This “productive” imagination is, psychologically speaking, probably what humans introspectively experience as “thinking in the inner eye”. People with no education in the cognitive sciences but also many cognitive researchers often believe that our ability to reason logically relies on such “productive imaginations”. 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. I start from Peirce’s work on the composition of relations. Peirce did not distinguish between formal and mental reasoning with relations and simply referred to all representations and thoughts

collectively as signs. He distinguished three properties of signs. A sign can be iconic, such as a visual image, if it is structurally similar to what it represents, it can be indexical, if it uses direct physical connections such as in an act of pointing, and it can be symbolic, such as in a verbal description (Peirce, 1931-1958). I will show that iconic representations do indeed play a role in human logical reasoning. Yet, they have other functions than those one might expect.

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. I will formulate some thoughts on the question of whether brain-events tell us something about mental logical reasoning and will argue that the huge explanatory gap between the behaviors we can observe and the brain activity we can measure can be closed by computational theories of human mental reasoning. The article ends with the conclusion that 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 yet sometimes we use visual mental images.

#### *Logical thinking from a neuro-cognitive perspective*

When we use the term “logical thinking” in daily life, we mean almost all kinds of thoughts, ranging from very elemental inferences up to the complex development of scientific theories and the creation of pieces of art. From the writings of Nietzsche’s biographers it is not really clear what they meant when they said that he was unable to think logically. Most likely, the witnesses used the term in a very broad sense too and in fact the biographers do not report any psychological tests that are today routinely used to test brain damaged patients logical skills. Probably, Nietzsche had no deficits in logical reasoning in the literal sense. To avoid such a 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 deduced from the premises are referred to as conclusions. The participants of an experiment can draw the conclusions in many different ways. I will refer to them as “inference verification” and two sorts of “active inference” (Knauff, Rauh, & Schlieder, 1995). To make the difference explicit we can introduce the notation  $\{\varphi_1, \varphi_1\} > \varphi_3$ , to denote the fact that the conclusion is compatible with the premises. Then the three paradigms may be written as follows:

(1) inference verification: does  $\{\varphi_1, \varphi_1\} > \varphi_3$  hold?

(2a) active general inference: find all  $\varphi_3$  such that  $\{\varphi_1, \varphi_1\} > \varphi_3$

(2b) active particular inference: find some  $\varphi_3$  such that  $\{\varphi_1, \varphi_1\} > \varphi_3$

It is essential to see that the term “conclusion” is only used as the last statement of a deductive problem. 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 on 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. The distinction is borrowed from the signal detection theory, which is an important statistical theory of psychophysics that is used when psychologists explore how individuals make judgments under uncertainty (Wickens, 2001). The model is comparable to the signal to noise ratio used in the engineering sciences, for instance when important signals of a machine must be separated from background noise. In this metaphor it becomes apparent that formal logic plays an essential role for the psychology of reasoning. The latter serves as the normative model for the former, or, in other words, for a reasoning researcher the logical validity has the same facticity as a physical signal. The researcher explores how well individuals can separate a logically

valid conclusion from other (invalid) alternatives and why they (sometimes) deviate from the norms of logic. The psychology of reasoning is descriptive and concerned with mental logical reasoning, whereas formal logic is normative and defines the criteria that must be satisfied to call such mental inferences logically valid. In this vein, an individuals' decision is counted as a correct response if a presented conclusion is logically valid and she or he says it is valid. This match between logic and decision is labeled 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 as incorrect – "false alarm". If, finally, the presented conclusion is logically valid but the participants of the experiment says it is in invalid this is also an incorrect response – a "miss". This important connection between logical validity and psychological correctness is summarized in Table 1.

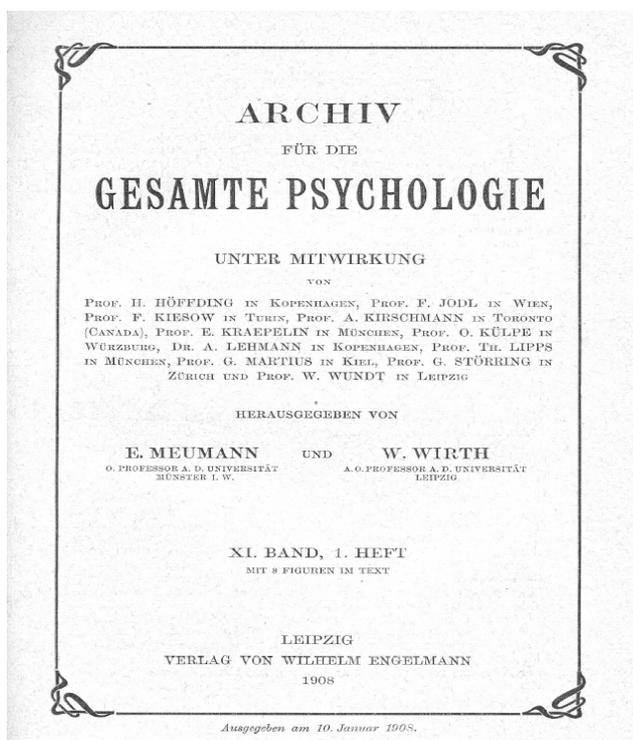
**Table 1. The connection between logical validity and psychological correctness.**

		reasoners' decision	
		valid	invalid
logical validity	valid	hit	False alarm
	invalid	miss	Correct rejection

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

In philosophy, Kant, Leibniz, and Descartes, and in logic George Boole, created a picture of human beings as an intelligent creature that thinks rationally and follows logical rules. From our daily life, however, we are often painfully aware that we are far away from this self-delusion. A superhuman creature such as a robot in a science fiction movie or a logical theorem proofer might follow the rules of standard logic, but human beings often make errors in logical reasoning and behave irrationally.

The first attempt to study human logical reasoning can be dated back to the ingenious work of the German scholar G. Störing (1908). His innovative experiments appeared in the *Archiv für die gesamte Psychologie*, only a few years after Nietzsche died from pneumonia and many strokes. Figure 3 shows the coversheet of the journal and the first page of Störing's article. Störing's work is only rarely recognized by modern reasoning researchers, although he was concerned with a very modern question. In the very first sentence of the paper he states that he asked himself what experimental-psychological research can contribute to the solution of debates between logicians. Then, in the progress of the (more than 130 pages long) article he describes a series of experiments in which he studied the logical abilities of volunteers under highly standardized conditions (in a darkened room, under precise-time control, etc.). He explored reasoning with syllogisms and reasoning with spatial, temporal, and abstract relations. His main conclusion from his studies was that people try solving the inference problems by imagining the content of the problem at the inner eye and that they try to bring the objects of the reasoning problem into a "mental" linear order. I will come back to this later.



### Experimentelle Untersuchungen über einfache Schlußprozesse.

Von  
**G. Störing.**

#### Einleitung.

Zu einer experimentellen Untersuchung der Schlußprozesse bin ich zunächst angeregt worden durch einige Streitfragen der Logiker, bei welchen es nabeliegt, an eine Entscheidung auf Grund experimentell-psychologischer Untersuchung zu denken. Dahin gehört vor allem die Auffassung von F. A. Lange, daß alles Schließen sich an der Hand räumlicher Anschauungen vollziehe. Ihm folgt neuerdings Kroman. Das ist eine Vorstellungsweise, die zu experimenteller Prüfung geradezu herausfordert.

Ähnlich hat auf mich die Kontroverse über die Bedeutung der Synthese der Beziehungsgedanken der Prämissen zum Zustandekommen des Schlußsatzes gewirkt. Einige Logiker behaupten bekanntlich, daß aus den Prämissen der Schlußsatz durch eine Synthese der Gedanken der Prämissen gewonnen werde. So Bradley und Schuppe. Letzterer sagt: »Werden die Prämissen unverbunden gedacht, so ist die Konklusion ein neues Urteil; werden sie in ihrer durch das identische Moment hervorgebrachten Verbindung als das eine *S* oder *P* gedacht, so ist die Konklusion nur der Ausdruck dieser Verbindung<sup>1)</sup>. Demgegenüber wird von anderer Seite behauptet, das Schließen bestehe darin, »daß durch die Vergleichung der beiden Prämissen die Notwendigkeit erkannt würde, dem Subjekt *S* ein Prädikat *P* beizulegen, und auf Grund dieser eingesehenen Notwendigkeit erst würde der Gedanke der Einheit *SP* wirklich vollzogen<sup>2)</sup>. Auch diese Kontroverse legt den Gedanken

1) Schuppe, Grundriß der Erk. und Logik. S. 53, vgl. e. Erk.-Log. S. 260.  
2) Sigwart, Logik. 2. Aufl. I. Bd. S. 443.

**Figure 3: Coversheet of the journal “Archiv für die gesamte Psychologie” from 1908 and the first page of Störing’s article “Experimental investigations on simple inferences”. Picture: Dwars; Source: Klassik Stiftung Weimar, Goethe Schiller Archiv.**

A while later the Swiss developmental psychologist Jean Piaget studied how children's ability to reason logically increases as they age. His research is much more acknowledged by the current reasoning community than that of Störing. It culminated in a theory of cognitive development, which stated that children of different ages are equipped with (or have access to) different inventories of inference rules as a basis for reasoning. Piaget's main assumption was that human reasoning relies on a mental logic consisting of formal inference rules (Piaget, 1936/1963).

Experiments, in which the psychologist presents logical problems to children or adults and then explores their reactions, still remain an indispensable method of reasoning research. Here I only report studies with adults that have already reached the level of formal thinking (c.f. Piaget, 1963). The experimental technique enables us to systematically observe behavior and then to draw inferences from the observed data about unobservable mental processes. Although these processes are realized by the neural tissue of the brain this functional analysis does not need a view into the brain itself. Rather, in such experiments the participants are typically placed in a laboratory setting in which a computer serves as an interface that administers the problem presentation and data collection. 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?

The first question has been explored in a countless number of experiments. Overall, 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, & Byrne, 1991; Evans, Newstead, & Byrne, 1993; Manktelow, 1999). In reasoning with conditionals, for instance, many reasoners misunderstand the logical meaning of the conditional expressions “if” and “then”. As summarized in Table 2, in the propositional calculus there are two logically valid and two logically invalid inferences. 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 2. The four inference schemas of conditional reasoning. MP and MT are logically valid; AC and DA are logical invalid.**

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, the q $\neg p$ ----- $\neg q$	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 $\neg q$ ----- $\neg p$	valid	If it rains, the street is wet. The street is not wet. It does not rain.

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.

Table3: Relative frequency [in %] of how often people treat the four inference schemas MP, DA, AC and MT as logically valid. Results are partly adopted from Evans, Newstead and Byrne (1993).

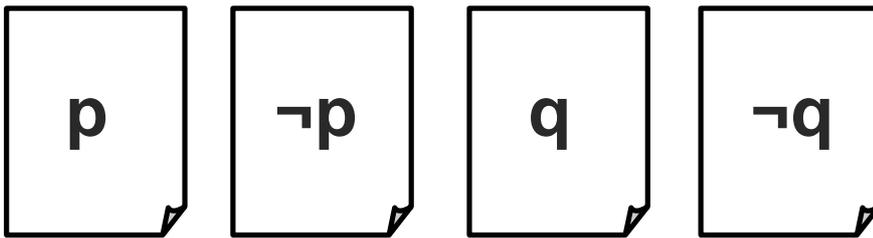
Experiment	n	MP	DA	AC	MT
Taplin (1971)	56	92	52	57	63
Evans (1977)	16	100	69	75	75
Marcus and Rips (1979) *	78	99	37	28	57
Kern, Mirels and Hinshaw (1983)	72	89	28	27	41
Markovits (1988)	76	100	52	42	59

\* Mean of three experiments

A conditional problem that cannot be ignored in an article on logical reasoning is the Wason-Selection-Task (WST). It was invented by the ingenious British psychologist Peter Wason and since then has been developed into one of the most important paradigms in the psychological research on human reasoning (Wason, 1966). In the task, the four cards shown in Figure 4a 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  $p$ ,  $\text{not-}p$ ,  $q$ , and  $\text{not-}q$ . Figure 4b shows the same problem with the four possible propositions as 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  $\text{not-}q$  (according to the MT a  $\text{not-}p$  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 role. This experimental finding has been replicated in dozens of experiments (e.g. Griggs, 1995; Feeney, & Handley, 2000).

**Figure 4a. 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”.**



**Figure 4b. The visible letters and numbers on the card correspond to the four possible propositions p, not-p, q, and not-q.**

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 %

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).

The second group of deductions explored in reasoning psychology are syllogistic inferences. In such tasks, people are usually confronted with problems that consist of generalizations. In natural language such generalizations are expressed in 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 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  $\forall$ , stand for “for all” and allows us to state that all elements in the scope of the quantifier have a certain property. The existence quantor  $\exists$  stand 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 the predicate logic. These inferences are the famous Aristotelian Syllogisms and in fact I would label them the “Aplysia” of reasoning research.<sup>1</sup> The syllogisms rely on the four “moods” presented in Table 5. The table also shows their formal definition in the predicate logic notion.

**Table 5. The four moods of a syllogism**

mood		Notion 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	$\neg \exists x (A(x) \wedge B(x))$	denying	universal
O	Some of the A are not B	$\exists x (A(x) \wedge \neg B(x))$	denying	particular

When the four moods of Table 5 are combined in two premises, then from a formal point of view we get exactly 15 different inferences (Salmon, 1983). However, in psychology it makes a difference if the terms in a premises are presented in the A – B or B – A order, if the premises are presented in a continuous (A-B, B-C) or discontinuous order (B-C, A-B), or if the conclusion that must be verified is presented in the order A-C or C-A (Garnham & Oakhill, 1994). Thus, Table 6 gives an overview of the 16 problems that can be obtained by changing the figure and the order of terms and premises. If the four

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

moods are taken into account too, this results in the 64 syllogisms that have been most extensively explored in reasoning research (there are up to 512 syllogisms if all combinations are taken into account).<sup>2</sup>

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

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

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

<sup>2</sup> 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).

All the professors are beekeepers.

All the beekeepers are athletics.

From these premises almost all individuals correctly deduce the conclusion

All the professors are athletics.

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

All the beekeepers are professors.

None of the athletics are beekeepers.

Only a few people, 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 are presented that can be used in the verification paradigm, i.e. if each possible premises is presented in one of the four moods A, E, I, O and the to-be-verified conclusion also consists of one of these moods. 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.

**Table 7: Relative frequency of selected conclusions [in %, rounded on an integer] for all syllogisms after Chater und Oaksford (1999). The first column denotes the syllogism and its figure, the second denotes the logically valid conclusion, and the columns 3 to 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.**

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

How can the difficulties many people have with the syllogisms be explained? The psychological literature provides numerous explanations and most of them make their specific contribution to our understanding of human logical reasoning. In fact, it seems that a cognitive account of syllogistic reasoning must incorporate theories that focus on the misunderstanding of quantifiers (e.g. Begg 1987; Chapman & Chapman, 1959; Revlis, 1975), the surface structure of premises that biases reasoners towards certain conclusions (e.g. Begg & Denny, 1969; Woodworth & Sells, 1935), the ease to match a presented conclusion with a mental calculation (Wetherick & Gilhooly, 1990), and the application of heuristics that save cognitive resources (overview in Evans, 1989). All such “mini” theories can be labeled as “error theories” of reasoning, as they try to explain reasoning difficulty and why some problems lead people into errors. 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 too much about the second question, which cognitive processes underlie our logical competences. So, which cognitive mechanisms enable us to reason logically? Here there still exists 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 and to provide explanations for reasoning errors that go far beyond that of the error theories. 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 & 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, & Byrne, 1991; Johnson-Laird, in press). 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, as this third class of deductive inferences uncovers some of the most important characteristics of human thinking. Consider, for example, the following problem:

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). For a proponent of the rule-based school of reasoning, the inference yields a transitive conclusion, in which its validity is dependant 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 of 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 & O'Brien, 1998).

Supporters of the model theory offer a completely different concept of how humans reason logically. Technically speaking, 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; Johnson-Laird, & Byrne, 1991, Johnson-Laird, 2001). Psychologically that means that human reasoning does not rely on syntactic operations like in the rule-based approaches, but rather on the semantic interpretation of the premises. Reasoning here relies on the construction of an integrated mental representation of the information that is given in the reasoning problem's 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, it represents in “small scale” how “reality” could be— according to what is stated in the premises of a reasoning problem. Model-theorists have studied problems of this type extensively, and reported findings that are difficult to explain based on formal rules but fit nicely with a model-theoretic view on reasoning. The first is that reasoners try to integrate the premise information into one unified mental representation, which would not be necessary for a rule-based inference process. An important prediction of model theory thus is that the ease of reasoning is a function of the difficulty to integrate the information from the premises into a unified representation. Hence, 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. This finding seems to be an

effect of the difficulty of integrating the information from the premises into a unified representation, because in the continuous and semi-continuous orders it is possible to integrate the information from the first two premises into one representation—a mental model—at the outset, whereas when they are presented with the discontinuous order, participants must wait for the third premise in order to integrate the information in the premises into a unified representation. Similar results are reported, for instance, in Carreiras and Santamaria (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 integrated at this point (see Table 8, Exp. 1 from Knauff, Rauh, Schlieder & Strube, 1998a).

**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. (1998a)**

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 strongest case for premise integration is the difference between determinate problems, in which only a single model can be constructed and indeterminate tasks that call for multiple models. Byrne and Johnson-Laird (1989) compared such problems and found that indeterminate problems (34 % correct) are reliably harder than determinate problems (61 % correct). According to the model theory, indeterminate problems are more difficult because the construction of more than one integrated representation is more difficult than constructing a single model.

In our group, we have extensively investigated reasoning with indeterminate problems, and may have found the most convincing evidence for premise integration and thus against rule-based approaches to reasoning. The model theory ought to explain the integration process as a serial process that always produces the same first mental model. Hence, we tested the assumption of the existence of generally preferred mental models (PMM) in an experiment, in which people had to determine possible relationships between objects based on the information given in the premises. The indeterminate problems called for three, five, or nine possible models. The results showed that whenever a reasoning problem has multiple models, reasoners prefer one of them and that individuals consistently prefer the solution that corresponds to this preferred mental model (PMM). This suggests that individuals indeed integrate the information from the premises and inspect unified mental representations to find new information not given in the premises (Knauff, Rauh & Schlieder, 1995; Rauh, Hagen, Knauff, Kuß, Schlieder, & Strube, 2005; Vandierendonck, Dierckx, & De Vooght, 2004). A detailed computational analysis of these PMM showed that they are those models among all that are the most parsimonious in terms of working memory and processing resources. They are the easiest to maintain in working memory and their processing is less expensive than that of the other models (Ragni, Knauff, & Nebel, 2005; Böddinghaus, Ragni, Knauff, & Nebel, 2006). Based on the fact that this PMM is the first one that is 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 PMM 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 models — often just the single PMM — and ignored other models that are also consistent with the premises. Not surprisingly, this led them into erroneous conclusions (Rauh, Hagen, Knauff, Kuß, Schlieder, & Strube, 2005). The cognitive resources do not limit the rule processing but

rather the consideration of possible models. Reasoners run into errors because they cannot consider all possible models and thus can be unaware of models that would contradict a conclusion the reasoner considered as true in some models.

*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 from brain injuries after infections, tumors, strokes, or accidents. Their destiny normally attracts less attention than Nietzsche's did, but, regrettably, these people also get painfully reminded that our mental capacities are a function of our brains and that reasoning abilities break down if parts of this "hardware" are destroyed. One possible way in which to investigate how reasoning processes are neurally implemented is to ask these patients to help science to explore which 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. For the large group of error theories the predictions are of course very heterogeneous and neuroscientific methods can, if anything, only in parts distinguish between all these "mini theories" (Knauff, 2006). The rule-based and model-based theories, however, come up with clearly distinguishable assumptions cornering the neural correlates of reasoning. On the broadest level they make different predictions regarding the involvement of the two cortical hemispheres. Such assumptions concerning the lateralization of cognitive processes are one of the oldest fields of brain research (e.g. Sperry, 1973). They rely on a simple—and only in parts correct!—job sharing between the "abstract" and "language-related" left hemisphere and the "holistic" and "visuo-spatial" right hemisphere (see Springer & Deutsch, 1981). Here is a representative quote:

*[Mental logic claims] that deductive reasoning is a rule governed syntactic process where internal representations preserve structural properties of linguistic strings in which*

*the premises are stated. This linguistic hypothesis predicts that the neuroanatomical mechanisms of language (syntactic) processing underwrite human reasoning processes . . . [Mental models claims] that deductive reasoning is a process requiring spatial manipulation and search. Mental model theory is often referred to as a spatial hypothesis and predicts that the neural structures for visuo-spatial processing contribute the basic representational building blocks used for logical reasoning. (Goel, 2003)*

The involvement of the two hemispheres has been a major concern of patient studies on logical reasoning (though the quote is taken from an imaging paper). 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. The background of this approach is that the cognitive system is composed of a set of modules, each dedicated to performing a particular cognitive function. 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 Hans Lukas Teuber (1955) to identify when lesions do have specific effects on distinct cognitive functions. One way to measure a brain damaged persons' logical reasoning abilities is to use standard intelligence tests which typically comprise subscales that seek to capture deductive reasoning. However, such tests are typically the result of a test construction strategy that has no reference to theories of cognitive psychology and they also use the term "logical reasoning" in the broadest sense (Wilhelm, 2005).<sup>3</sup> Another method frequently used by reasoning researchers is to confront patients with deductive reasoning problems usually used with healthy people. In this account, patients with different types of brain damage are asked to solve syllogistic-, conditional-, or relational

---

<sup>3</sup> An interesting exception is a logical reasoning test that has been developed by Wilhelm (2005).

reasoning problems and the researcher compares the 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. 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.

**Table 9: Summary of findings from Golding (1981).**

Card selected	RHP	LHP
p q	40	15
p	0	50
p, q, $\neg$ q	20	0
p, $\neg$ q	30	5
other combinations	10	20

Golding's experiment is a little odd because a straight forward 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, because Golding formulated an interference hypothesis and thus her prediction that right hemisphere brain lesions would facilitate insight into the logical problem was upheld. In addition all patients who had a specific perceptual classification deficit solved the problem which indicates that the visual coding of the classification task interfered with the verbal processes of reasoning.

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 while recovering from transitory ictal suppression of one hemisphere by electroconvulsive therapy (ECT; that 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.

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 weaker than the RHP, but, interestingly, the RHP were more impaired with the incongruent conclusions.

### *Logical reasoning in the intact brain*

The patient studies have been frequently interpreted in favor of the rule-based theories of reasoning, as they show that reasoning occupies a language-related fronto-temporal neural network (Goel, et al. 1997, 1998). However, more sophisticated experimental methods show that this is probably just one side of the coin, and that individuals typically use different cognitive strategies to solve different reasoning problems. A more precise method to explore the cortical substrates of cognition are brain imaging methods. These techniques enable researchers to explore the intact brain in action and to locate neural activity with about 1 millimeter precision. Positron-emission-tomography (PET), for instance, uses small amounts of radioactively marked tracer molecules such as glucose to measure differences in cerebral blood flow (CBF). However, most of the research on logical reasoning has been performed by means of functional magnetic resonance imaging (fMRI). In such experiments, a volunteer lies on their back in the fMRI device and can see (by means of a mirror system) a projection screen at the end of the device's bore. Typically, the premises and conclusions of the inferences are presented on this screen and the participants evaluate the conclusion by pressing keys on a response box he/she holds in hand. Alternatively, the problems can be presented as spoken sentences via headphones and the volunteers also use the response box for the evaluation of logical validity. In both cases, the device measures the local increase in oxygen delivery in the activated cerebral tissue while the person contemplates the problem. 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. Physically, the fMRI technique relies on the understanding that deoxyhemoglobin is paramagnetic relative to oxyhemoglobin and the surrounding brain tissue, and that a local increase in oxygen delivery is thought to be correlated with brain activation (Logothetis, Pauls, Augath, Trinath & Oeltermann, 2001; Logothetis, 2002). The principle of fMRI experiments is to measure brain activation of quickly repeated intervals and to explore differences among them. Typically, the baseline activity is measured when the participant is at rest, and other measurements are taken when the participant performs certain cognitive tasks. In the simplest experimental design, the activity in the baseline condition is then subtracted from the activity measured during the performance of the cognitive tasks. The resulting data can be statistically analyzed. Areas in which statistically significant differences were measured are presumed to have been activated by the cognitive task. In more sophisticated experiments, combinations of experimental conditions are compared to other combined conditions. To illustrate the results, the patterns of activation are usually transferred into fMRI images, in which the most visible regions correspond to the areas activated by the cognitive task.

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. 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. As model-based reasoning theories suggest reasoning is a visuo-spatial process these theorists view the occipital cortex as an essential brain structure for reasoning. Here is a representative quote from our own group:

*[According to the model theory] the primary visual cortex, or at least nearby visual regions, should be evoked by reasoning without a specific assumption concerning hemispheric differences (Knauff, et al. 2002).*

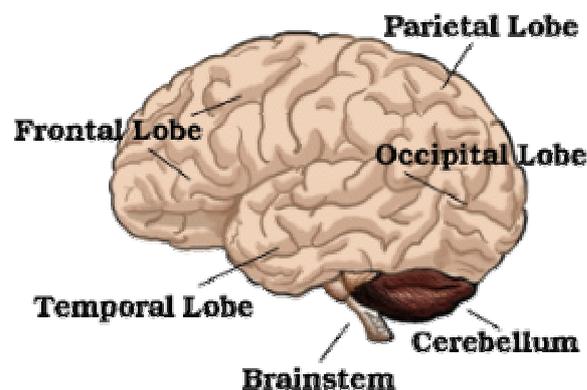
The temporal lobes, which are located one on each side of the brain at about the level of the ears, are involved in the processing of auditory signals and hearing. They are also involved in high-level auditory processing including speech processing. Particularly, the Wernicke's area plays a key role in language understanding. It is clear that rule-based theories must assume that areas of these temporal cortices are active during reasoning, probably with a left hemispheric prevalence (Goel, et al., 1997). Other areas of the temporal cortices are not only involved in comprehension and verbal memory but also in the processing of information regarding the form and identity of visual objects.

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. This area is probably the most important for the model theory of reasoning.

*On the neuroanatomical level, the sentential theory predicts that the language processing regions of the brain are involved in reasoning, whereas the [model] theory predicts that the cortical areas involved in spatial working memory, perception, and movement control are evoked by reasoning. The sentential theory, furthermore, predicts a dominance of the left hemisphere, whereas the spatial theory assumes a right hemispheric prevalence (Knauff, et al. 2002, p 204).*

In the second part of the paper, when I explain which role visual mental images play in human reasoning, I will come back to the fact that model theories predict an involvement of both the parietal and the occipital cortices. In the present context no differentiation between the two lobes and their role in reasoning is necessary.

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. It is involved in executive processes in working memory and typically implicated when several pieces of information need to be monitored and manipulated. No researcher would deny that the frontal cortex plays a central role in logical reasoning, but the two theories do not make different predictions.



**Figure 5. The four lobes of the human brain from the side (lateral) view. (a permission for the reproduction of this Figure is needed).**

The available brain imaging studies provide strong evidence against the idea that logical reasoning is a purely rule-based process as suggested by the patient studies. 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 occipitoparietal–frontal network, including parts of the prefrontal cortex the parietal cortex, and the visual association cortex. It is obvious that these findings are difficult to explain based on purely linguistic processes, as the activated brain areas are implicated in the processing of visual and spatial information and visuo-spatial working memory (cf. Andersen, 1997; Smith & Jonides, 1997; Jonides, et al. 1993). 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.

Possibly the most convincing evidence that not only linguistic but also model-based processing plays a role in logical thinking comes from studies on relational reasoning. In the study 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. It is essential to know that these areas are often identified with “seeing in the minds eye” (e.g. Kosslyn, 1994; Kosslyn & Thompson, 2003). I will come back to this point later, but for the moment it is sufficient to notice that there is no reason why a rule-based purely abstract inference process should activate these brain areas. The same findings are reported in Goel and Dolan (2001) who addressed the question of model-based- and rule-based inferences by using 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. Reasoners also rely on these areas even if the content of the problem is abstract and in no way pushes the individual towards using something like models or visual mental images. It seems more that reasoning does not rely on a single cognitive mechanism but rather on a combination of linguistic processes that probably work analogues to the proofs of formal logic and visuo-spatial processes that can be described as the construction and manipulation of mental models that capture what is described in the premises of the reasoning problem.

*Do we reasoning logically by using mental images?*

The reported experimental findings provide strong evidence that in many we situations reason logically by constructing and inspecting mental models. It only stands to reason that cognitive psychologists and psychological laymen think that such models might be something like “mental pictures” in the “mind’s eye” and that we can “look” at these pictures to find new, not explicitly given information. Such an idea to identify logical models with visual images might appear bizarre to a logician. In psychology, however, a very vigorous debate revolved 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. From the many psychological definitions of visual mental imagery I prefer to say that it is 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.

As I have already mentioned in the introduction the American logician Pierce formally explored how relations can be combined. For him a representation, no matter in which medium, can be iconic, such as a visual image, if it structurally similar to what it represents. However, it can be indexical too, if it relies on direct physical connections such as in an act of pointing, and finally it can be symbolic, such

as in a verbal or as purely symbolic description. In psychology, a debate between supporters of iconic mental representations on one side and proponents of purely symbolic representations on the other started during the early decades of the last century. On the one hand, in 1910 Cheves Perky discovered that mental imagery supports visual perception and that people often merge mental images and what is actually seen. In other words, visual imagination can be so similar to real perceptions that they can be mistaken for the latter (Perky, 1910). On the other hand, German psychologists belonging to the “Würzburger Schule” promoted the assumption that thinking is possible without imagination. The claim was supported in an experiment by Karl Bühler, who asked individuals, for instance, “Does a man have the right to marry the sister of his widow?” and afterwards asked them what had happened in their mind. Not one of the participants reported experiencing visual images. From his findings, Bühler concluded that thinking is possible without seeing in the mind’s eye (Bühler, 1909). However, other authors criticized the idiosyncrasy of Bühler’s problems and assumed—as Störing did—that individuals reason by constructing an iconic mental representation of the problem that they then can scan for new information not given in the premises. For a long period of time, for most researchers it was a matter of fact that thinking calls for “imagination” in the literal sense — that is, the activity of envisaging objects and scenes in their absence (e.g. Titchener, 1909).

Later, in Anglo-American psychology, publications on mental imagery engendered much controversy. Cognitive psychologists initially avoided the concept of imagery, given the harsh criticism it had received from behaviorists (Watson, 1913). In contemporary psychology, however, a wide range of evidence is compatible with the assumption that imagery is a vital part of human cognition, including the well-known studies of mental rotation, the mental scanning of images (cf. Kosslyn, 1980; Shepard & Cooper, 1982), and studies on the relationship between imagery and creative problem-solving, suggesting that visualization facilitates innovative solutions (Suler & Riziello, 1987; Antonietti, 1991; recent results in: Denis, Logie, Cornoldo, de Vega, & Engelkamp, 2001). Moreover, subsequent to the

well-known imagery debate in the 1980's (overview in: Block, 1981; Tye, 1991), the majority of cognitive researchers agree on the assumption that cognitive processes can rely on a number of different representational formats.

Starting in the 1960's, cognitive psychologists also began to explore the role of visual images in relational reasoning. Such a study was firstly carried out by De Soto, London, and Handel (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, Pierson, & Lang, 1975; Clement & Falmagne, 1986). However, several studies have failed to detect any effect of imageability on reasoning. Johnson-Laird, Byrne, and Tabossi (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, Pollard, and Griggs (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, for a long time the results from many behavioral studies have been inconclusive and have left many questions unresolved.

In our own laboratory we have been involved with the role of visual mental images in logical reasoning for about five years now. Within all of these projects my colleagues and I aimed at unifying 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 I will briefly discuss in the last part of paper.

To understand the following thoughts it is necessary to come back to the functional organization of the brain. As already mentioned the model theory has often been identified with the activation of both the parietal and the occipital cortices. The point is that this implies two quite different views on the internal structure of mental models and the format in which they are represented in the human brain. In much the same way as external representations (such as diagrams and pictures), a mental representation in the brain can also reflect different sorts of information. An external diagram of a machine, for instance, might explain by arrows how it works, by indicating in which direction an element of the machine moves (Hegarty, 2004). Though, we can also have a picture of the machine that exactly depicts its size, shape, color. A representation stands for something else or makes something present that is absent but the question is which aspects of the representandum are represented in the representation. Starting with Palmer (1978) an army of scientists were concerned with the question of what is represented in a mental representation. In the present context it is sufficient that we can identify the two cortical structures implicated in model-based reasoning with two sorts of representational formats. While the parietal cortex (or at least parts 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. 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 turns 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, Petit, Mazoyer, Denis, & Tzourio, 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, Thompson, & Alpert, 1997).

The parietal cortex, in contrast, has often been identified with more abstract spatially organized mental models (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 my sense can be underspecified in many respects as long as it represents the meaning of the premises. It is more abstract than an image, because it captures the logical properties of the problem.

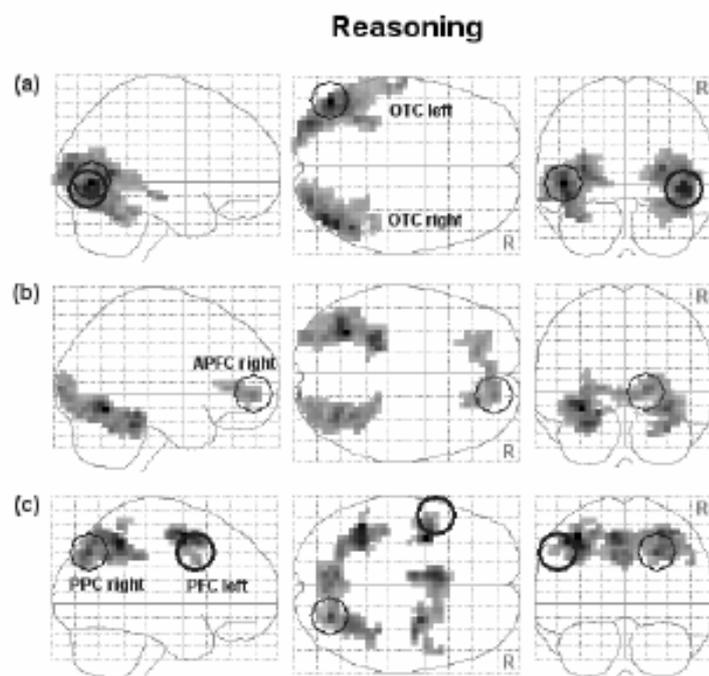
Some of the studies on logical reasoning reported above found little evidence that visual brain areas (in occipital cortex) are involved in reasoning (Goel et al., 1997; 1998). Then, however, an increasing number of studies reported activity in primary and secondary visual areas when participants were engaged 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, Kassubek, Mulack, Salih, & Greenlee (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 occipitoparietal-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, Knauff, Fangmeier, and Spreer (2003). Here we scanned the brain activity of our participants and also measured their visuo-spatial ability 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, Knauff, Fangmeier, & Spreer, 2003).

A main 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 in a blocked fashion (e.g. Knauff, Mulack, Kasubek, Salih, & Greenlee, 2002) or just compared the neuronal processes during the conclusion of the reasoning problem with the presentation of irrelevant control sentences (e.g. Goel, & Dolan, 2001). In both paradigms it is impossible to determine whether the activity in occipital brain areas pointing to 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 inseparably mixed. 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 confounds in the previous studies on the neuronal basis of imagery and reasoning. In the study, we scanned the brains of our volunteers while they solved relational reasoning problems (Fangmeier, Knauff, Ruff, & Sloutsky, 2006). To avoid the need to read the premises and conclusions we replaced the sentences with graphical arrangements describing the spatial relations between three objects. The reasoning problems contained two premises and a conclusion and the participants had to decide whether the conclusion logically (necessarily) followed from the 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 Figure 6. The darker a region in the image indicates that 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 Figure 6a). 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 Figure 6b). 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 Figure 6c).



**Figure 6. Results from a study by Fangmeier, Knauff, Fuff, & Sloutsky (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. The darker a 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.**

What does that mean for the role of visual, spatial and sentential representations in logical reasoning? For proponents of a strictly sentential account are our results frustrating. In fact, we found much neural activation in brain areas that are related to visual and spatial processing. It is difficult to find an immediate explanation for these findings base on purely syntactic rules of inference. However, the results also have significant implications for visual theories of reasoning. While many studies have implied that visual images play a key role in the reasoning process, it now appears that the activity in visual brain areas is for the most part an effect of premise comprehension. This stage relies on neural processes in the occipito-temporal cortex that are known to be involved in visual mental imagery and visual working memory (Postle, Stern, Rosen, & Corkin, 2000; Courtney, Ungerleider, Keil, & Haxby, 1996; Kosslyn, Alpert, Thompson, Chabris, Rauch, et al., 1994, Kosslyn, Ganis, & Thompson, 2001). The most reasonable account for this finding is that the processing of premises spontaneously elicits visual imagery. As suggested by Peirce iconic “signs”, the resulting visual images are structurally similar to a real visual perception and rely on similar brain functions. Like a visual percept, it might represent colors, shapes, and metrical distances. It probably can be rotated and scanned and it might have a limited resolution (cf. Kosslyn, 1994; Johnson-Laird, 1998). It is reasonable to assume that these representations of the premises are responsible for the experience of visual images during reasoning. But do these images play a role in the further processing? Our data suggests the answer to this question is “no”. The essential finding is that in latter phases vision-related activity in the occipital cortex disappears and is replaced by large activated clusters in spatial brain areas in the parietal cortex. The most plausible explanation for this finding is that further reasoning is based on spatial representations

and the visual images are not pertinent to this phase of the reasoning processes. The spatial representations seem to be “models” in the strictest logical sense. They represent the information pertinent to reasoning by means of spatial relations, and in the inferential tasks, the resulting spatial representations are likely to exclude visual detail, and to represent only the information relevant to the inference. The model’s parts correspond to the parts of what it represents, and its structure corresponds to the structure of the reasoning problem (Johnson-Laird, 2001; see also Knauff, & Schlieder, 2005).

### *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 model 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, Fangmeier, Ruff, Johnson-Laird, 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 & May, 2006) we empirically identified four sorts of relations: (1) visuo-spatial relations that are easy to envisage visually 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. Then we started by conducting a series of behavioral experiments in which participants solved linear syllogisms with these relations (Knauff &

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 & Johnson-Laird, 2002). We then performed an fMRI study using the same sorts of problems. As can be seen in Figure 7, 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, because individuals might have the facility to construct mental models from all sorts of relations. Such models will be spatial in form for visuospatial and spatial relations, and, as long-standing evidence suggests, even relations such as “smarter” are also likely to elicit spatial models (see, e.g., Johnson-Laird, 1998; De Soto et al., 1965). However, only the 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 construction calls for additional activity in visual cortices and retards the construction of a spatial mental model that is essential for the inferential process.

*Where does all this leave us?*

We have traveled a long and meandering road, so a recapitulation from a more distant point of view may not be out of place. The intention of this paper has been to create a link between cognitive brain research and formal logic. I have tried to encourage researchers from both fields to pursue such

connections by answering some questions and by highlighting new open issues. First, I have shown that overall people are pretty good in solving logical reasoning problems. With the MP, they typically perform almost perfectly and with many syllogisms even logically untrained people have few problems. However, we sometimes run into errors or completely fail to think in agreement with the norms of formal logic. We have seen that this is often caused by the fact that reasoning performance is limited by the capacities of the cognitive system, misunderstanding of the premises, ambiguity of problems, and the application of resource-saving heuristics. Overall, I have tried to make it clear that the dispute between the different reasoning theories is quite unproductive, because not one of the accounts alone can explain the experimental findings.

Human reasoning in the propositional calculus—conditional reasoning—seems to rely on more than one cognitive mechanism. For instance, the rule-based camp reports experimental findings showing that humans tend to have difficulty with MT, because they are not equipped with a primitive rule such as the rule for MP. A rule for MT inferences must be derived from the MP and other rules each time it is used, what makes the inference more prone to errors (Braine, Reiser, & Rumin; 1984; Rips, 1994). Another point is that psychological theories often still regard material implication as the normative theory of the conditional. Over the last two decades, however, in the philosophy of logic and language the idea that material implication can account for everyday conditionals has been subject to severe criticism (Oaksford & Chater, 2003). On the other hand, I have also shown that behavioral and neuro-scientific results support the model theory of conditional reasoning. It seems that much of the researchers applying patient studies initially were biased towards the supposedly obvious role of the left hemisphere for reasoning. However, the findings did not really offer a case for the rule-based theories of reasoning beyond saying that the processing of verbal premises involves language-related brain areas. Lesions to the left hemisphere might result in a deficit to process the linguistic elements of the problems and thus in impaired overall performance. This, however, does not say that the actual reasoning process is affected

by damage to the left hemisphere. It is likely that left-hemisphere lesions lead to an inability to process the linguistic aspects of a reasoning problem, but that for the pure reasoning process the right hemisphere is important. This interpretation would also explain most of the findings. For instance, in the study by Caramazza et al. (1976) and Read (1981) the patients in fact had problems in logically deducing the converse of relations. Moreover, Whitaker et al. (1991) examined conditional reasoning in patients that had undergone a unilateral anterior temporal lobectomy, one group to the right hemisphere and the other group to the left hemisphere. The content of the problems was related to the participants' prior knowledge of the world. Given the premises

If it rained, the streets will be dry.

It rained.

the RHP had a strong tendency to conclude “The streets will be wet” while the LHP concluded “The street will be dry”. In other words, the RHP were unable to perform the deduction in isolation from their prior knowledge, while the LHP relied on the linguistic content of the problem and their prior knowledge.

Overall, the role of prior knowledge is essential for the psychology of reasoning. I have not mentioned it so far, although it is one of the central branches of reasoning research (see point 3 in the enumeration of research issues at the beginning of the paper), and also supports my “many-mechanisms theory” of reasoning. In the psychological literature on reasoning the issue of “content effects” pops up everywhere and many studies have shown that what an individual knows about the field of discourse 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 – in this case the inference is supported. Or, the formal truth value conflicts with the truth value of the prior knowledge – then the inference is more difficult, which means it results in more errors or takes significantly longer. If an inference generated by an individual is biased towards the truth value of the

prior knowledge or even overwritten by it, this is called belief bias (Evans, 1989). Belief biases have been shown in all areas of deductive reasoning. The most well-known are all related to the WST. Remember that in the original version of the problem most individuals fail to select the logically valid card  $p$  and  $\neg q$ . However, an influential study was done by Johnson-Laird, Legrenzi and Legrenzi (1972) using the “postal version” of the WST. The participants were all Italians and they were asked to imagine they were postal workers sorting the mail. They had to check the envelopes of letter (instead of cards) which violate the postal rule “If a letter is sealed, then it has a 50 lire stamp on it”. The logical structure of the problem was identical to the classical WST, but now about 90% of all people selected the correct envelopes: the sealed one ( $p$ ) and the one without a 50 lire stamp ( $\neg q$ ). This is only one classical example of docents that nowadays are discussed under the label “deontic reasoning” (Bucciarelli, & Johnson-Laird, 2005; Manktelow, & Over, 1995). Many researchers have argued that the effect goes back to the fact that the content activates a falsification strategy (Johnson-Laird, Legrenzi & Legrenzi, 1972) or found other explanations in the rule-based and in the model-based account of reasoning (Beller & Spada, 2003; Bucciarelli, & Johnson-Laird, 2005; (see also Sperber & Girotto, 2003). Others developed an apparently new view on content effects in reasoning and even reasoning in general. The idea was inspired by the domain-specificity of so-called “expert-systems” in the early years of artificial intelligence research (cf. Russell, & Norvig, 2003). The pragmatic reasoning schemas theory (Cheng and Holyoak, 1985; Cheng, Holyoak, Nisbett and Oliver, 1986) and the social contract theory (Cosmides, 1989) claim that people use domain-specific inference rules that result from their past experience with a (deontic) content. Holyoak and collaborators have presented a few pragmatic schemas that reflect permissions and obligations (e.g. If the action is to be taken, then the precondition must be satisfied. If the action is not to be taken, then the precondition need not be satisfied. etc.), and Cosmides and other developed the idea that human beings are equipped with “Darwinian algorithms” that promote their own survival and that of the species. Such approaches, completely deny that we have any competence in

formal logic. All is knowledge and past experience about the reward and punishment of breaking and accepting social rules (Fiddick, Cosmides, & Tooby, 2000; Fiddick, 2004). The theories brought up some interesting experimental findings that again show that human are very flexible in using different cognitive strategies to solve a logical problem. Sometimes we even avoid reasoning and just “recognize” what follows from the premises. However, since these approaches lie completely outside the logical framework and also focus primarily on conditional reasoning, they are almost irrelevant for the question on how formal and mental logic is related to each other.

Reasoning in the predicate calculus also withstands all tries to explain it with a single cognitive mechanism. So far, psychologists have focused on syllogisms which have two premises each with one quantifier (for an exception see Johnson-Laird, Byrne und Tabossi, 1989). Such problems have advanced the “aplysia” of reasoning research. There is no doubt that people often make errors in syllogistic reasoning, in part because of the existence of various biases (Evans, 1989). No theory of reasoning can abandon that belief biases (e.g., Oakhill et al., 1990), the atmosphere effect (e.g., Woodworth & Sells, 1935), and the conversion error (Chapman & Chapman, 1959) are needed to explain erroneous inferences. Rule-based theories also have some important advantages (Yang, Braine, & O'Brien, 1998). For instance, it is well known that in some reasoning problems, people tend to derive “All B are A” from “All A are B”, even when this inference is logically invalid because the quantifier is non-symmetric. However, Newstead and Griggs (1983) were able to show that about one third of experimental reasoners claims, incorrectly, that “all B are A” follows from “all A are B”, and about two thirds that the conversion of “Some A are not B” follows (Newstead & Griggs 1983). Following Rips this points to general rule of conversion, which is sometimes misapplied (Rips, 1994). Also, the problems rule-based theories have in explaining the indeterminacy effect (problem with more than one model are more difficult, see above) depend only on the set of rules. As van der Henst (2002) has shown there seems to be no (formal) reason to presuppose that a rule-based procedure could not deal with

indeterminate relations, and applying a specific rule-based procedure to indeterminate relations result in greater difficulty, too (van der Henst, 2002). Another advantage of rule-based theories is that they can explain many findings by the difficulty to translate natural language expression into a logical formula, and vice versa (cf. Grice, 1975; Hodges, 1977). A counterargument, however, is that surely all experimental findings can be model by rules as a rule-based system is equivalent to a Turing machine and thus can perform each computational function. The question is, however, if the related assumptions about cognitive processes are psychologically plausible. Moreover, why there should be, for instance, a mental rule of conversion and why people sometimes should be mistaken about the applicability of this rule and sometimes not (Geurts, 2003). The model theory's answer is simply that such a rule does not exist in the first place. The model theory also can refer to many experimental findings showing that reasoners use the semantics of quantifiers to construct and manipulate models of the premises. In particular, the model theory can provide a psychologically plausible explanation—working memory limitations—why the difficulty of a syllogistic problem increases with the number of models that can be constructed from the premises (Johnson-Laird & Byrne, 1991). The theory can also account for the figural effect (Johnson-Laird & Bara, 1984) and many other results from the field of syllogistic reasoning (see Evans, Newstead & Byrne, 1993; for an overview). Moreover, it is remarkable that most of the “rule-based approaches to syllogistic reasoning”, are in important parts more model-based than rule-based (Garnham & Oakhill, 1994). This is true for instance for the theory of verbal reasoning introduced by Polk (1995), and theories based on Euler-circles and Venn-diagrams (e.g. Stenning, 2002; Stenning & Oberlander, 1995). It speaks for the present multi-processing view on reasoning that the two approaches can be brought together in such a way and it is also crucial that models can be represented by sets of sentences, and there are well-known emulations of the one kind of talk in the other (Stenning & Oberlander, 1995).

Goel (2003) also reports findings indicating that different neural mechanisms are involved in

syllogistic reasoning. These studies again have to do with the role of content in reasoning. He presents data from three neuroimaging studies of syllogistic reasoning that point to two neural pathways for human reasoning. His opinion is that a frontal-temporal system processes familiar, conceptually coherent material while a parietal system processes unfamiliar, nonconceptual or incoherent material. The former engages language and long-term memory while the latter engages visuo-spatial processing systems. There is much in this position with which I agree. In particular, I agree that different processes are involved in mental reasoning and I also agree that the visuo-spatial subsystems are involved in reasoning. What I disagree with is the conclusion these authors draw from their findings. In particular, I disagree 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 my 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” 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.

Relational reasoning seems to be the via regia to many questions of reasoning research as people are often very good in solving such inferences. No doubt, the ability to reason with relations is as essential for the survival of our species, as it is for our daily life (Ragni, Knauff, & Nebel, 2006). In fact, these days it is difficult to find one researcher who sincerely maintains that the application of formal rules of inference alone is able to explain how human beings solve relational inferences. In some cases they might, for instance, if a person is trained in formal logic (Klauer, Meiser, & Naumer, 2000) or if the

premises consist of a long chain of identical relations, such as in “A is to the left of B, B is to the left of C, C is to the left of D, D is to the left of E.” use a simple transitivity rule rather than construct a model (Schlieder, 1995). Overall, however, the mental model theory, and in particular the theory of preferred mental models, has a much stronger explanatory power than rule-based approaches and the consistent explanation of experimental findings could hardly be gained by any other theory of reasoning (Held, Knauff, & Vosgerau, 2006). Goodwin & Johnson-Laird (2005) have argued that such models fulfill the criteria that Peirce defined for his “iconic signs”. However, as the studies from our laboratory suggest, this by no means implies that the models are “visual mental images”. It seems much more likely that we do not use a single deterministic strategy even if we use “models”. The models can be either “images” in the literal sense or they can be more abstract as they represent the premise information spatially but avoid excessive visual details. Overall, it seems that the ease to visualize the content of a reasoning problem affects whether we rely exclusively on spatial mental models or if we construct visual mental images of the problem content.

*Do brain-events tell us something about mental logical reasoning?*

Some researchers might declare that I am only telling fairy stories! Many philosophers of the mind have convinced us that the behavioral sciences must live with the fact that a view into the brain as such does not seem very promising to gain more knowledge about the building blocks of cognition (e.g. Block, 1980; Fodor, 1975; Newell, 1980; Putnam, 1961; Pylyshyn, 1984). So, do brain-events tell us something about mental logical reasoning? In fact, there is a many to many mapping between cortical regions and cognitive functions, and it is naïve to believe neuroanatomical data alone are powerful enough to formulate cognitive theories of human reasoning I have already mentioned this in the introduction and in Knauff (2006). An even more important concern is that there is a huge explanatory gap between the behavior we can observe and the brain activity we can measure. It is not that the

interpretation of differences in oxygenation is an inappropriate measure for cognitive activities as Logothetis and colleagues have convincingly shown that this is not the case (Logothetis, Pauls, Augath, Trinath & Oeltermann, 2001; Logothetis, 2002). Rather, we must explain how the involved brain areas perform the computations that are needed to solve a cognitive task. If models can be represented by sentences and vice versa, (Stenning & Oberlander, 1995) we need to know a lot more about what kinds of “linguistic”, “spatial”, and “visual” processing are localized, which returns us to the old argument against attempts to draw conclusion from brain events to cognitive processing (K. Stenning, personal communication 2006). So, do brain-events after all do not tell us something about mental logical reasoning? I think they do given that we are able to close the explanatory gap between brain and mind. First, after approximately 50 year brain-abstinent reasoning research, the independence of the computations from the neural structures in which they are implemented seems to me to be open to discussion (Giunti 1997; Goel, 2005). Second, we are today in the situation to have precisely defined hypotheses and so many behavioral results are available that limit the arbitrary interpretation of brain data that it is worth having a closer look into the reasoning brain in action. Third, and above all, however, we have the possibility to close the mentioned gap by defining computational models that precisely determine which algorithms are run in the cerebral cortex. They are absolutely not obsolete. Brain imaging studies might tell us something about the format of representations, but without assumptions concerning the processes that work on them this is quite useless. This idea is motivated by the famous “mimicry argument” that has been more generally expressed by Anderson (1978), who argued that to realize any system of representation, one needs to know not only the format of the representations, but also the processes that operate on these representations.

To bridge the mind-brain-gap in reasoning research, in our group we have developed a functional neuro-computational theory of human relational reasoning that describes which format of mental representations we postulate and which computational processes we believe to work in the brain. In the

**SRM** theory (Spatial reasoning by models) we assume that the models are represented in spatial working memory which is mainly located in the posterior parietal cortex. This spatial working memory contains the representation of the premises as objects with their relations and is conceptualized as a spatial array. In this spatial array the spatial information is represented only in relational – not metrical like images – terms, thus following the line of spatial representations and rejecting concepts of visual mental images.

Binary spatial relations are defined as a triplet  $(X, r, Y)$  in which

X is the referent,

r is a binary relation, and

Y is the relatum.

X is called the “to be located object” (LO) and Y the “reference object” (RO) (Miller & Johnson-Laird, 1976). As relations “r” we use only the most parsimonious one: “left of”, “right of”, “in front of”, or “behind”. We interpret these relations uniquely, i.e. no two of these relations interfere with one another so that “left of”, for instance, indicates that the considered objects (the RO and the LO) are in the same horizontal line with any number of (zero to many) empty or filled cells between them. As the neural counterpart of the spatial area we think of the receptive fields of neurons in the parietal cortex. The receptive field consists of any stimulus that changes the neuron's firing rate. Hence, we think that a mental model is represented in such a way that neighboring regions of the representation represent neighboring regions of the model. All operations on the spatial array – namely the reasoning processes – are considered as moves of a spatial focus. This focus can place an element into the model or inspect the model to find new information (this focus shift works almost automatically under the control of parietal brain structures). We assume that the reasoning process proceeds in three steps. In the construction phase, in the parietal cortex a mental model is constructed that reflects the information from the premises. In agreement with many experimental findings, we assume that if new information is encountered during the reading of the premises, it is immediately used in the construction of the model

(Johnson-Laird & Byrne, 1991). In the inspection phase, processes in the anterior-prefrontal cortex (see Figure 6b) come into play with which this model is inspected to find new information that is not explicitly given in the premises. The third phase, the validation phase, is controlled by neural processes in the dorso-lateral prefrontal cortex (see Figure 6c) that check the validity of the putative conclusion in other models. Here we implemented our own account by saying that there is no iteration process, in which model are constructed and inspected in turn, but rather a process that starts from the preferred mental model (PMM) and then varies this model to find alternative interpretations of the premises (Rauh, Hagen, Knauff, Kuß, Schlieder, & Strube, 2005). The model is the one that is easier to construct and to maintain in working memory than other possible models (Knauff, Rauh, Schlieder, & Strube, 1998).

We have currently implemented two versions of this model-based component of human logical reasoning. One of these works as a stand-alone-system (Ragni, Knauff, & Nebel, 2005). The other is implemented as a new component of the ACT-R theory (Böddinghaus, Ragni, & Nebel, 2006), which is at present the best-known cognitive architecture of cognitive psychology and a theory for simulating and understanding human cognition. ACT-R strives to understand how people organize knowledge and produce intelligent behavior (Anderson & Lebiere, 1998). As such our account is integrated into a more general framework of cognition that also uses syntactic production rules of inference to solve logical problems and that offers specific memory modules in which background knowledge can be appropriately modeled. The developers of ACT-R for most of the time were concerned with the prediction of response latency and on this basis have postulated a series of cognitive components like memory retrieval and rule conflict resolution that are involved in the performance of a task and offer explanations for the factors that control how long these components take (overview in Anderson, & Lebiere, 1998). In recent time however, this model has also been shown to be very helpful in predicting neural activities during cognitive tasks. An impressive example is the reconstruction of neural activities

during the solution of algebraic equations (Anderson, Qin, Sohn, Stenger, & Carter, 2003; 2004) The detailed modeling of mental logical reasoning is probably even more ambitious. However, the goal for the future from my point of view is to bring together in a similar way what is known about the cognitive components of logical reasoning and what has been found in the cognitive neuroscience of reasoning. This would bridge the still existing gap between the mind and brain in reasoning. I believe that progress in this field will provide additional evidence that we do not use a single deterministic strategy for solving logical reasoning problems. The present paper offers initial support for the assumption that we sometimes reason logically analogous to the proofs of formal logic, sometimes think logically by using models in the strict logical sense, and yet sometimes use visual mental images. Such an account will have the potential to resolve many fruitless disputes about how humans reason logically and why we sometimes deviate from the norms of formal logic.

#### Acknowledgements

The author thanks the German National Research Foundation (DFG) for supporting this work by a Heisenberg Award.

#### References

- Andersen, R.A. (1997). Multimodal integration for the representation of space in the posterior parietal cortex. *Philosophical transactions of the Royal Society of London. Series B: Biological Sciences*, 352, 1421-1428.
- Anderson, J. R. (1978). Arguments concerning representations for mental imagery. *Psychological Review*, 88, 249-277.
- Anderson, J. R., & Lebiere, C. (1998). *The atomic components of thought*. Mahwah, NJ: Erlbaum.
- Anderson, J. R., Qin, Y., Sohn, M.-H., Stenger, V. A., & Carter, C. S. (2003) An information-processing model of the BOLD response in symbol manipulation tasks. *Psychonomic Bulletin and Review*, 10, 241-261.
- Anderson, J. R., Qin, Y., Sohn, M.-H., Stenger, V. A., & Carter, C. S. (2004). The Relationship of Three Cortical Regions to an Information-Processing Model. *Journal of Cognitive Neuroscience* 16 (4) 637-653.

- Antonietti, A. (1991). Why does mental visualization facilitate problem-solving? In R. H. Logie & M. Denis (Eds.), *Mental Images in Human Cognition* (pp. 211-227). Amsterdam: Elsevier Science Publishers.
- Begg, I. (1987). Some. *Canadian Journal of Psychology*, 41, 62-73.
- Begg, I. & Denny, J. (1969). Empirical reconciliation of atmosphere and conversion interpretations of syllogistic reasoning. *Quarterly Journal of Experimental Psychology*, 81, 351-354.
- Beller, S., & Spada, H. (2003). The logic of content effects in propositional reasoning: The case of conditional reasoning with a point of view. *Thinking and Reasoning*, 9(4), 335-378.
- Block, N. (1980). What is Functionalism (Eds.) *Readings in Philosophy of Psychology*. Cambridge Massachusetts, Harvard UP, 1980, S.171-184.
- Block, N. (1981). *Imagery*. Cambridge, MA: MIT Press.
- Böddinghaus, Ragni, Knauff, & Nebel, B. (2006). Simulating spatial reasoning using ACT-R. In Proc. of the 7th International Conference on Cognitive Modeling.
- Braine, M.D.S. (1978). On the relation between the natural logic of reasoning and standard logic. *Psychological Review*, 85, 1-21.
- Braine, M.D.S. (1990). The "natural logic" approach to reasoning. In W.F. Overton (Ed.), *Reasoning, necessity and logic: Developmental perspectives*. Hillsdale, NJ: Erlbaum.
- Braine, M.D.S., B.J. Reiser, and B. Rumain (1984). Some empirical justification for a theory of natural propositional logic. In: G.H. Bower (ed.), *The Psychology of Learning and Motivation*. Academic Press, New York.
- Braine, M. D. S., & O'Brien, D. P. (Eds.). (1998). *Mental logic*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Bucciarelli, M., & Johnson-Laird, P. N. (2005). Naïve deontics: A theory of meaning, representation, and reasoning. *Cognitive Psychology*, 50, 159-193.
- Bühler, K. (1909). Zur Kritik der Denkeexperimente. *Zeitschrift für Psychologie*, 51, 108-118.
- Byrne, R. M. J., & Johnson-Laird, P. N. (1989). Spatial reasoning. *Journal of Memory and Language*, 28, 564-575.
- Caramazza, A., Gordon, J., Zurif, E. B., & DeLuca, D. (1976). Right-hemispheric damage and verbal problem solving behavior. *Brain and Language*, 3, 41-46.
- Carreiras, M., & Santamaría, C. (1997). Reasoning about relations: spatial and nonspatial problems. *Thinking and Reasoning*, 3, 191-208.
- Chapman, L.J. & Chapman, A.P. (1959). Atmosphere effect re-examined. *Journal of Experimental Psychology*, 58, 220-226.
- Chater, N. & Oaksford, M. (1999). The probability heuristic model of syllogistic reasoning. *Cognitive Psychology*, 38(2), 191-258.
- Cheng, P.W. & Holyoak, K.J. (1985). Pragmatic reasoning schemas. *Cognitive Psychology*, 17, 391-416.
- Cheng, P. W., Holyoak, K. J., Nisbett, R. E., & Oliver, L. M. (1986) Pragmatic versus syntactic approaches to training deductive reasoning. *Cognitive Psychology*, 18, 293-328.
- Clement, C. A., & Falmagne, R. J. (1986). Logical reasoning, world knowledge, and mental imagery: Interconnections in cognitive processes. *Memory & Cognition*, 14, 299-307.
- Cosmides, L. (1989). The logic of social exchange: Has natural selection shaped how humans reason? Studies with the Wason selection task. *Cognition*, 31, 187-276.
- Deglin, V. L., & Kinsbourne, M. (1996). Divergent thinking styles of the hemispheres: How syllogisms are solved during transitory hemisphere suppression. *Brain & Cognition*, 31, 285-307.

- Courtney, S. M., Ungerleider, L. G., Keil, K., & Haxby, J. V. (1996). Object and spatial visual working memory activate separate neural systems in human cortex. *Cerebral Cortex*, 6, 39-49.
- Denis, M., Logie, R., Cornoldo, C., de Vega, M., & Engelkamp, H. (2001). *Imagery, language and visuo-spatial thinking*. Hove (U.K): Psychology Press.
- De Soto, C. B., London, M., & Handel, M. S. (1965). Social reasoning and spatial paralogic. *Journal of Personality and Social Psychology*, 2, 513-521.
- Ehrlich, K., & Johnson-Laird, P. N. (1982). Spatial descriptions and referential continuity. *Journal of Verbal Learning and Verbal Behavior*, 21, 296-306.
- Evans, J. St. B. T. (1989). *Bias in human reasoning*. Hove, UK: Lawrence Erlbaum.
- Evans, J. St. B. T., Newstead, S. E. & Byrne, R. M. J. (1993). *Human Reasoning*. Hove, UK: Lawrence Erlbaum Associates.
- Fangmeier, T. Knauff, M., Ruff, C. C. & Sloutsky, V. (2006). fMRI Evidence for a three-stage-model of relational reasoning. *Journal of Cognitive Neuroscience*. 18(3), 320-334.
- Feeney, A. and Handley, S.J. (2000) The suppression of q card selections: Evidence for deductive inference in Wason's Selection Task. *Quarterly Journal of Experimental Psychology*, 53A, 4,1224-1242.
- Fiddick, L., Cosmides, L., & Tooby, J. (2000). No interpretation without representation: The role of domainspecific representations and inferences in the Wason selection task. *Cognition*, 77, 1-79.
- Fiddick, L. (2004). Domains of deontic reasoning. *Quarterly Journal of Experimental Psychology*, 57, 447-474.
- Fishman, R. S. (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/ishn2002.htm>.
- Fodor, J. A. (1975). *The language of thought*. New York: Crowell.
- Garnham, A. & Oakhill, J. (1994). *Thinking and reasoning*. Oxford, UK: Blackwell.
- Geurts, B. (2003). Reasoning with quantifiers. *Cognition*, 86, 223–251.
- Giunti, M. (1997). *Computation, dynamics and cognition*. New York: Oxford University Press.
- Goel, V. (1995). *Sketches of thought*. Cambridge, MA: MIT Press.
- Goel, V. (2005). Cognitive Neuroscience of Deductive Reasoning. In K. Holyoak & R. Morrison. (Eds.). *Cambridge Handbook of Thinking & Reasoning* (pp. 475-492) Eds. Cambridge University Press.
- Goel, V. (2003). Evidence for Dual Neural Pathways for Syllogistic Reasoning. *Psychologica*, Vol. 32, pp. 301-309.
- Goel, V., Büchel, C., Frith, C. & Dolan, R.J. (2000). Dissociation of mechanisms underlying syllogistic reasoning. *NeuroImage* 2000, 12, 504-514
- Goel, V., & Dolan, R. J. (2001). Functional neuroanatomy of three-term relational reasoning. *Neuropsychologia*, 39, 901-909.
- 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. *Journal of Cognitive Neuroscience*, 10, 293-302.
- Golding, E. (1981). The effect of unilateral brain lesion on reasoning. *Cortex*, 17, 31-40.
- Goodwin, G. P. & Johnson-Laird, P.N. (2005). Reasoning about relations. *Psychological Review*. 112, 468-493
- Grafman, J. & Goel, V. (2002). Neural Basis of Reasoning. *Encyclopedia of Cognitive Science*. Macmillan.

- Grice, H.P. (1975). Logic and conversation. In P. Cole & J.L. Morgan (Eds.), *Syntax and semantics*. (Vol. 3: *Speech acts*, pp. 41-58). New York: Academic Press.
- Griggs, R.A. (1995). The effects of rule clarification, decision justification and selection instruction on Wason's abstract selection task. In S.E. Newstead & J.St.B.T. Evans (Eds.), *Perspectives on thinking and reasoning. Essays in honour of Peter Wason*. Hove, UK: Lawrence Erlbaum Associates.
- Hartshorne, C., Weiss, P., & Burks, A. (Eds.). (1931–1958). *Collected papers of Charles Sanders Peirce* (Vols. 1–8). Cambridge, MA: Harvard University Press.
- Hegarty, M. (2004). Mechanical reasoning by mental simulation. *Trends in Cognitive Science*, 8(6), 280-285.
- Held, C., Knauff, M. & Vosgerau, G. (Eds.) (2006). *Mental Models and the Mind: Current Developments in Cognitive Psychology, Neuroscience and Philosophy of Mind*. North-Holland: Elsevier.
- Hodges, W. (1977). *Logic*. Penguin Books, New York.
- Hodges, W. (1997). *A Shorter Model Theory*, Cambridge, CA: Cambridge University Press.
- Huttenlocher, J. (1968). Constructing spatial images: A strategy in reasoning. *Psychological review*, 75, 550 - 560.
- Johnson-Laird, P. N. (1972). The three-term series problem. *Cognition*, 1, 57-82.
- Johnson-Laird, P. N. (1983). *Mental models*. Cambridge: Cambridge University Press.
- Johnson-Laird, P. N. (1998). Imagery, visualization, and thinking. In J. Hochberg (Ed.), *Perception and Cognition at Century's End* (pp. 441-467). San Diego, CA: Academic Press.
- Johnson-Laird, P.N. (2001). Mental models and deduction. *Trends in Cognitive Science*, 5, 434-442.
- Johnson-Laird, P. N. (in press). *How we reason*. Oxford: Oxford University Press
- Johnson-Laird, P. N., & Bara, B. G. (1984). Syllogistic inference. *Cognition*, 16, 1-61.
- Johnson-Laird, P. N., & Byrne, R. M. J. (1991). *Deduction*. Hove (UK): Erlbaum.
- Johnson-Laird, P. N., Byrne, R. M. J., & Tabossi, P. (1989). Reasoning by model: The case of multiple quantification. *Psychological Review*, 96, 658-673.
- Johnson-Laird, P.N., Legrenzi, P & Legrenzi, M.S. (1972). Reasoning and a sense of reality. *British Journal of Psychology*, 62, 395-400.
- Jonides, J., Smith, E. E., Koeppe, R. A., Awh, E., & Minoshima, S. (1993). Spatial Working Memory in Humans as Revealed by PET. *Nature*, 363, 623-625.
- Kant, I. (1787/1998). *Kritik der reinen Vernunft*. Hamburg: Meiner Verlag.
- Klauer, K. C., Meiser, T., & Naumer, B. (2000). Training propositional reasoning. *Quarterly Journal of Experimental Psychology*, 53A, 868-895.
- Knauff, M. (2006). A neuro-cognitive theory of relational reasoning with mental models and visual images. In C. Held, M. Knauff & G. Vosgerau (Eds.). *Mental Models and the Mind: Current Developments in Cognitive Psychology, Neuroscience, and Philosophy of Mind* (pp. 127-149). Oxford (UK): Elsevier.
- Knauff, M., Fangmeier, T., Ruff, C. C. & Johnson-Laird, P. N. (2003). Reasoning, models, and images: Behavioral measures and cortical activity. *Journal of Cognitive Neuroscience*, 4, 559–573.
- Knauff, M., & Johnson-Laird, P. N. (2002). Visual imagery can impede reasoning. *Memory & Cognition*, 30, 363-371.
- Knauff, M., Kassubek, J., Mulack, T., & Greenlee, M. W. (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. *Quarterly Journal of Experimental Psychology*, 59, 161-177.

- Knauff, M., Mulack, T., Kassubek, J., Salih, H. R., & Greenlee, M. W. (2002). Spatial imagery in deductive reasoning: A functional MRI study. *Cognitive Brain Research*, 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). Mahwah, NJ: Erlbaum.
- Knauff, M., Rauh, R., Schlieder, C., & Strube, G. (1998). Mental models in spatial reasoning. In C. Freksa, C. Habel & K. F. Wender (Eds.), *Spatial Cognition - An interdisciplinary approach to representing and processing spatial knowledge* (pp. 267-291). Berlin: Springer.
- Knauff, M. & Schlieder, C. (2005). Spatial inference: No difference between mental images and models. *Behavioural and Brain Sciences*, 27, 589-590.
- Kosslyn, S. M. (1980). *Image and mind*. Cambridge, MA: Harvard University Press.
- Kosslyn, S. M. (1994). *Image and brain*. Cambridge, MA: MIT Press.
- Kosslyn, S. M., Alpert, N. M., Thompson, W. L., Maljkovic, V., Weise, S. B., Chabris, C. F., Hamilton, S. E., Rauch, S. L., & Buonanno, F. S. (1993). Visual mental imagery activates topographically organized visual cortex: PET investigations. *Journal of Cognitive Neuroscience*, 5, 263-287.
- Kosslyn, S. M., Ganis, G., & Thompson, W. L. (2001). Neural foundations of imagery. *Nature Reviews Neuroscience*, 2, 635-42.
- Kosslyn, S. M., & Thompson, W.L. (2003). When is early visual cortex activated during visual mental imagery? Theory and meta-analysis. *Psychological Bulletin*, 129, 723-746.
- Kosslyn, S. M., Thompson, W. L., & Alpert, N. M. (1997). Neural systems shared by visual imagery and visual perception: a positron emission tomography study. *NeuroImage*, 6, 320-334.
- Lashley, K. S. (1929). *Brain mechanisms and intelligence: A quantitative study of injuries to the brain*. Chicago, IL: University of Chicago Press.
- Leevers, H. J. & Harris, P. L. (1999). Persisting effects of instruction on young children's syllogistic reasoning with incongruent and abstract premises. *Thinking and Reasoning*, 2, 145-173.
- Logothetis, N.K. (2002). The Neural Basis of the BOLD fMRI Signal. *Phil. Trans. R. Soc. Lond.* (357), 1003-1037
- Logothetis, N.K., J. Pauls, M.A. Augath, T. Trinath and A. Oeltermann (2001). Neurophysiological Investigation of the Basis of the fMRI signal. *Nature*, 412, 150-157.
- Manktelow, K.I. (1999). *Reasoning and Thinking*. Hove (UK): Psychology Press.
- Manktelow, K. I., & Over, D. E. (1995). Deontic Reasoning (pp. 91-114). In: Newstead, St. E., & Evans, J. St. B. T. (Eds.). *Perspectives on Thinking and Reasoning. Essays in Honour of Peter Wason*. Hove (UK): Lawrence Erlbaum Associates.
- Mellet, E., Petit, L., Mazoyer, B., Denis, M., & Tzourio, N. (1998). Reopening the mental imagery debate: Lessons from functional anatomy. *NeuroImage*, 8, 129-139.
- Miller, G. A. & Johnson-Laird, P. N. (1976). *Language and perception*. Cambridge, MA: Harvard University Press.
- Möbius, P.J. (1902). *Über das Pathologische bei Nietzsche*. Wiesbaden: Bergmann.
- Newell, A. (1980). Physical Symbol Systems. *Cognitive Science*, 4, 135-183.
- Newstead, S.E. & Griggs, R.A. (1983). The language and thought of disjunction. In J.St.B.T. Evans (Ed.), *Thinking and reasoning: Psychological approaches*. London: Routledge & Kegan Paul.
- Newstead, S. E., Pollard, P., & Griggs, R. A. (1986). Response bias in relational reasoning. *Bulletin of the Psychonomic Society*, 2, 95-98.
- Oakhill et al., 1990. S.47???????
- Oaksford, M., & Chater, N. (2003). Optimal data selection: Revision, review, and reevaluation. *Psychological Bulletin & Review*, 10, 289-318.

- O'Brien, D. P. (2004). Mental-Logic theory: What it proposes, and reasons to take this proposal seriously. In J. Leighton R.J. & Sternberg (Eds.), *The nature of reasoning* (pp. 205-233). New York: Cambridge University Press.
- Palmer, S. E. (1978). Fundamental aspects of cognitive representation. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization* (pp. 259-303). Hillsdale, NJ: Erlbaum.
- Perky, C. W. (1910). An experimental study of imagination. *American Journal of Psychology* 21, 422-452.
- Piaget, J. (1936, 1963) *The origins of intelligence in children*. New York: W.W. Norton & Company, Inc.
- Plaut, D. C. (1995). Double dissociation without modularity: Evidence from connectionist neuropsychology. *Journal of Clinical and Experimental Neuropsychology*, 17, 291-321.
- Postle, B. R., Stern, C. E., Rosen, B. R., & Corkin, S. (2000). An fMRI investigation of cortical contributions to spatial and nonspatial visual working memory. *Neuroimage*, 11, 409-23.
- Putnam, H. (1961). Brains and behaviour. In R. J. Butler (Ed.). *Analytical Philosophy*. Oxford: Blackwell ( and in N. Block (1980).
- Pylyshyn, Z. W. (1984). *Computation and Cognition: Toward a Foundation for Cognitive Science*. Cambridge, MA: MIT Press.
- Ragni, M., Knauff, M. & Nebel, B. (2005) A. Computational theory of spatial reasoning with mental models. *Proc. of the 27th Annual Meeting of the Cognitive Science Society*.
- Rauh, R., Hagen, C., Knauff, M., Kuß, T., Schlieder, C. & Strube, G. (2005). From preferred to alternative mental models in spatial reasoning. *Spatial Cognition and Computation* (in Druck).
- Read, D. E. (1981). Solving deductive-reasoning problems after unilateral temporal lobectomy. *Brain and Language*, 12, 116-127.
- Revlis, R. (1975). Two models of syllogistic inference: Feature selection and conversion. *Journal of Verbal Learning and Verbal Behavior*, 14, 180-195.
- Rips, L. J. (1994). *The psychology of proof*. Cambridge, MA: MIT Press.
- Ruff, C. C., Knauff, M., Fangmeier, T., & Spreer, J. (2003). Reasoning and working memory: Common and distinct neuronal processes. *Neuropsychologia*, 41, 1241-1253.
- Russell, S. & Norvig, P. (2003). *Artificial Intelligence: A Modern Approach* (2. Auflage). Upper Saddle River, NJ: Prentice Hall.
- Salmon, W. C. (1983). *Logik*. Stuttgart: Reclam.
- Sax, L. (2003). What Was the Cause of Nietzsche's Dementia? *Journal of Medical Biography*. 11, 47-54.
- Schain, R. (2001). *The Legend of Nietzsche's Syphilis*. Westport, CT: Greenwood Press.
- Schlieder, C. (1995). *The construction of preferred mental models in reasoning with Allen's relations* (IIG-Report 5/95). Freiburg: Universität Freiburg.
- Shaver, P., Pierson, L., & Lang, S. (1975). Converging evidence for the functional significance of imagery in problem solving. *Cognition*, 3, 359-375.
- Shepard, R. N., & Cooper, L. A. (1982). *Mental images and their transformations*. Cambridge, MA: MIT Press.
- Simchowitz, S. (1900). Der kranke Nietzsche. *Frankfurter Zeitung*, 7. August 1900.
- Sperber, D., & Girotto, V. (2003). Does the selection task detect cheater-detection? In J. Fitness & K. Springer, S. P. & Deutsch, G. (1981). *Left Brain, Right Brain*. San Francisco: W. H. Freeman and Company.
- Smith, E. E., & Jonides, J. (1997). Working memory: A view from neuroimaging. *Cognitive Psychology*, 33, 5-42.

- Sperry, R. (1973). Lateral Specialization of Cerebral Function in the Surgically Separated Hemispheres. In F. J. McGuigan and R. A. Schoonover (Eds.), *The Psychophysiology of Thinking* (pp. 5-19). New York: Academic Press.
- Stenning, K. (2002). *Seeing reason*. Oxford: Oxford Univ Press.
- Stenning, K., Oberlander, K. (1995). A cognitive theory of graphical and linguistic reasoning: Logic and implementation, *Cognitive Science*, 19, 97-140.
- Störing, G. (1908). Experimentelle Untersuchungen über einfache Schlussprozesse. *Archiv für die gesamte Psychologie*, 11, 1-27.
- Suler, J. R., & Riziello, J. (1987). Imagery and verbal processes in creativity. *Journal of Creative Behaviour*, 21, 1-6.
- Teuber, H-K. (1955). Physiological psychology. *Annual Review of Psychology*, 6, 267-296.
- Titchener, E. B. (1909). *Lectures on the experimental psychology of the thought processes*. New York: Macmillan.
- Tye, M. (1991). *The imagery debate*. Cambridge, MA: MIT Press.
- Uttal, W. R. (2001). *The new phrenology*. Cambridge, MA: MIT Press.
- van der Henst, J. B. (2002). Mental model theory versus the inference rule approach in relational reasoning. *Thinking and Reasoning*, 8, 193-203.
- Vandierendonck, A., Dierckx, V. & De Vooght, G. (2004). Mental model construction in linear reasoning: Evidence for the construction of initial annotated models. *Quarterly Journal of Experimental Psychology*, 57A, 1369-1391.
- Warrington, E.K. & Taylor, A.M. (1973) Contribution of the right parietal lobe to object recognition. *Cortex*, 9, 152-164.
- Wason, P.C. (1966) *Reasoning*. In Foss, B.M. (Ed.) *New Horizons in Psychology*. Harmondsworth, Middx: Penguin.
- Wason, P. & Johnson-Laird, P. (1972). *Psychology of Reasoning: Structure and Content*. Cambridge, MA: Harvard University Press.
- Watson, J. B. (1913). Psychology as the behaviorist views it. *Psychological Review*, 20, 158-177.
- Wedin, M. V. (1988). *Mind and Imagination in Aristotle*. New Haven: Yale University Press.
- Wetherick, N. & Gilhooly, K. (1990). The figural effect in syllogistic reasoning. In K. Gilhooly, M.T.G. Keane, R. Logie & G. Erdos (Eds.), *Lines of thought: Reflections on the psychology of thinking* (Vol. 1, pp.99-108). London: Wiley.
- Wharton, C., and Grafman, J. (1998). Deductive reasoning and the brain. *Trends in Cognitive Science*, 2, 54-59.
- Whitaker, H.A., Markovits, H., Savary, F., Grou, C., Braun, C. (1991). Inference deficits after brain damage. *Journal of Clinical and Experimental Neuropsychology*, 13, 38.
- Wickens, T. D. (2001). *Elementary Signal Detection Theory*. Oxford: Oxford University Press.
- Wilhelm, O. (2005). Measuring Reasoning Ability. In O. Wilhelm, & R. W. Engle (Eds.), *Understanding and Measuring Intelligence* (pp. 373-392). London: Sage.
- Woodworth, R.S. & Sells, S.B. (1935). An atmosphere effect in formal syllogistic reasoning. *Journal of Experimental Psychology*, 18, 451-460.
- Yang Y., Braine, M.D.S. & O'Brien, D.P. (1998). Some empirical justification of one predicate-logic model. In M.D.S. Braine & D.P. O'Brien (Eds.) *Mental Logic* (pp. 333-365). Mahwah, NJ: Erlbaum.