UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE FILOSOFIA E CIÊNCIAS HUMANAS
PÓS-GRADUAÇÃO EM PSICOLOGIA

INFERÊNCIAS DA LÓGICA MENTAL PREDICATIVA NA COMPREENSÃO DE TEXTO

Jitka Soskova

Dissertação apresentada ao Programa de pós-graduação em Psicologia da Universidade Federal de Pernambuco, como parte dos requisitos necessários obtenção do grau de Doutor em Psicologia

Orientador: : Dr. Antonio Roazzi

Recife
2003

INFERÊNCIAS DA LÓGICA MENTAL PREDICATIVA NA COMPREENSÃO DE TEXTO

Jitka Soskova

Banca examinadora:

Dr. David P. O'Brien, City University of New York, Baruch College

Dr. José Aires de Castro Filho, UFC, Dep. de Fundamentos de Educação

Dra. Maria da Graça B. B. Dias, UFPE, Pós-Graduação em Psicologia

Dr. Jorge da Rocha Falcão, UFPE, Pós-Graduação em Psicologia

Dra. Mª do Carmo Tinoco Brandão de Aguiar Machado, UFPE, Dep. de Ciências Sociais

Dra. Selma Leitão Santos, UFPE, Pós-Graduação em Psicologia

Coordenação do curso:

Dra. Maria de Conceição Diniz Pereira de Lyra

A todos, que direta ou indiretamente contribuíram para a realização deste trabalho, meu MUITO OBRIGADO!!!

To everybody who directly or indirectly contributed to the realization of this project,

THANK YOU!!!

ABSTRACT

The model of mental propositional logic of M.D.S. Braine, B.J. Reiser, and B.Rumain (1984) proved to be useful for explaining logical inferences during text comprehension. The use of mental predicate logic proposed by M.D.S. Braine (1998) in text comprehension has not been tested yet. The present study verified the predictions of errorless, effortless and automatic use of the schemas of mental predicate logic (MPL). Experiment 1 presented undergraduate students with short texts containing premises of single core schemas of the MPL and showed that inferences are drawn from these schemas errorlessly. The recognition task tested the relative difficulty of the MPL schemas as compared with schemas of formal logic and paraphrases. MPL inferences were perceived as so easy that they were often confounded with paraphrases of the text and judged as significantly easier that control sentences containing inferences valid in formal logic but not predicted by mental logic. The texts of Experiment 2 contained premises for 2 to 3 core schemas of the MPL which required the subjects to apply them in a line of reasoning predicted by the Direct Reasoning Routine (DRR) of the MPL theory. The participants applied the DRR and came to the correct conclusions practically errorlessly. The recognition task confirmed the effortless application of the MPL schemas. Experiment 3 used the naming task to test the on-line application of the or-elimination schema of the MPL (For example: *everyone on the meeting wanted tea or juice; the secretary did not want juice / .: the secretary wanted tea*). The results showed that readers draw these inferences automatically at the moment the premises appear conjointly in the working memory. These findings corroborated the predictions for the use of mental predicate logic in text comprehension. The discussion focuses on the connection between mental logic and text comprehension theories.

RESUMO

O modelo da lógica mental proposicional de Braine, Reiser, e Rumain (1984) tem sido útil para prover uma base de explicação das inferências lógicas na compreensão de texto. Até ao momento, o uso da lógica mental predicativa (LMP) do Braine (1998), na compreensão do texto, não foi investigado. O presente estudo teve como objetivo averiguar as predições da LMP de que os esquemas são aplicados sem esforço, sem erro e automaticamente. No Experimento 1, foram apresentados aos estudantes universitários pequenos textos que incluíram premissas para apenas um dos esquemas centrais da LPM. Os resultados mostraram que as inferências foram concluídas praticamente sem erro. A tarefa de reconhecimento investigou a dificuldade de realizar inferências da LMP, comparando os com as inferências da lógica formal e paráfrases do texto. As inferências da LMP foram julgadas significativamente mais fáceis do que as frases-controle da lógica formal e, tão fáceis que eram erroneamente confundidas com paráfrases do texto. No Experimento 2, os textos incluíram premissas para aplicação de dois ou três esquemas Centrais da LMP, que requereu aplicação do Raciocínio de Rotina Direto (RRD). Os resultados mostraram que os sujeitos aplicaram RRD e chegaram a conclusões lógicas válidas, praticamente sem erro. A tarefa de reconhecimento confirmou a aplicação dos esquemas da LMP sem esforço. O Experimento 3 constou da tarefa de nomeação para testar a aplicação automática do esquema "ou-eliminação" da LMP (Por exemplo: *todos os participantes da reunião querem suco ou chá, a secretária não quer suco / ∴ a secretária quer chá*). Os resultados indicaram que os leitores realizam estas inferências automaticamente, no momento que as premissas encontravam-se, simultaneamente, na memória do trabalho. Estes achados confirmaram as predições da teoria da lógica mental sobre o uso das inferências lógicas na compreensão de texto. Os achados foram discutidos em relação à integração entre lógica mental e as teorias de compreensão do texto.

Preface

Mental logic theory explains how people in their everyday reasoning integrate information coming at different moments from different sources, choose and apply a suitable logical schema, and draw inferences that go beyond the original information. This thesis has the objective to test the suitability of Braine's (1998) Mental Predicate Logic model of for logical inferences drawn during text comprehension.

In the Chapter 1 the theoretical background of mental logic theory is reviewed. First, deductive reasoning is defined and compared with other types of reasoning. Next, the relation between reasoning and logic is explored. Do the rules of formal logic describe human reasoning? Different theoretical approaches provide different possible answers to this question: The research on heuristics and biases focuses on the part of human reasoning that is not in line with the rules of logic. Content-bound theories state that human reasoning cannot be described with the use of abstract logical rules but is rather closely tied with specific pragmatic contents and contexts of the reasoning process. Without discrediting these arguments, there is empirical evidence showing that human reasoning has a logical domain-general basis. Two theoretical approaches are based on such a premise: the mental models theory and mental logic theory. The basic difference between these two theories is related to the type of representation of the information which they believe human logical reasoning works with. The mental models theory assumes that we reason from iconic, model representation of the premises. Mental logic theory deduces that logical inferences are drawn from a propositional format of representation. Empirical evidence gives more support to the mental logic theory. Also, the mental logic theory is more suitable for explaining logical inferences in text and discourse comprehension than the mental models approach.

The idea that people posses mental logic has been suggested several times in the history of deductive reasoning research, one of the most famous proposals having been made by Piaget. Piaget's as well as other proposals of mental logic are compared with the theory of mental logic of Braine and O'Brien (1998a).

The mental logic theory of Braine, O'Brien and their colleagues has gradually gained substantial empirical support. The theory consists of two parallel models: Mental Propositional Logic, and its extension to Mental Predicate Logic. These models are described in detail, reviewing the basic parts of the mental logic theory: the logical schemas, and the reasoning

program. The theory makes several basic predictions about the schemas of mental logic: These schemas are to be applied basically without error, without effort, automatically, universally across cultures, and are acquired early in childhood. Substantial empirical evidence supports these predictions, although the research up till today focused mostly on the Mental Propositional Logic model. The Mental Predicate Logic model has not been thoroughly tested yet.

As the theory claims that it is suitable for explaining logical inferences during text comprehension, the next section of the Introduction focuses on the theories of text comprehension. Inferences, the core of the comprehension process, are analyzed in greater detail, reviewing all the important factors of the message and of the reader that can influence their occurrence.

The last part of the introduction reviews the research of Lea and his colleagues (Lea, 1995; Lea, O'Brien, Fisch, Noveck, & Braine, 1990), as it provides evidence of the use of the Mental Propositional Logic schemas in text comprehension. Lea also showed that such schemas are applied automatically. This issue seems to be quite controversial, as logical inferences are forward inferences and their automatic application during reading is not predicted by most of the text comprehension theories.

The objectives of the experiments were set to clarify these issues as well as to test other predictions mentioned above, using the Mental Predicate instead of Propositional Logic model. Four experiments were designed in order to test the following: First, the errorless and effortless application of the individual schemas; second, the linking of several schemas in the reasoning program; and third, the automatic drawing of logical inferences during reading. Each experiment is described in a separate chapter.

Experiment 1a in Chapter 2 applied the validity task to assess the error rate of inferences related to a single schema introduced in short texts.

Experiment 1b in Chapter 3 applied the recognition task that permitted to infer the relative difficulty of application of the individual mental logic schemas. Similarly to Experiment 1a, the eight core schemas were tested in short stories, one at a time.

Experiment 2 in Chapter 4 introduced several schemas of the model conjointly in short texts in order to see whether the readers are able to link the logical inferences according to the reasoning program, and come to the correct conclusion without much effort. This experiment applied again the validity and recognition task to test the errorless and effortless use of the

mental logic schemas.

Chapter 5 describes Experiment 3 that aimed to detect whether one of the schemas of the mental Propositional Logic model is applied automatically during reading at the moment the necessary premises are conjointly held in the working memory. The naming task based on priming of the result of the logical inferences provided evidence relevant to this objective.

The results of each of the experiments in the previous four chapters are discussed individually. In addition to that, Chapter 6 presents a general discussion of all the experiments. The findings of the experiments are compared with each other and with relevant theories, and analyzed in a broader framework of the relation between logic and language. Issues necessitating further research and clarification are pointed out. The mental logic model proposal is evaluated in relation to the competing theories of deductive reasoning. The conclusion confirms that the mental logic theory provides a useful framework for understanding our everyday reasoning and the use of logic in text comprehension.

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION

Several competing theories attempt to explain human deductive reasoning. Between them, the mental logic theory proposed by Braine, Reiser and Rumain (1984), further developed in Braine and O'Brien (1998a), has gradually gained substantial empirical support. The theory consists of two parallel models: Mental Propositional Logic, and its extension to Mental Predicate Logic. Empirical evidence confirms that the propositional logic model can account for logic inferences people routinely make in text and discourse comprehension (Lea, 1995; Lea et al., 1990). This project intends to test the use of mental predicate logic model for logical inferences drawn during text comprehension.

Types of Reasoning

Human reasoning lies at the core of higher-level cognition processes. Most of contemporary cognitive psychology sees reasoning as processing of information and manipulation of mental representations. Analyzing reasoning from this point of view means "…understanding how the organism transforms, organizes, stores and uses information arising from the world in sensorial data or memory" (Sternberg, 1994, p.xv).

What are the different kinds of human reasoning? Literature on reasoning can be basically divided into two big groups: studies of deduction and studies of induction. In deductive arguments a set of premises is defined, and the problem solver has to apply logical inferences to come to a conclusion. If the premises are true and the inferences are valid, the conclusion is true. Validity is therefore a function of the relation between the premises and conclusion. It is based on the claim that the premises provide absolute grounds for accepting the conclusion.

Arguments in which the premises provide only limited grounds for accepting the conclusion are inductive arguments. We use induction when we try to discover rules governing a certain situation, detecting which properties generalize in the required manner and which do not. Inductive arguments are not logically certain; they can only be evaluated for plausibility or reasonableness. In contemporary cognitive science inductive processes are viewed as

responsible for generating concepts.

Human reasoning can also be categorized using a different terminology. A widely used division of reasoning is that between formal and informal reasoning, applied for resolution of well-defined and ill-defined problems, respectively (Garnham & Oakhil, 1991). Formal, or well-defined reasoning is identical to deductive reasoning. Informal reasoning is used when the problems are poorly defined; that means, when either the starting state of the situation, the desired final configuration, or the reasoning steps to be taken to overcome the problem "barrier" are not well defined (e.g., Frensch & Funke, 1995). The problem solver has to resort to heuristics, using mainly induction to solve these problems.

Several attempts have been made to summarize all the research in the area of human reasoning. One of the most extensive projects was concluded by Carrol (1993). Carrol reanalyzed 460 data sets from a wide variety of tests of reasoning conducted by many different authors, and submitted them to factor analysis. The results suggest that human reasoning can be described by three factors: Sequential Reasoning, Induction, and Quantitative Reasoning. The factor of Sequential Reasoning points to deductive reasoning, which Carrol (1993) describes as "…reasoning involved in tasks or tests that require subjects to start from stated premises, rules or conditions and engage in one or more steps of reasoning to reach a conclusion that properly and logically follows from the given premises." (p.62). The second factor resulting from Carrol's analysis – Induction – involves discovering the rules that govern the materials, or illustrate certain similarities or contrasts. The third factor - Quantitative Reasoning - operates in tasks or tests that require subjects to reason with concepts that involve quantitative or mathematical relations. Quantitative Reasoning can be either inductive or deductive or both. What distinguishes it from the other two factors is that it relies heavily on skills acquired by formal schooling.

A neuroimaging study conducted by Goel, Gold, Kapur, and Houle (1997) revealed that different brain loci are activated during inductive and deductive tasks. Also, brain structures responsible for deductive and probabilistic reasoning appear to be substantially distinct. (Osherson, Perani, Cappa, Schnur, Grassi, & Fazio, 1998). The brain processes involved in deductive reasoning are therefore distinguishable from processes involved in other types of reasoning.

All this evidence clearly confirms that logical (deductive, sequential, formal, or well defined) reasoning is one of the basic types of human reasoning, clearly distinguishable from

other types of thinking.

Reasoning and Logic

Are Humans Rational or Irrational?

The idea that people have a logical mind has been present for millennia. Aristotle laid down the rules of formal logic as a complete set of rules governing human thought, and this idea remained unchanged for 2000 years. This historical view of logic saw logic as manipulation of symbolic forms, and early psychological research assumed that human mind proceeds in the same way.

The initial research of human deductive reasoning focused on studies of classical syllogisms, where people were asked to assess logical arguments or generate valid conclusions. Research of deductive reasoning has undergone a dramatic acceleration in the 1960s, thanks to the influence of the Piagetian theory. Piaget incorporated the philosophical tradition of *logicism* (Henle, 1962) into his theory of cognitive development, proposing that adults eventually developed a formal operational thinking on the basis of abstract logical structures (Inhelder, & Piaget, 1958). Many influential psychologists at this time were prone to argue that human reasoning was invariably and inevitably logical.

During the 70s and 80s research on heuristics and biases in decision making and deductive reasoning became fashionable, investigating all the logical errors and fallacies people commit in everyday reasoning. The encountered differences between formal logic conclusions and everyday reasoning strategies and outcomes, reverted the scale in favor of the idea that human reasoning does not, or hardly ever, follows the lines of formal logic. Therefore, people started to be seen as rather irrational computing devices. Nevertheless, even during these "dark times of rationality", several defenders of the logical base of human reasoning continued to develop the theories of mental logic, competing fiercely with those supporting the theory of mental models. Research on deductive reasoning became a major field of cognitive psychology by the end of 20th century.

The research of human reasoning can therefore be seen as a competition between two camps: One group (e.g., Braine & O'Brien, 1998a; Braine et al., 1984; Lea, 1995; Rips, 1983) starts from the assumption that humans are rational and aim to discover what is the mental logic like: which logical schemas are used correctly, when and how are they applied,

where is the interface between logical and non-logical reasoning processes, how does mental logic interact with other areas of human reasoning, such as language comprehension, problem solving, e.t.c. Another group of scholars (e.g., Cheng & Holyoak, 1985; Cosmides, 1989; Evans, 1982; Wason, 1983) propose that the areas where human reasoning corresponds to formal logic rules is extremely restricted or focus their research on the aspects of reasoning that divert from the rules of formal logic. This research stresses the irrational part of human reasoning.

The rationality debate stems from the obvious fact that sometimes people behave rationally and other times irrationally. We have developed mathematics, all the different fields of science and high technology including complex computing systems – all that based on empirical evidence and logic as the basic principles of scientific knowledge. In everyday life we plan our actions, anticipate their logical consequences, and are able to take rational decisions. On the other hand, we do not behave that way all the times: We take risks without calculating the exact probability of success, and let sudden impulses and emotions take lead in our decisions.

The evidence of irrationality led Evans (2002) to criticize the deduction paradigm and argue that formal logic should not even be taken as a measure of human rationality: "The study of deductive reasoning, defined narrowly by logic, appears to the casual observer to stand apart from normal cognitive psychology, which deals in the nature of mental processes without judging their correctness relative to a normative standard" (p.978).

In logical arguments when premises are true and the correct logical schema is applied, the conclusion must be true. Therefore, in a logical reasoning task, one must: assume premises true, base reasoning on these premises without considering any prior knowledge, and accept the logician's interpretation of the words *some, if, or, not*, regardless of how such words are used in ordinary discourse. In everyday reasoning, this is indeed not common: Human thinking is often based on beliefs, in which there are varying degrees of confidence leading to conclusions that may be probabilistic and provisional in nature (for example, Evans, 1982; Klauer, Munsch, & Naumer, 2000). A number of heuristics and biases influence our everyday reasoning (see, for example, Evans, 1982, 1993; Kahneman & Tversky, 1972, 1982; Tversky & Kahneman, 1973, 1974, 1983). Moreover, content and context of the reasoning process may influence the success or failure of the reasoning task (Cheng & Holyoak, 1985; Griggs & Cox, 1982; Wason, 1977; Wason & Shapiro, 1971; and others).

On the other hand, even researchers who focus on the irrational part of human reasoning admit that "…there is evidence of an irreducible minimal deductive

competence for which psychologists must provide an explanation" (Evans, 2002, p.982).

Mental logic theories are aiming to provide such explanation. The basic argument of the defenders of rationality is that the evidence that certain behavior is irrational does not lead to the conclusion that there is no mental logic at all. Noveck, Lea, Davidson, and O'Brien (1991) explore this argument, concluding that there is no reason to abandon the classical view that human nature includes rationality. This rationality includes mental logic that accounts for our basic logical intuitions. For example, when my doctor says that I have caught either dengue or a flue and a blood test would exclude dengue, then my mental logic allows me to infer that I have a flue. O'Brien (1993) lists evidence for such basic logical competence and posits that failure to solve complex reasoning problems does not count as evidence against mental logic.

However, the debate between defenders of rationality and irrationality of human thought was (and continues to be) rather fruitful; as a result, many different approaches to the study of human deductive reasoning have been taken. The following section will review the main theories in the area of deductive reasoning.

Content-bound Theories

The issue whether reasoning reflects domain-specific or domain-general processes has been debated so thoroughly, that some scholars see this distinction as the chief organizing contrast of human reasoning (Markman & Gentner, 2001). Domain-general theories, like mental logic theory or mental models theory, look for universal rules and models of human reasoning. These theories aim to capture phenomena that are generalizable for all deductive reasoning or at least for a certain general category of reasoning problems (like, for example, syllogisms, or spatial relationships, or propositional logic problems). On the other hand, there are theories claiming that content and context are fundamental to reasoning and that human thinking processes are specified by the content of the problems to be solved (e.g., Cheng & Holyoak, 1985; Newell & Simon, 1972; Tooby & Cosmides, 1989). Some have taken an extreme opposite of the domain-general theories, arguing that there is no utility to a general notion of representation or process. One such view is the situated cognition approach that assumes that all thinking is fundamentally context governed (Carraher, Carraher, & Schlieman, 1989; Suchman, 1987).

Most theories that study the influence of content and context on reasoning competence refer to the Wason selection task (Wason, 1960, 1968), "probably the most investigated single

problem in the whole literature on the psychology of reasoning" (Evans, 2002, p.980).

In this task people are told that there are four cards on the table and each has a letter on one side and a number on the other. The following rule applies to the cards: If there is a vowel on one side of the card then there is an even number on the other side. The four cards displayed might have the values A (vowel), T (consonant), 4 (even number), and 7 (odd number) on their visible sides. The instructions ask the reasoner to choose only those cards that need to be turned in order to decide whether the mentioned rule is followed or not.

This task requires the reasoner to understand first the falsifying principle: It is necessary to turn over cards that might falsify the rule. In this case the falsifying combination is a vowel together with an odd number, so the only cards that lead one to verify the rule are the A and the 7. However, this solution is very hard to discover and only 10% or less of university students participating in such an experiment find the correct answer. A typical choice is the cards A and 4, which were interpreted by Wason as confirming an erroneous verification principle.

The results of the experiments with the card selection task led Wason to question the Piagetian theory, arguing that people readily succumb to fallacies.

Successive research showed that there is a remarkable difference in performance on this task when presented to the subjects in a algebraic content version (cards with letters and numbers) and a version taken from real-life context. One of such contexts is as follows: Four persons are sitting in a bar, one is drinking beer, one is drinking Coke, one is 16 years old and another is 25. The task is to verify whether the drinking age rule is followed: "If a person drinks alcohol, then he/she must be over 19 years old." Sixty to seventy percent of the subjects were able to give correct answers in the real-life content versions against the 10% of correct solutions in the algebraic version (e.g., Cheng & Holyoak, 1985; Griggs & Cox, 1982; Wason & Shapiro, 1971). Cheng and Holyoak (1985) concluded that reasoning relies on generalized knowledge structures, which are acquired by induction in situations reflecting permission, obligation or causality relations. The knowledge structures, or *schemata*, contain context-sensitive production rules, which, once established, determine the inferences in situation that trigger the schema. These rules are pragmatic, rather than syntactic, in that they serve the achievement of certain goals in a given context.

Another big group of content-bound theories are the *ecological theories*. The ecological approach to reasoning, offered by evolutionary psychologists, specifies inference mechanisms

for narrowly defined classes of social situations. The first major contribution of this kind to the theories of reasoning was made by Cosmides (1989) who proposed that some contexts, developed during evolution as forms of social exchange, substantially facilitate deductive reasoning. These mechanisms are based on social contracts that have a conditional form: If one takes a benefit, one must pay a cost. Gigerenzer and his colleagues (e.g., Gigerenzer & Hug, 1992) specified one of such mechanisms in a hypothesis of a cheater detection algorithm.

Over recent years this evolutionary account of reasoning has been developed significantly (e.g., Cosmides & Tooby, 2000), although it has been mostly restricted to the selection task.

These theories raised the question whether adaptive reasoning derives directly from evolution, or is primarily based on learning in the lifetime of the individual. Cosmides presented in her work an evolutionary argument for the operation of an innate reasoning module in such contexts.

Another empirical route in human reasoning has emphasized non-logical processes and biases in deductive reasoning (e.g., Evans, 1972, 1982, 1993), heuristics, and errors in probabilistic reasoning (e.g., Kahneman & Tversky, 1972, 1982; Tversky & Kahneman, 1973, 1974, 1983). A bias is defined as a systematic influence of some logically irrelevant feature of the task (Evans, 2002).

One of the content effects of long standing in the history of deduction research is the so-called *belief bias:* People endorse far more conclusions as valid when they accord with prior belief than if they disbelieve it (e.g., Evans, Baston, & Pollard, 1983; Klauer et al., 2000). There is an interesting interaction between logical validity and believability - the belief effect is far more marked on invalid conclusions. For example, the experiments of Evans, Baston, and Pollard (1983) showed that in case of a valid syllogism subjects accepted a believable conclusion in 92% of the cases and an unbelievable conclusion in 46%. In case of invalid syllogisms the effect of belief bias was significantly stronger: The subjects accepted an invalid but believable conclusion in 92% and an invalid and unbelievable conclusion only in 8% of the cases.

Klauer et al. (2000) discuss at which stage of information processing does the bias arises, whether during input, processing or output. First, a belief could distort the interpretation and representation of the premises. Second, it could directly affect the reasoning process. Finally, the belief could bias the respond stage. Klauer et al. analyzed results from previous

studies and designed several experiments in order to specify better the mechanism of belief bias. The results indicated that belief bias might come about through response bias.

Evans (2002) argues that people do not treat premises necessarily as true or false but seem to have degrees of beliefs in premises and conclusions. People respond to degrees of belief rather that making absolute deductions about truth and falsity.

Many more different biases have been detected. In the context of the research on Wason selection task a *matching bias* has been described as a tendency to pay attention to features of the problem that have a lexical match with items explicitly named in the conditional statement (Evans, 1972). Evans (1982) suggests that in doubt people prefer to endorse negative conclusion rather than affirmative in the interest of caution, which constitutes the *negative conclusion bias.* Also, people have difficulty in processing double negation (*double negation effect* described in Evans, Clibbens, and Rood, 1995), and a general difficulty in processing implicit negations (Schaeken & Schroyens, 2000). The influence of context on some reasoning tasks led Stanovich (1999), in Evans (2002) to propose a *fundamental computational bias* comprising a tendency to contextualize all problems with regard to prior knowledge and belief. Syllogistic reasoning might also biased by syntactic factors, like the nature of the quantifiers used in premise (Chater & Oaksford, 1999), and the order in which terms are mentioned (Johnson-Laird, & Bara, 1984).

However, all these theories seem to ignore that people are able to think logically in everyday reasoning and text and discourse comprehension, even when there is none of the content bound schemas involved. The scholars working in the area of content-bound theories offer no comprehensive theory explaining correct responses. Experiments working with the mental logic model of Braine and O'Brien (1998a) show that a certain part of human logical reasoning is universal, not bound by any specific content or context.

Mental Models Theory

The theory of Johnson-Laird and his colleagues (e.g., Johnson-Laird, 1980, 2001; Johnson-Laird & Byrne, 1991; Johnson-Laird, Byrne, & Schaeken, 1992) posits that during reasoning subjects construct a *mental model* of the given information, and then they reason from the model. For instance, given the premises A is inside B and B is inside C, a subject imagines a state of affairs corresponding to the premises; the conclusion that A is inside C can then be read off from the image. According to this theory, reasoning consists of three stages: forming a

model as a representation of the situation consistent with the premises (*comprehension stage*), reading off a tentative conclusion (*description stage*), and then testing the conclusion by trying to construct alternative models consistent with the premises (*validation stage*). The last validation stage is the only deductive stage where reasoners assumingly search for alternative models of the premises that falsify the putative conclusion. If a falsifying model is found, it is necessary to return to the second stage to search for another putative conclusion. If no falsifying model is found, the conclusion is considered true.

Mental models theory has been extensively developed to account for responses to categorical syllogisms (e.g., Johnson-Laird, 1980), inferences about spatial and lexical relations (Johnson-Laird, 1980), relational, and quantified reasoning (Johnson-Laird, 1999), temporal reasoning (Schaeken, 1996), modal and probabilistic reasoning (Tabossi et al., 1999 in Johnson-Laird, 2001). It has been urged that it may suffice also for propositional reasoning (Johnson-Laird, 1999; Johnson-Laird et al., 1992). In a recent update Johnson-Laird (2001) explains how mental model can account for reasoning about probability, necessity and possibility in his theory: A conclusion is necessary if it holds in all the models of the premises. If it holds in a proportion of models, its probability is equal to that proportion, and if it holds at least one model, the conclusion is possible given the premises.

One of the basic principles that underlies mental models theory is the principle of truth: mental models represent what is true according to the premises, but by default not what is false. Johnson-Laird (2001) postulates that individuals by default do not represent what is false, in some cases they make 'mental footnotes' about the falsity of clauses, and if they retain the footnotes they can flesh out mental models into fully explicit models, which represent clauses even when they are false.

Johnson-Laird (2001) explains that mental models are not imitations of real-world phenomena; they are simpler. "…each model represents a possibility…, it captures what is common to the different ways in which the possibility might occur" (p. 434). A model is an iconic representation where its parts correspond to the part of what it represents and its structure corresponds to the structure of the possibility. Models can represent relations among three-dimensional entities or abstract entities; they can be static or kinetic, they underlie visual images although many components are not visualizable.

A mental model works with an iconic representation of the premises; nevertheless, Johnson-Laird (1983) admits that reasoning must start with propositional representations encoded in a verbal format. These propositional representations are neither integrated

nor elaborated with other information held in memory, therefore they do not enable a person to perform operations like generalization or inference on the information encoded in them. The propositional representations are mapped into mental models by process called *procedural semantics*: For each new item of incoming information a search is made to ensure that the proposition is consistent with any earlier information encountered. If an appropriated model can be found to accommodate the proposition, the model will be cued, applied, and perhaps modified. If no appropriate model can be found, the relevant procedures will be employed to construct a mental model from scratch (construction *de novo*). The evaluation of a whether a proposition is true or false is conducted in terms of *meaning* attached to the proposition, not syntactic rules of logical truth tables.

O'Brien (1993) points out some weak points of Johnson-Laird´s theory. First, it does not provide a clear description of what a mental model is. Mental models could be images but clearly are intended to go beyond images. Second, according to the theory, reasoning does not proceed from propositions but from iconic representations of the mental models. As models are not propositional, they do not include variables, but always refer to specific instances. Therefore, the mental models theory has some problems with representations of, for example, the universality of a certain property (*all p are q*). How do we create a mental model of *all* the possible *p* (for example, even those which we are still to encounter in future)? Another difficulty arises with representations of negative instances. The theory states that we do not make models of negative instances. How comes that when we are presented with the premises *not both p and q; p*; then we are able to infer *not q* without any difficulty? (see Lea & Mulligan, 2002). In sum, there are many sorts of propositions that are difficult, if not impossible to represent with mental models. Moreover, Lea (1995) showed that the mental model theory could not provide correct predictions of logical inferences in text processing.

Braine et al. (1995) provides a detailed analysis and discussion of the difference between predictions made by mental logic and mental models theory in relation to several experiments. For example, in reasoning tasks, where subjects are presented with a set of premises and a conclusion and asked to write down everything that follows, several intermediate inferences are routinely reported by the subjects. Should the reasoners read the conclusion from the model, no intermediate inferences would be reported, only a *true* or *false* answer.

Some of the basic predictions of how the reasoning based on mental models works are not supported by empirical research. For example, Evans, Handley, et al. (1999) could not find evidence for the prediction that during the validation stage subjects will search for

counterexamples of the model of premises together with the putative conclusion. Doubts about conclusion validation are also expressed in the study of Newstead, Handley, and Buck (1999).

The mental models theory also contrasts with theories of language comprehension in respect to the representational format. Kintsch's (1998) Construction-Integration model of text comprehension is based on a network of propositions. The propositional format of representation is, according to Kintsch (1998), very well suited for drawing inferences, generalizing or other cognitive operations, including deductive inferences (Lea, 1995).

Mental Logic Approach

The idea that natural reasoning incorporates logic was extensively developed by Piaget and his colleagues (e.g., Inhelder & Piaget, 1958, 1964). Piaget derived his theory from philosophical and psychological tradition of *logicism* (Henle, 1962) that defends logic as the basis for rational human thought. During the 1960s and 1970s the influence of Piagetian theory was enormous. The Piagetian theory proposed that adults eventually developed formal operational thinking on the basis of abstract logical structures. Piaget's model of mental logic is based on a semantic truth-functional system of 16 propositional relations, together with the transformations of the INRC group (identity, inversion, reciprocity, and correlativity, respectively). Each connective, such as *and, or, not, if…then,* is defined as a disjunction of all the possibly true conjuncts of the component proposition, that is, of all the lines of the truth table corresponding to the use of the connective.

Piaget's proposal was criticized from several points: some experiments show that children, who did not reach the stage of formal operational thought already have some ability to reason logically (e.g., Braine & Rumain, 1983; Brainerd, 1977; Ennis, 1975, 1976; Fisch, 1991, in O'Brien, 1993). Other researchers argue that adults fail to solve certain logical tasks. (e.g., Evans, 1982, 1991, 1993; Wason, 1977). O'Brien (1987) points out that Piaget's model does not include predicate logic working with quantifiers even though some of his tasks are clearly requiring this competence. Piaget's suggestion, that the structure of concrete-operational thought corresponds to logic of classes, based on a complete set of formal logic rules, overestimated the similarity between formal and mental logic.

In recent years a consensus has been developed that human logical reasoning uses a set of elementary deductive steps, which are only a subset of formal logic rules. It proceeds

through the application of sound inference schemas. Inference schemas are procedures that specify which propositions can be derived from particular assumed propositions. The soundness of these propositions ensures that propositions derived from true assumptions inherit that truth. Several theories proposed such a set of logical inference schemes, which would account for these elementary deductive steps, forming a "natural logic" (Braine, 1978; Braine & O'Brien, 1998a; Braine et al., 1984; Osherson, 1975; Rips, 1983, 1994)

The mental logic theory of Braine and O'Brien (1998a) assumes that human beings are in principal rational and they use logical reasoning routinely in their everyday activities. The logic adopted by these theories is seen as *propositional* and *intentional* (O'Brien, 1993).

What does it mean that logic is intentional? Propositions take truth values, that is, a proposition can be either true or false. Sentences are not propositions. Any sentence can be true or false depending on who is the speaker, and what is the context of the utterance. For example, the sentence "My computer is not working now" can be true if it is asserted by a certain person in a certain situation and false in another context. Once the utterer and the circumstance of a certain sentence are known, the sentence becomes a proposition. The speaker knows which computer and which moment he is speaking about therefore he/she, as well as any informed listener, are able to make judgments about truth or falsity of this proposition. For example, he/she can believe that this proposition is true, doubt that it is true, claim that it is false, and so on. Such propositional activities concern intentional states of affairs.

Assumptions and conclusions of logical arguments are propositions and do take truth values. Images are not propositional; although an image might be an accurate or an inaccurate representation, it can neither be true or false. Propositions that refer to images, though, are true or false.

Propositions can be atomic or compound, that is, atomic propositions can be negated, or joined in a conjunction, disjunction, conditionality, and so forth. Forming a compound proposition requires an inference – one does not observe a disjunction or a conditional. We use compound propositions in our everyday reasoning so we need to have an account of how we form and use them, which is the subject of investigation of the mental logic model.

Because propositions have truth values, their inference procedures must be truth preserving. This means that, given a set of propositions assumed true, further propositions drawn from them by logical procedures also would be true. Logicians refer to this property as

logical soundness: A set of inference procedures is sound if, and only if, given a set of true propositions, the inference procedures will provide true conclusions only.

Propositions can be true or false but in ordinary logical thinking we proceed only from propositions assumed true. It seems not to have sense to draw conclusions from untrue or contradictory premises. Thus ordinary reasoning proceeds not from premises but from assumptions – premises assumed true (Braine & O'Brien, 1991; Braine et al., 1984; O'Brien, 1993; Politzer & Braine, 1991). This observation points out one of the differences between standard and mental logic: In standard logic it is possible to make a valid argument starting from false or contradictory premises. An argument is valid unless there is a possible assignment of truth values such that its premises taken conjunctively are true while its conclusion is false.

O'Brien (1993) summarizes that human logical reasoning is propositional, and not a mere manipulation of symbols. Propositional activities, such as asserting, doubting, believing, denying, and so forth, require intentional state of affairs. Mental logic applies sound inference procedures to propositions assumed true, inferring new propositions that inherit that truth.

Braine and O'Brien (1998a) argue that a complete theory of propositional reasoning must include more then logic inference schemas. Schemas provide only a repertoire of steps available to the reasoner; it does not itself generate a chain of reasoning. Braine and O'Brien propose two more components to the theory: The first necessary component is the comprehension mechanism, which understands natural language and translates the logical schemas in their semantic representations. Any type of linguistically based reasoning must start with decoding the verbal information into the representations used in schemas. This is the point where mental logic theory meets with text processing theories (Lea, 1998).

The second component is a reasoning program consisting of routines and strategies that can put together a chain of inferences, selecting the schema that is to be applied at each point in the reasoning.

A possible additional element is a set of non-logical or quasi-logical fallback procedures that determine a response when the reasoning program fails to deliver a solution to a problem.

Braine and O'Brien (1998c) explain that all these components – the schemas, the comprehension mechanism, the reasoning program, and the fallback procedures – are interrelated.

The mental logic approach does not claim that all of human logical reasoning is described by a set of content-free formal rules. O'Brien (1993) explains that for certain types of

problems people probably do use mental models for reasoning. Inferences from mental models would cohabit with inferences from mental logic. Moreover, mental logic schemes cohabit easily also with content-bound reasoning processes, studied extensively by Cosmides (1989), and Cheng and Holyoak (1985). Content-specific schemas, like permission scheme or obligation scheme, act to enrich the inferences of the mental logic content-free schemas (Braine & O'Brien, 1991).

Rips' theory (1983, 1994) is very similar to the ML theory of Braine and O'Brien, as it shares important general characteristics. Rips suggest that reasoning involves a stage of translation of the information from the surface structure into *mental sentences* (Rips' term), and that reasoning takes place in the language of mental sentences. This is essentially similar to the Braine and O'Brien's theory of language of thought. Rips' deductive reasoning theory also presents the three basic components: a representational system and a set of inference schemas that are part of a reasoning program. There are, however some significant differences both in the representational system and in the reasoning procedures.

Rips' (1994) system, called PSYCOP, is a revision and extension of an earlier system, ANDS (Rips, 1983). PSYCOP has two kinds of schemas: forward and backward. Forward schemas resemble the core schemas of the Braine and O'Brien's (1998c) model in that their use is not restricted – they apply whenever they can be applied. The use of backwards schemas is restricted by goals to deduce certain propositions and by features of the problem environment. In problems without a conclusion only the forward schemas can be applied – every such schema that can be applied does so and SPYCOP recycles until no further inferences can be drawn. In problem with a conclusion to be evaluated the forward schemas apply first until no further inference can be drawn using them: then, if the conclusion has not already been deduced, it is set as a goal and backwards schemas that can apply do so (with some priorities in order of application) in a depth-first search until derivation is found or PSYCOL runs out of search paths.

Braine and O'Brien's theory distinguishes between core schemas and feeder schemas in order to avoid infinite loops of processing in the reasoning program. PSYCOP avoids infinite loops by omitting schemas that could give rise to them (like *and-introduction* – feeder schema no.7 on page 23). In summary, PSYCOP includes some schemas that Braine and O'Brien's system does not and omits others. Rips' rules are based on a division between forward and backward types, the latter being dependent for their operation on the existence of appropriate goals. It also lacks a mechanism that detects contradiction between a premise set and a

conclusion given. Braine (1998) also criticizes the way PSYCOP handles quantifiers in predicate logic arguments – Rips' system is based on domain-general quantification, whereas Braine's quantification is domain specific, which provides a better fit to natural language. Experiments presented in Braine and O'Brien (1998a) are discussed in relation to the Rips' model of deductive reasoning and conclude that the theory of Braine and O'Brien (1998c) and Braine (1998) is able to provide better predictions for the results. Rips' model sometimes predicts inferences subjects do not regularly draw and other times omits inferences that appear in the protocols of the subjects.

In the next chapter the mental logic theory of Braine & O'Brien (1998a) will be described in more detail. The logical schemes that are included in the theory will be listed and the main aspects of the reasoning program outlined.

Mental Predicate and Propositional Logic

The ML theory follows the example of work in standard logic by Gentzen (1935/1964). Gentzen suggested that a certain basic set of inference schemas form a "natural deduction system". These schemas allow people to integrate logical information concerning conditionals, conjunctions, disjunctions and negations, conveyed in English language as *if*, *and*, *or*, and *not*, respectively. Building on these ideas Braine et al. (1984) proposed a system of propositional inference schemas, which could describe human deductive reasoning in everyday life as well as in text and discourse comprehension. This system was further developed by Braine and O'Brien (1998c).

Mental logic schemes form a limited subset of formal Aristotelian logical inference rules. This subset includes, for example, the *modus ponens* scheme: (if p then q; p∴ therefore q), because this scheme is used naturally and effortlessly in everyday reasoning. On the other hand, another rule associated with *if – modus tolens* (if p then q; not q∴ therefore not p) – is omitted from mental logic, because the use of this rule is far from universal. About half of college-students fail to give a correct answer on this type of problems (e.g., Rumain, Connell, & Braine, 1983).

The ML theory distinguishes between direct reasoning and indirect reasoning. Direct reasoning is the natural and automatic part of our everyday reasoning, as well as text and discourse processing, as it is applied routinely and effortlessly. Direct reasoning can be captured by a set of core, feeder and incompatibility schemas applied according to a Direct Reasoning

Routine (Braine & O'Brien, 1998c).

Indirect Reasoning Program includes strategies, which account for the pragmatic type of reasoning. The indirect reasoning schemas are applied less easily and automatically, as they require a certain level of cultural sophistication, such as practice or formal education.

The reasoning programs use the following types of schemas: core schemas, feeder schemas, incompatibility schemas, and some others. The core schemas would be applied automatically when premises are active in the working memory and they are considered true. The feeder schemas are used automatically when their output feeds the core schemas, and the incompatibility schemas define contradictions. The application of feeder schemas is limited by the condition of providing input for the core schemas. This is because the nature of these schemas could give rise to infinite loops of processing. For example, should the feeder schema *p; q / ∴ p and q* , or the feeder schema *p and q / ∴ p* apply automatically without such restriction, they could feed each other over and over again infinitely.

Braine and O'Brien (1991, 1998a) further developed the ML theory and extended the system of propositional mental logic also to the predicate mental logic model. Predicate logic describes logical inferences with propositions in the form of predicate – argument. The propositions include quantifiers, like *all, some, none*, and they deal with the scope of these quantifiers.

In the list below the relevant schemas from the mental logic model are presented in the propositional version in the first row, followed by the predicate logic version(s). Braine (1998) developed a new notation for predicate logic, which corresponds better to the structure of the natural language, especially with regard to the quantifiers and their scope.

Mental Predicate Logic Notation

"S[All X]" means that the X or Xs all satisfy the condition S. "[Some X]" indicates some unspecified X or Xs. "[a]" indicates an argument of any form whatever that is not in the scope of a negation. "[q X]" means X modified by any quantifier, e.g., *all, each, many, few, some*, etc. (but not *no, none, not any*); "[q X]" includes pro-forms with quantified antecedents (e.g., "[PRO-All X]", "[PRO-Some X]").

"[PRO-a]" is a pro-form whose antecedent is a. "S_1[All X] OR S_2[PRO-All X]" means that each X is either S_1 or S_2; "S_1[Some X] & S_2[PRO-Some X]" means that some X or Xs are S_1 and the same X or Xs are also S_2. Read "[All X: S[PRO]]" as 'all the Xs that are such that

they satisfy S', i.e. 'the Xs that are S'.

In "NEG S[All X]" [All X] is outside the scope of the negation. "[~. . .~]" means that "[. . .]" is inside the scope of a negation, e.g. "NEG S[~All X~]" means 'it is not the case that all the Xs are S'; similarly, "NEG S[~Some X~]" means that no Xs are S ('it is not the case that some X is S'), as opposed to "NEG S[Some X]" 'Some X is not S'. Thus NEG S[Some X] ¶ NEG S[~All X~].

The notation is briefly illustrated in the first two schemas; a complete description is available in Brian (1998).

Mental Logic Schemas

Core Schemas:

1. p or q; ~p / ∴ q

(a) S1[All X] OR S2[PRO-All X]; NEG S2[α]; [α] ⊆ [X] / ∴ S1[α]

(b) S1[All X] OR S2[PRO-All X] / ∴ S2[All X: NEG S1[PRO]]

Schema 1 is called *or-elimination* schema, and its propositional version is shown on the first line: When one of two alternatives is false, the other must be true. The first of the predicate-logic versions (on the second line) can be rendered in English as “All of the Xs satisfy predicate S1 or they satisfy S2; some particular object or set of objects, α, does not satisfy S2; α is included among the Xs; one can conclude that α satisfies S1.” (The “PRO” notation usually is realized as a pronoun. For discussion of the notational system, see Braine, 1998). The second predicate-logic version (on third line) can be rendered as “All of the Xs satisfy predicate S1 or they satisfy S2; one can conclude that all of the Xs such that they do not satisfy S1 satisfy S2.”

An example of a problem of the sort that uses the first predicate logic version of this schema could be as follows: *The boys either played with girls or fought with girls; Tom and Dick did not play with girls.* Application of the schema leads to the inference that *Tom and Dick fought with girls.*

An example of a problem that uses the second predicate logic version of this schema could be: *The boys either played with girls or fought with girls.* The inferred conclusion is that *The boys who did not play with girls fought with girls.*

2. If p THEN q; p / ∴ q

(a) S[All X]; [α] ⊇ [X] / ∴ S[α]

(b) NEG S[~Some X~]; [α]⊆ [X] / ∴ NEG S[α]

At the propositional level Schema 2 is standard logic's *modus ponens*. The first of its predicate-logic versions can be rendered as "All of the Xs satisfy S; some particular object or set of objects, α, is among the Xs; it can be concluded that α satisfies S. An example of a problem that uses the first predicate-logic versions of the schema would be the proposition *The girls all wore red jeans*. The application of the schema leads to the conclusion *The girls in sneakers wore red jeans*.

The second version can be rendered as "There is no X that satisfies S; some particular object, α, is included among the Xs; it can be concluded that α does not satisfy S." This schema could be illustrated by the inference *None of the boys wore striped shirts / ∴ Sam and Henry did not wear striped shirts*.

The tildes around "Some X" indicate that it is within the scope of the negation and can be instantiated. "NEG S[Some X]" would indicate that "some X is not S." One could not then conclude that S is not an X. Note that the meaning of the quantifier in a schema is given by the inferences about instantiation, i.e., which objects can or cannot satisfy the predicate.

3. ~(p & q); p / ∴ ~q

At the propositional level Schema 3 is a standard's logic *not-both* schema.

(a) NEG E[~Some X :S1[PRO-All X] &S2[PRO]~]; S2[α];[α]⊆ [X] / ∴ NEG S1[α]

Example: There were no boys who wore sandals and blue jeans ; The boys that played with Mary wore blue jeans / ∴ The boys that played with Mary did not wear sandals.

(b)NEG (S1[All X] & S2[PRO-All X]]) / ∴ NEG S2[All X: S1[PRO]]

Example: There were no boys who wore sandals and blue jeans / ∴ The boys that wore blue jeans did not wear sandals.

4. p OR q; If p THEN r; If q THEN r / ∴ r

S1[All X] OR S2[PRO-All X]; S3[All X: S1[PRO]]; S3[All X: S2[PRO]] / ∴ S3[All X]

Example: All the cars in the lot have stickers or the guards tow them away. The cars that have stickers are Toyotas. The cars that the guards tow away are Toyotas / ∴ All the cars in the lot are Toyotas.

5. p or q; IF p THEN r; IF q THEN s / ∴ r OR s

S1[All X] OR S2[Pro-All X]; S3[All X: S1[PRO]; S4[All X; S2[PRO]] / ∴S3[All X] OR S4[PRO-All X]

Example: All the cars in the lot have stickers or the guards tow them away. The cars that have stickers are Datsuns. The cars that the guards tow away are Toyotas / ∴ The cars in the lot are all Toyotas or Datsuns.

6. IF p OR q THEN r; p / ∴ r

no predicate-logic version

Feeder schemas

7. p; q / ∴ p & q

S_1[All X]; S_2[All X] / S_1[All X] & S_2[PRO-All X]

Example: The boys wore blue jeans; The girls played with the boys / The boys wore blue jeans and the girls played with them.

8. p & q / ∴ p

S_1[q X] & S_2[PRO-q X] / S_2[q X]

Example: Many of the boys wore blue jeans and the girls played with them / The girls played with many of the boys.

Incompatibility Schemas:

9. p; NEG p / ∴ INCOMPATIBLE

(a) S[All X]; NEG S[q X] / ∴ INCOMPATIBLE

Example: The boys are all wearing sneakers; Some of the boys are not wearing sneakers / INCOMPATIBLE.

(b) S[q X]; NEG S[All X] / ∴ INCOMPATIBLE

Example: Some of the boys are wearing sneakers; None of the boys are wearing sneakers / INCOMPATIBLE

10. p OR q; NEG p & NEG q / ∴ INCOMPATIBLE

(a) S_1[All X] OR S_2[PRO-All X]; NEG S_1[q X] & NEG S_2[PRO-q X] / ∴ INCOMPATIBLE

Example: The cars all had stickers or the guards towed them away; Some of the cars did not have stickers and the guards did not tow them away / ∴ INCOMPATIBLE.

(b) S_1[q X] OR S_2[PRO-q X]; NEG S_1[All X] & NEG S_2[All X] / ∴ INCOMPATIBLE

Example: One of the boys wore a striped or a spotted shirt; None of the boys wore a striped shirt and none wore a spotted shirt / ∴ INCOMPATIBLE.

Other Schemas:

11. Given a chain of reasoning of the form:

Suppose p

- - -

- - -

∴ q

One can conclude: IF p THEN q

Given a chain of reasoning of the form:

Suppose S_1[Some X]

- - -

- - -

∴ S_2[PRO-Some X]

One can conclude: S_2[Any X: S_1[PRO]])

12. Given a chain of reasoning of the form:

Suppose p

- - -

- - -

INCOMPATIBLE

One can conclude: ∴ NEG p

13. S[All X]; [α] ⊆ [X] / ∴ S[Some X])

Example 1: Many of the girls in spotted shirts wore red jeans / Some of the girls wore red jeans.

Example 2: All the girls played with boys in green jeans / All the girls played with children in green jeans.

14. S[All X] / S[Some X]

Example: All the girls wore green jeans / Some of the girls wore green jeans.

Reasoning Program

The mental logic schemas are used in reasoning programs. Braine and O'Brien (1998c) explain direct reasoning as follows: The reasoner starts with the premises, makes an inference from the premises and then successively makes further inferences from the premises together with the propositions already inferred, until the conclusion or a proposition incompatible with it is reached. A special case of direct reasoning is an if-then statement. In this case, the reasoning is considered direct if the reasoner first adds the antecedent of the conclusion to the premises as an additional starting formula and takes the consequent as the conclusion to be reached. Then he/she solves the reformulated problem by successive inferences as described earlier, starting with the premises together with the antecedent. The Direct Reasoning Routine, defined by Braine and O'Brien (1998c) is described in Table 1.

The program applies both to problem situations where there is a conclusion given whose truth is to be evaluated, and when no such conclusion is given (i.e., when subjects are just making inferences from the information they have). In the latter case, the Preliminary Procedure and the Evaluation Procedure of the Direct Reasoning Routine are inapplicable: Then the routine comprises the Inference Procedure only.

The program begins with the Direct Reasoning Routine. The routine terminates when the conclusion is evaluated, or when no new propositions are generated by the Inference Procedure. If the routine terminates without evaluating the conclusion, then available indirect reasoning strategies are applied.

The functioning of the DRR can be illustrated on the example of two parallel problems presented in O'Brien et al. (1994). Premises referred to letters written on an imaginary blackboard:

Problem 1		Problem 2	
(a)	*N or P*	(a)	*Not both Z and S*
(b)	*Not N*	(b)	*If H then Z*
(c)	*If P then H*	(c)	*If P then H*
(d)	*If H then Z*	(d)	*Not N*
(e)	*Not both Z and S.*	(e)	*N or P*

The mental-logic theory makes the following predictions for these two problems. On Problem 1 the DRR applies Schema 1 of Table 1 to the first two premises (a) and (b), deriving *P*; Schema 2 of Table 1 then is applied when premise (c) is read, deriving *H*, which allows Schema 2 to be applied again when premise (d) is read, deriving *Z*, which allows Schema 3 to be applied when premise (e) is read, deriving *not S*. In Problem 2 the same premises are presented in the reverse order. When the premises of Problem 2 are read, the DRR is unable to apply any of the core schemas until all of the premises have been read, but then it applies Schema 1 to premise (e), *N or P*, and premise (d), *Not N* (now the last two premises encountered), to infer *P*, which then allows Schema 2 to be applied (to the output of Schema 1 together with premise (c) *if P then H*) to infer *H*, which then leads to application of Schema 2 again to derive *Z* when premise (b) *If H then Z* is considered, and then finally to application of Schema 1 when premise (a), *Not both Z and S*, is considered to derive *not S*. The prediction thus is that the two problems will lead to exactly the same lines of reasoning, with the same inferences being made in the same order on the two problems. Thus, because the order of the predicted inferences is determined by the order in which the core schemas become available and not by the order in which the premises are presented the two problems will have the same output even though the information is received in the same order. O'Brien et al. found that the order in which participants wrote down inferences on both problems corresponded to those predicted by the DRR.

In general, indirect reasoning may require some heuristics and learned strategies to find the successful line of reasoning. Examples of indirect reasoning strategies adopted from Braine, O'Brien (1998) are presented in Table 2. The difficulty of a problem will likely reflect the difficulty of the specific heuristic or strategy used.

Table 1

Direct Reasoning Routine

PRELIMINARY PROCEDURE. (i) If the given conclusion is an *if-then* statement, add the antecedent to the premise set[a], and treat the consequent as the conclusion to be tested. (ii) Use the Evaluation Procedure to test the conclusion (the given conclusion or the new one created at Step (i)), against the premise set. If the evaluation is indeterminate, proceed to the Inference Procedure.

INFERENCE PROCEDURE. For each of the Core schemas (Schemas 1 through 7, Schema 1 in the left-to-right direction only), apply it if its conditions of application[b] are satisfied or if its conditions of application can be satisfied by first applying one or a combination of the Feeder schemas (Schemas 8, 9, 14, and 1 in the right-to-left direction). Add the propositions deduced to the premise set. When there is a conclusion to be evaluated, use the Evaluation Procedure to test the conclusion against the augmented premise set; if the outcome of the evaluation is indeterminate, repeat the Inference Procedure. When there is no conclusion to be evaluated, just repeat the Inference Procedure. (In executing the Inference Procedure, no schema is applied whose only effect would be to duplicate a proposition already in the premise set.) In reading out conclusions inferred, one-time use of Feeder schemas is optional[c].

EVALUATION PROCEDURE. To test a given conclusion against a premise set, respond "True" if the conclusion is in the premise set or can be inferred by applying one or a combination of the Feeder schemas; respond "False" if the conclusion, or an inference from it by Schema 9, is incompatible (by Schemas 10 or 11) with a proposition in the premise set, or with a proposition that can be inferred from the premise set by applying one or a combination of the Feeder schemas.

Notes: [a] The "premise set" at any point comprises the original premises together with any propositions that have been added by the Preliminary and Inference Procedures.

[b] The conditions of application of a schema are satisfied when the premise set contains propositions of the form specified in the numerator of the schema; to apply the schema is to deduce (generate) the corresponding proposition of the form specified in the denominator of the schema. Schemas that are equivalences are applicable when part (or all) of a proposition in the premise set matches the form specified on one of the sides of the equivalence; application consists in substituting the proposition of the indicated form for the matching part.

[c] For example, if propositions *a* and *b* are inferred independently, it is optional to use Schema 8 and read these out as *a* & *b*; if a conjunction is in the premise set, it is optional to use Schema 9 to read out the conjuncts separately.

Adapted from M. D. S. Braine and D. P. O'Brien (1998c), p.82.

Table 2

Some Indirect Reasoning Strategies

SUPPOSITION-OF-ALTERNATIVES STRATEGY. If the premise set contains a disjunction (or if one is obtained by applying Schema 9), and if some of the propositions of the disjunction do not occur as antecedents of conditionals in the premise set, then suppose each of these in turn and try to derive a conditional with it as antecedent, using Schema 12.

STRATEGIES OF ENUMERATION OF ALTERNATIVES A PRIORI: E.g., if the premise set contains one or more conditionals of the form *If p then* . . . or *If not p then* . . ., add the proposition *p or not p* to the premise set and return to the Inference Procedure.

REDUCTIO AD ABSURDUM STRATEGY. Limited form: If there is a conjunction or disjunction embedded within a premise proposition or within the conclusion, then suppose the conjunction or disjunction as per Schema 13 and use the evaluation procedure to test its compatibility with the premise set; if the evaluation is "false", add the negation of the conjunction or disjunction to the premise set; use the evaluation procedure to test the conclusion against the augmented premise set, and if the evaluation is indeterminate, return to the Inference Procedure.

Stronger form: To test the falsity of a conclusion given, or of any proposition embedded within a premise or the conclusion, add the negation of that proposition to the premise set and try to derive an incompatibility as per Schema 13, using the Inference Procedure, any available other strategies, and the Evaluation Procedure.

Adapted from M. D. S. Braine and D. P. O'Brien (1998c), p.83.

Pragmatic Principles

The ML theory assumes that reasoning starts with a comprehension process that translates the information into semantic representations. Comprehension process is influenced by many pragmatic factors. These factors affect the information from which inferences are drawn, so to understand the result of logical inferences one must consider the possible pragmatic factors affecting the comprehension process. Braine and O'Brien (1991) summarize these factors into three groups: *pragmatic reasoning schemas* that affect conditional inferences (mentioned in chapter Content-bound Theories), *gricean cooperative principles*, and *invited inferences.*

The first factor concerns how the content of propositions affects the way they are constructed. The pragmatic reasoning schemas suggested by Cheng and Holyoak (1985) define a class of plausible interpretations into which the content of a conditional could be assimilated. For example, the *permission schema* defines rules of the form "If an Action *x* is to be done, then Condition *y* must be satisfied". In general, a sentence with content that falls in the domain of the schema will be constructed with a representation that includes, but is richer than, the representation that could be gotten from the lexical entry alone. For example, the contrapositive ("If the Condition is not satisfied, then the Action cannot be done") becomes more accessible to the subject than if permission schema is not elicited. The schema would facilitate drawing of the appropriate conclusions from this sentence.

The second factor that influence the comprehension process is based on theory of Grice (1975, 1979), which is considered "one of the most valuable descriptions of conversational logic" (Cooren & Sanders, 2002, p.1045). Grice's basic assumption, called the Cooperative Principle, is that every speaker intends to meet the conversational demand. Thus, when a speaker's utterance doesn't seem to satisfy the conversational demand at that moment, the interpreters infer further propositions that the speaker's utterance implicates, in order to uphold the Cooperative Principle. These further propositions are inferences that are made by interlocutors to understand each other when what is literally uttered differs from what it would have been incumbent on the speaker to say in order to fulfill the conversational demand at that moment.

For example, it is well known that a conditional *if p then q*, invites the inference *if not p then not q* (Geis & Zwicky, 1971), which leads to standard fallacies in conditional reasoning. Rumain, Connell, and Braine (1984) showed the influence of conversation implicatures in an experiment, where they used syllogisms like, for example:

If it is raining, Fred gets wet.

Given the second premise *Fred gets wet*, subjects erroneously conclude *It is raining.* Grice (1975) explains that in everyday communication people expect to be provided with exactly the right quantity of information they need to be able to draw the conclusion their interlocutor expects them to draw. That means there should be provided neither more nor less information than necessary. In this case, the subjects assumed that rain is the only reason for Fred getting wet. Rumain et al. (1984) added another premise to the syllogism:

If it is raining, Fred gets wet.

If it is snowing, Fred gets wet.

Fred gets wet.

In this case, the subjects were much more likely to argue that there is no valid conclusion to such a problem, which is the correct answer to this problem.

Grice (1975) presents four specific maxims for talk in conversational information exchanges that allow us to access the cooperativeness of the utterance:

- *Quantity* ("Make your contribution as informative as it is required; do not make your contribution more informative than is required"),
- *Quality* ("Try to make your contribution one that is true - do not say what you believe to be false and do not say that for which you lack adequate evidence"),
- *Relation* ("Be relevant"),
- and *Manner* ("Be perspicuous: avoid obscurity of expression, avoid ambiguity, be brief, be orderly.").

When a maximum is flouted, the interlocutor takes it to mean that the speaker might have something more in mind that meets the conversational demand. All these maxims imply a structure of expectation in which somebody is supposed to produce a specific performance. In this connection, it is not surprising that some scholars, like Sperber and Wilson (1995) reduced all the maxims to that of relation ("be relevant").

The third pragmatic principle is that logical particles often carry invited inferences (Geis & Zwicky, 1971). For instance, *if p then q* often invites the inference *if not p then not q*, an or-statement invites the inferences "not both", and a statement of the form *Some F are G* invites *Some F are not G*. According to Geis and Zwicky , people make such pragmatically invited

inferences unless they have reason to believe them inappropriate, that is, unless they are countermanded either explicitly, or implicitly by some property of the discourse content or context.

As Braine and O'Brien (1991) suggest, invited inferences could be seen as special cases of conversational implicatures.

Errors of Reasoning

In order to be accepted as a valid theory of deductive reasoning, the mental logic theory has to be able to explain also reasoning errors. Braine and O'Brien (1998a) list three possible sources for reasoning errors: comprehension errors, heuristic inadequacy errors, and processing errors.

A comprehension error is an error of construal of the premises or of the conclusion: The starting information used by the subject is not that intended by the problem setter. Henle (1962) pointed out that many errors occur because people misunderstood or misrepresent the problem, even if they apply the correct logical inference to it. As argued earlier, subjects do not accept the logical task when they focus on the truth or falsity of the conclusion instead of relating the conclusion purely to the preceding premises (see chapter Content-bound Theories). The same restriction counts for premises: Ordinary reasoning proceeds not from premises, but from assumptions, that is, from premises that are assumed true (O'Brien, 1993; Leblanc & Wisdom, 1976; e.t.c.). Other pragmatic principles that influence the interpretation of the problem are listed in the previous chapter Pragmatic Principles.

Heuristic inadequacy errors occur when the subject's reasoning program fails to find a line of reasoning that solves a problem, that is, the problem is too difficult for the subject. Problems using the schemas of the ML model, formulated in low to moderate complexity should not trigger heuristic inadequacy errors.

Processing errors comprise lapses of attention, errors of execution in the application of schemas, failure to keep track of information in working memory, and the like. Braine & O'Brien (1998a) assume that the probability of a processing error increases with problem complexity, but overall tends to be low and essentially vanishes in the simplest problems where processing load must be assumed to be minimal.

Evidence for the Mental Logic Model

The four principal claims of the theory of mental logic, confirmed by empirical evidence, can be summarized as follows:

ML theory can predict which reasoning problems were solved and which were not solved,

ML theory can predict the relative difficulty of those problems that were solved,

ML theory can predict the intermediate steps made in lines of reasoning leading to solution of those problems.

The mental logic schemas are used relatively errorlessly and effortlessly and universally.

Several studies have provided direct tests of the propositional mental logic model's claim that the core and feeder schemas together with the direct-reasoning routine are readily available.

In one of the first empirical tests of the theory, Braine et al. (1984) asked the subjects to solve two types of problems. In one type of problems subjects were provided propositions that refer to letters written on an imaginary blackboard, for example, *'On the blackboard there is either a T or an X.'*. On another type of problems the propositions refer to boxes containing toy animals and fruits, for example, *'In this box there is either a lion or an elephant.'*

In the first step of the experiment the subjects were presented assumptions together with conclusions to be evaluated. The mental logic model predicted successfully which problems were solved correctly, relative response times on simple problems and subjects' judgments about relative problem difficulty. Lea et al. (1990), Fisch (1991), in O'Brien (1993), O'Brien and Lee (1992), O'Brien, Braine, and Yang (1994), and Braine, O'Brien, Noveck, Samuels, Lea, Fisch, and Yang (1998) used two types of tasks to test the model. In one task the subjects were presented problems with conclusions to be evaluated, and they were asked to write down every intermediate inference they drew on the way to evaluating the conclusion. In another task the problems presented assumptions without any conclusions, and subjects were asked to write down everything they could infer from the assumptions. On both sorts of problems, the mental logic model predicted successfully which inferences subjects wrote down, and the order in which they were written down. Subjects almost always wrote down the output of the core schemas in the order predicted by the model, but they almost never wrote down the output of the feeder schemas, even though the output of the core schemas often depended on

the previous output of the feeder schemas. A few subjects, perhaps responding to the instructions to write down everything, wrote down the output of the feeder schemas, and when they did, this output was in the order predicted by the model.

These experiments confirm that the inferences that the mental propositional logic model predicts should be made effortlessly and routinely were made routinely and with little apparent effort. These findings are not limited to American undergraduate students. Fisch (1991), in O'Brien (1993), found that 9 and 10 year olds make the basic mental propositional logic model inferences as easily as do adults, and O'Brien and Lee (1992) found the same results when American college students were presented problems in English and Hong Kong college students were presented the same problems in Chinese. Although there may be other inferences also made routinely, those included in the model appear secure.

There is substantial empirical evidence in favor of the propositional mental logic model, but up to date only few studies have addressed the predicate mental logic model. Yang, Braine, and O'Brien (1998) conducted an initial test of the mental predicate logic schemas, using monadic predicates (i.e., predicate that takes a single argument) and dyadic predicates (predicates that take two arguments). They examined whether the propositional mental logic model can predict the relative difficulties of problems.

Subjects were presented with a set of problems in which the schemas and number of reasoning steps varied. All problems concerned the presence or absence of beads of different shapes, materials and colors in a sack. In each problem, one or more facts were given, followed by a conclusion. Subjects had to decide whether this conclusion was true or false, given those facts, and then to rate its subjective difficulty right after they marked a truth value. The problems were all soluble by the Direct Reasoning Routine, but they varied in the number of inferential steps required.

The results showed that subjects made relatively few errors overall, and almost none on one-step problems. The mental logic approach has been able to construct a set of logical reasoning problems that people solve routinely. Two predictions were made concerning the relative difficulty of the problems. First, it was predicted that the number of reasoning steps required to solve a problem would correlate with observed mean difficulty ratings, and second, it would correlate even more strongly with the sum of the difficulties of the schemas applied in the reasoning steps to solve each problem. Both of these predictions were confirmed.

Yang et al. (1998) suggest that the relative difficulty of individual core schemas depends

on the type and number of logical particles used in the schema summarized the results concerning relative difficulty of the individual schemas of mental predicate logic. Application of a schema with only one particle from among *and, if, some* and *not* (e.g., schema 2 on page 18) is perceived as relatively effortless; application of a schema with *or* is perceived to take somewhat more effort (e.g., schema 1 (b) on page 17), and application of a schema that presents two particles (e.g., schemas 3(b), 5, and 6 on page 18) is perceived as relatively more difficult.

Mental Logic and Language of Thought

Braine and O'Brien (1998b) stated that there is more to mental logic then a set of logical schemes and a reasoning program. The verbal information coming at different times and from different sources must be decoded into the representations used in schemas. Braine and O'Brien (1998b) agree with Fodor (1975) that the reasoner translates statements from natural language into a representational system of *language of thought* and then reasons in this syntax or language of thought.

The above mentioned theories argue that there must be an innate representational format of language of thought available, which is "filled in" by words specific to a given language, when the child starts to acquire the language. For example, to represent a situation when two things are alternative, an innate syntax for disjunction would be filled in, for example, by the word *or* in English, *ou* in Portuguese, or *nebo* in Czech.

The language of thought is the syntax of the basic elements of reasoning of a linguistic sort – cognitive primitives. Cognitive primitives are elements of reasoning, which are universally available for all cultures and languages, acquired very early in the child development, and used effortlessly in everyday reasoning. Braine and O'Brien (1998b) propose that such cognitive primitives would include, for example: the predicate/argument distinction (a format to distinguish between entities and their properties), conjunction (representing the situation when two properties are present), disjunction (two things are alternative), negation (an expected property is absent), the true/false distinction, and some quantifying and referring devices. The listed cognitive primitives are the constituting elements of the mental logic. Braine and O'Brien suggest that the syntax of thought include mental logic.

Braine and O'Brien (1998b) see three possible empirical routs that could investigate the existence and nature of mental logic and syntax of thought: The first route leads through the

reasoning processes: identifying the elementary inference schemes and testing their automatic, errorless and effortless use. The second possibility is a linguistic approach: checking different languages and cultures to see if the cognitive primitives identified by Braine and O'Brien indeed constitute the universals of language. The third possibility is a developmental route: we would like to know whether the elements of mental logic are included in what is semantically and syntactically most primitive in language acquisition.

All three routes of investigation are gradually gaining empirical support. The current project investigates mental logic through reasoning and is designed to test the claims of automaticity, and errorless and effortless use of the predicate mental logic theory schemas in text comprehension.

Text Comprehension

A great amount of effort was invested and many pages were written on the topic of text comprehension, "…one of the elusive, controversial constructs in cognitive science." (Graesser, Singer & Trabasso, 1994. p.373). There is a common agreement on the basic issues, like that comprehension is a complex cognitive process that consists of construction of multi-level representations of texts and improves when the reader has adequate background knowledge to assimilate the text.

According to the construction-integration model of Kintsch (Kintsch, 1998; Kintsch & van Dijk, 1978), comprehension is proposed to proceed in cycles. In each cycle, the understander considers the succeeding chunk of text, that provides a limited number of additional information.

Each cycle consists of two phases: construction and integration. During construction, the message from the text is coded into propositions that are arranged in a network. These propositions could be explicit text ideas, coherence-preserving inferences, and generalizations that capture the gist of the situation. In this phase the network includes also close associates of the text ideas, not necessarily relevant to the gist of the text. For example, Swinney (1979) presented readers with text about spying. When the word "bug" appeared in the text, the associate "insect", a relevant associate, though inappropriate in this context, was briefly activated. In this initial phase of text processing, all relevant associates are connected to the ideas they validate.

During the integration phase, activation spreads among the ideas of the current processing cycle. Activation tends to accumulate among the ideas that are highly interconnected with one another. As a consequence, any irrelevant propositions, that were introduced during construction, such as "bug" in the presented example, receive negligible activation. After each processing cycle, a few highly activated propositions are carried over for further processing in limited-capacity working memory buffer.

Although Kintsch (1998) takes the proposition as the unit of analysis at the semantic level, he argues that texts must be represented at multiple levels of representation. At the surface level, texts must be represented as strings of words, phrases, and sentences, and at the deeper level as a model of the situation described by the text (situational model). Thus, the mental representation of a story might consist simultaneously of a string of words and sentences; a semantic representation that captures the meaning of these words in terms of a propositional structure; and a situational model that might invoke images, spatial and temporal information, as well as propositional elements. In addition to local representation, texts have a global representation, corresponding to the notion of gist, and formalized in the construction-integration model by the concept of a macrostructure (Kintsch, 1998).

Text-base provides only a shallow representation; deeper understanding is achieved only after the reader draws all the necessary inferences to be able to construct causes and motives that explain why events and actions occurred. In general terms, to comprehend a text means to understand the global message, or the point. This is however impossible without understanding the pragmatic context of the text, such as who wrote the text, why it was written, who read the text and why it was read.

Inferences

It seems rather clear that understanding of any discourse or utterance raises and falls with the capacity to draw appropriate inferences. Inferences are the core of the text understanding process.

For McKoon and Ratcliff (1992), inference is defined as any piece of information that is not explicitly stated in a text. This definition includes relatively simple inferences as well as complex elaborative inferences that add new propositions to the text as well as those that connect pieces of text.

Singer (1990) doesn't include in the study of inferences the low level processes

contributing to spoken and written word recognition. “True” inferences are the processes of adding propositions to the resulting network representing the text.

Another important distinction is the difference between inference and activation. The words in a message activate related concepts. For example, the word nail activates the concept NAIL, which activates a related concept HAMMER. The activation of the concept HAMMER may contribute to the encoding of the argument HAMMER in the text base, but it doesn’t guarantee it. Even associated concepts unrelated to the meaning of the message are activated. Swinney (1979) showed that during encoding of an ambiguous word, all possible meanings of such a word are activated. Within a second after reading such a word in the text, only the appropriate meaning is still active and assumed to be incorporated in the network. An inference would not be a merely activated concept; it is the proposition, which was added to the network.

Singer (1990) also eliminates the high-level processes, such as complex problem solving, reasoning, and the construction of space-situational models from the analysis of inferences. Finally, reconstructive discourse memory, although inferential in nature, is seen as a strictly retrieval phenomenon.

Kintsch’s (1998) position in this question is quite different: The problem-solving processes, starting from premises from which some conclusions are drawn, can justly be called inferences. A different set of processes occurs when gaps in the text are being bridged by a piece of preexisting knowledge. Kintsch sees these processes as knowledge retrieval. Both proper inferences and knowledge retrieval can be either automatic and unconscious, or controlled/ conscious. But the basic distinction is that generation, or inference, produces new information by deriving it from information in the text by some inference procedure, whereas retrieval adds information already pre-existing in the long-term memory.

Kintsch also discusses whether inferences proceed from mental models as defended by Johnson-Laird and his colleagues (e.g., Johnson-Laird, Byrne, & Schaecken, 1992) or from rules as suggested by mental logic models. Kintsch agrees with the suggestion of Braine and O'Brien (1998a) that inferences based on rules and on imagery cohabit with each other. In truly symbolic, abstract domains inferences must be by rule, and there where the basic representation is an action or perceptual representation inferences must involve operations on mental models. “Inferences in the linguistic domain, where the representation is at the narrative level, may be based on mental models (to the extent that the language is embodied…) but also could involve verbal inference rules” (Kintsch, 1998, p.192).

Types of inferences

Singer (1990) categorize the implications of a message as are either logical or pragmatic in nature. Logical implications are 100% certain, and are used on some identifiable set of rules, for example, the rules of arithmetic. Pragmatic implications are probable but not certain. They are based on our knowledge and experiences rather than on formal rules.

The logical implications of a sentence are more likely to be true that the pragmatic implications. Nevertheless, this doesn't mean that logical implications are more likely to accompany comprehension than are pragmatic inferences. The logical-pragmatic distinction helps to categorize inference types but doesn't give answer to the question of which inferences are drawn in which situation.

There is a central problem of research in this area: does inference processing accompany comprehension or does it occur only later? Two alternative hypotheses can be formulated: some inferences could be drawn during comprehension, or on-line. Others will occur off-line only during retrieval at the test time, for example, during the sentence recognition test. McKoon and Ratcliff (1992) term the on-line inferences as *automatic* and the off-line inferences as *strategic.*

The literature on inferences distinguishes between *bridging* and *elaborative inferences*, also called *backward* and *forward inferences*, respectively. A bridging inference is a proposition or argument that is constructed to bridge two sentences, such as:

a. The tooth was pulled painlessly.

b. The dentist used a new method.

In the given example the bridging inference to be drawn is that the dentist was the agent who pulled the tooth. Constructing a bridge between the current sentence and the preceding discourse preserves the coherence of the discourse.

Singer (1990) illustrates the importance of drawing a bridging inference is by a contrasting sequence: *The tooth was pulled painlessly. The tailor used a new method.* Because there is no obvious knowledge that bridges TAILOR and PULLING A TOOTH, the sequence strikes us as incoherent.

Elaborative inferences add information to the text base but they are not essential to the

coherence of the message. For example, if we read only the sentence *The tooth was pulled painlessly***,** we could draw the elaborative inference that it was pulled by a dentist.

Singer (1990), Thorndyke (1977) and many others suggest that there is a link between when and what type of inferences are drawn: bridging inferences would accompany the encoding process, whereas elaborative inferences not. On the other hand, the more recent investigations show that this distinction between the type of inference and the moment when it is drawn is not so simple (e.g., Graesser et al., 1994; McKoon & Ratcliff, 1992). Graesser et al. (1994) list thirteen different types of inferences and compare several different theoretical positions, each of which predicts smaller or bigger number of inference classes to be drawn on-line. The analyzed theories agree that at least inferences required for local coherence (bridging), and those based on easily available information in the memory are drawn automatically (*minimalist theory*, McKoon & Ratcliff, 1992).

Graesser et al. (1994) state that each of the classes of inferences could potentially be generated on-line if the experimenter tuned the instructions, task, and materials properly. The goal of the reader and the amount of time and cognitive resources the reader has strongly determine the amount of inferencing that occurs (Kintsch, 1998).

Message Characteristics that Guide Inference Processing

The form of a message constrains the inferences that are computed. Some of the important factors that determine what inferences are drawn are: coherence requirements, aspects of causal connections in the text, thematic structure, distance among ideas, interestingness, linguistic cues, and the presence of negations.

Inferences are used to establish local and global coherence of a text. There is a general agreement that bridging inferences are those necessary for maintaining local coherence, whereas elaborative or strategic inferences serve the reader to construct globally coherent model of the text. McKoon and Ratcliff (1992) explain that local coherence is defined for those propositions that are in working memory at the same time; in other words, propositions that are no farther apart in the text then one or two sentences. These are assumed to proceed automatically. Global inferences connect widely separated pieces of textual information, linking the elements of the story grammar or explicit pieces of information into an overall causal chain or network. McKoon and Ratcliff's (1992) minimalist hypothesis states that inferences necessary to provide global coherence are not drawn automatically, but other researchers (e.g. Graesser et al., 1994) feel that this position underestimates the amount of inferencing that

occurs during normal reading.

Inferring causal connections among the ideas are especially important for discourse comprehension. Singer (1990) presents considerable experimental evidence that inferences about the causal connections are drawn during comprehension. For example, Keenan et al. (1984), in Singer (1990), collected norms that specified four degrees of causal relatedness of information in the text and demonstrated that the reading time varied as a function of the degree of causal relatedness.

Trabasso and Sperry (1985) analyzed the causal structure of a story counting the number of implicit and explicit causal connections in which each story proposition participated. Readers had to judge the importance of these propositions. The results revealed that rated importance varies systematically with the number of causal connections of a proposition. Because only a subset of the causal connections was explicitly stated in the story, the readers must have inferred the others. The propositions that are causally connected with a large number of other propositions are perceived as important. According to the construction-integration model of text comprehension of Kintsch (1998) such propositions would receive more activation than others and are therefore maintained as part of the macrostructure, or gist, of the story.

Inference processing is concentrated on the ideas related with the gist, or the thematic ideas in a discourse (Singer, 1990). Walker and Mayer (1980) presented readers with different versions of a text where the same ideas were presented either as thematic or peripheral. Later, the participants were to judge the truth of implied facts. They were more accurate in their judgments if the inferences were derived from thematic facts, as opposed to the situation when inferences were related to peripheral facts.

An explanation of why inference processing focuses on thematic ideas could be found referring to the construction-integration model of Kintsch (1998). According to the model, the thematic ideas constitute high-level proposition of the text, which are likely to be adopted as macro-propositions. As such they retain in the working memory, available to be inferentially combined with subsequent discourse ideas.

Singer (1990) suggests that the distance between discourse ideas should affect the likelihood of constructing bridging inferences between them. It should be easier to identify a referent of a word when the referent appeared recently in the text, than when it has appeared

much earlier. Again, according to the model of Kintsch (1998), most of the propositions are deleted from the working memory shortly after they are presented in the message. If the distance between two propositions is big, it is less likely that the earlier one will remain in working memory to provide the reference necessary to complete bridging inferences.

Linked to the motivation of the understanders, the interestingness of the message may guide the inference processing. Schank (1979) demonstrated that topics as power, sex, money, and death, plus events of personal relevance, are inherently interesting. Moreover, unexpected, surprising events are also interesting. Kintsch (1980) discusses interestingness in terms of the relation between the understander's knowledge and the events described in the discourse. To be most interesting, the discourse has to be somewhat but not overly familiar to the reader.

Several authors (Kintsch & Keenan, 1973; Millis & Just, 1994; Murray, 1997) pointed out the crucial role of linguistic cues in the text. Words like *but* or *because* are specifically requiring a bridging inference, they invite the readers to make a specific connection. Their presence in the text significantly speeded up the inference making.

Negation can affect various levels of comprehension. High level processes are affected by the presence of negatives (e.g., Clark & Chase, 1972; Just & Carpenter 1971). Fragment completion time and verification latencies were slowed when the sentences participants were completing contained a negation. This effect was obtained even when manipulating inherently negative words, like *forget*. There is a general consensus from this work that extra processing is required to accommodate negations. More recently, negation has been suggested as one discourse factor that can induce suppression of activation of a concept in the reader's representation of the text (Gernsbacher & Jescheniak, 1989). Specifically, negated propositions become less active and therefore they are more difficult to verify as having appeared in the text. MacDonald & Just (1989) summarize the effects of negation in the following three points: (a), negation does not affect initial encoding (reading times), (b) negation does produce discourse level effects on the negated concepts, (c) as established in earlier literature, negation does produce large effects on high-level processes that compute truth value (statement verification times). MacDonald and Just propose that on the discourse level negation shifts discourse focus from the negated item to one that has not been negated.

Lea and Mulligan (2002) investigated the effect of negation on the production of deductive inference during reading, suspecting that readers would be less likely to make inferences about what did not or will not happen than they are to make inferences about what has or will happen. Lea and Mulligan compared two of the ML schemas, an *or*-

elimination and *not-both elimination*. Or elimination inference produces an affirmative conclusion (*a* or *b*, not *a*; therefore *b*) and not-both elimination produces a negated conclusion (not both *a* and *b*, *a*; therefore not *b*). Their results showed that naming times for negated concepts were slowed, but the reading time provided strong evidence that participants do draw the not-both inference; they were much slower to read a sentence when it contradicted a not-both inference than when the sentence appeared in a control version of the story that did not sanction the inference.

Lea and Mulligan (2002) conclude that: first, negation does not seem to inhibit the inference making process. Second, it appears that negation does affect the inference after it is made – inference concepts that are negated are less accessible than comparable inferences that are not negated.

Reader Characteristics that Guide Inference Processing

Individuals vary in their cognitive abilities, experiences, knowledge, age, motivation and goals, or available cognitive resources. All these factors affect significantly inference generation during reading.

Reading can have many different objectives or tasks: one can read with the intention of learning, solving a problem, enjoying, or making a summary of a scientific text. Determining the time and cognitive resources the reader is going to invest in comprehension the goal of the reader plays a crucial role in specification of the type of inferences to be drawn.

Kintsch (1998) argues that time and cognitive resources the reader has strongly determine the amount of inferencing that occurs. Experimental studies show that under restricted time limits readers make only few inferences, whereas when more time is given, they make strategic inferences, as causal antecedent inference, or superordinate goal inference.

The amount of cognitive resources determines if a stimulus can be processed at deep or shallow levels. Deep processing involves the extraction of meaning, whereas shallow processing refers to the examination of the superficial features of a stimulus. In reading, for example, counting the number of nouns in a text would be on the level of shallow processing, and judging the activity conveyed by the text would exemplify deep processing. It was proposed that deeper processing result in stronger and longer-lasting memory traces.

Singer (1990) presents various studies showing that shallow processing doesn't result in weaker memory traces but rather in different type of memory, for example the memory for

precise wording of the passage. It seems inevitable that if one prevents the reader from attending to the text meaning by giving him the task to count the nouns in the text, the inferences he draws will be of a different type then when the task is, for example to judge the moral of the story. The goals of the reader, associated with a certain task will guide construction of different text representations and result in different patterns of inference processing (Kintsch, 1998).

Drawing an inference means to combine the information in the text with the knowledge of the reader. Numerous experiments confirm the strong influence of the knowledge of the understander on the type and number of inferences he draws. The understander's knowledge guides the construction of macropropositions, schemas and the judgment of relevance of certain information permit to determine whether or not a certain proposition will be included in the macrostructure, which, in turn affects the course of relevant inference processes.

Logical Inferences in Text Comprehension

The well-known large-scale models of text comprehension (Graesser et al., 1994; McKoon & Ratcliff, 1992; Kintsch, 1998) either do not take any account on how logical inferences are made during reading, or they negate their automaticity during text comprehension. For example, Graesser et al. (1994) believe that "…some classes of inferences are normally difficult to generate and therefore off-line. First, there are logic-based inferences that are derived from systems of domain-independent formal reasoning, such as propositional calculus, predicate calculus, and theorem proving…" (p.376).

Logical inferences are elaborative or forward, as they cannot be considered necessary to bridge two sentences, or to maintain local coherence. Several studies approached the question whether predictive inferences are drawn during reading and the opinions are divided. A number of studies have provided evidence that predictive, or forward, inferences are not drawn during reading (Potts, Keenan, & Golding, 1988; Singer & Ferreira, 1983), other studies have provided evidence of just the opposite (Calvo & Castillo, 1996; Keefe & McDaniel, 1993; Klin, Guzmán, & Levine, 1999; Lea, 1995; Murray, Klin, & Myers, 1993). To explain these different findings a number of methodological variables have been identified. For example, Whitney et al (1992) compared the effectiveness of three different tasks for detecting forward inferences and concluded that word stem completition and lexical decision were well suited for determining what information was inferred whereas naming task was not. In contrast, Murray et al.

(1993) and Keefe and McDaniel (1993) concluded that facilitation could be found to a word naming probe representing a forward inference, but only if the probe was presented while the inferred event was highly in focus. Keenan et al. (1990) advice to use naming task for detecting forward inferences as it is immune to post-lexical context checks that can affect lexical-decision latencies.

Less is known about the influence of text variables on forward inferences. Under what conditions forward inferences are drawn on-line? According to both *minimalist* (McKoon & Ratcliff, 1992) and *constructionist* (Graesser et al., 1994) theories, forward inferences are unlikely to be generated on-line. The minimalist theory assumes that predictive inferences should be more probable if they are readily available based on general knowledge, and if few alternative consequences are available.

The constructionist theory depicts a relatively active reader who attempts to construct a meaningful representation of the text that (a) addresses readers' goals, (b) is coherent both locally and globally, and (c) explains why actions, events, and states are mentioned in the text. Graesser et al. aimed to explain the knowledge-based inferences the readers add to the text base propositions. They clarify that information in the text activates knowledge structures in the long-term memory and part of that stored information is added to the representation of the text. According to the authors, most of these knowledge structures in long-term memory contain contextually rich information that is grounded in experience, such as scripts and schemas.

According to Keenan et al. (1990), describing an inference involves specifying both the unit and the level of the inference. The unit of inferencing can vary from "and activated concept, to a set of concepts constituting a proposition, to … a higher order knowledge structure such as schema" (p.382). Inferences can function on the level of an activated concept, to maintenance on working memory, to encoding in long-term memory. Klin, Guzmán, and Levine (1999) suggest that forward inferences consist of more than the momentary activation of a single concept. Their experimental texts introduced an entire episode between the inference and a line contradicting the presumably inferred event. The reading times suggested that the inferred proposition influenced the processing of the contradiction line. Klin et al. (1999) conclude that forward inferences take form of an entire proposition encoded into the text representation.

Braine and O'Brien (1998a), argue that mental logic schemes are routinely used in text and discourse comprehension. The mental logic theory makes predictions contrary to

those mentioned suggested by theories of text and discourse processing. It states that mental logic inferences during text processing are made errorlessly and effortlessly. Moreover, the inferences based on the schemas of mental logic are automatic; therefore they must be drawn on-line.

The work of Lea and his colleagues (Lea, 1995; Lea et al., 1990) brought some clearance into this controversy. Lea et al. (1990) confronted the reader with short texts, which included logical propositions, represented in English language by the words *if*, *and*, *or*, and *not*. The texts contained premises, which fitted in the schemas of the mental logic model. The participants had to indicate whether or not the final sentence of the story made sense in the context of the story (the validity task). In order to be able to make such a judgment, the subjects necessarily had to use 2 to 3 different schemas of the propositional mental logic model of Braine and O´Brien (1998c). Over the 12 texts, 95% of the responses about the sensibility of the final sentence were logically valid. This result indicates that the logical inferences of propositional ML are drawn routinely and errorlessly during text comprehension. Presentation of ML propositions in text seems to trigger the use of propositional logic schemas in the sequence defined by the Direct Reasoning Routine.

In the second part of the experiment, Lea, et al. (1990) tested the facility of drawing mental logic inferences by looking at recognition and recall intrusions. This methodology is often used in comprehension research (e.g., Bower, 1979). In a typical study, subjects read a passage of text and then take a memory test. The common result is that if the inference is relatively easy to be drawn, the readers confound information that was presented in the text with the results of easy inferences, and they mistakenly “remember” having read information that they actually must have inferred (recall intrusions).

When we read a text, we are usually able to recognize sentences from the text quite good. However, recognition memory for sentences is a complex phenomenon. Singer and Kintsch (2001) summarize the factors that influence sentence recognition. First, whether the probe sentence is a verbatim copy of a sentence from the text, or a paraphrase, a plausible inference or merely topically related information, makes a difference (Kintsch, Welsch, Schmalhofer, & Zimny, 1990). Second, the delay between reading the text and recognition is important (Kintsch et al., 1990). Also, important sentences related to the main topic of the text are remembered more accurately than unimportant details (Walker & Yekovich, 1984); and response strategies affect the results (Reder, 1982, in Singer & Kintsch, 2001).

Relying on the recall intrusion effects of the recognition task, Lea, et al. (1990) constructed three sentences for each story: One was a paraphrase of information that was presented explicitly in the text; the second was a logical inference that the model predicts readers would make while reading the story; and the third one was a 'foil' item. The foil item was an inference valid in formal logic but one that the model does not predict readers would make. The participants were shown these three sentences and asked to indicate whether the information contained in the sentences had been presented explicitly in the text they had just read, or whether they had to infer that information. This was called the recognition task.

Lea et al. (1990) predicted that participants would (correctly) recognize the paraphrase as having been presented in the text they just read, that they often would (incorrectly) recognize the model-predicted logical inference as having been presented in the text, and that they would (correctly) reject the foil inference as having been presented in the text. The false alarms triggered by the model-predicted logical inference would occur because readers would make those logical inferences and incorporate the resulting information into their mental representation of the texts so easily that only moments later they would have difficulty determining whether they had inferred that information or read it. Lea et al.'s results supported this prediction: 69% percent of the time participants thought that the model-predicted inferences had been presented explicitly in the texts, whereas only 15% of the time did they think the logically valid non-predicted inferences (foil items) had been presented in the stories. Eighty-nine percent of the time participants accurately identified the paraphrases as containing information presented explicitly in the passages.

Thus, the results from the validity task and the recognition task provide strong evidence that people are able to make the logical inferences described in the model very accurately and easily enough that they often do not realize that they are making inferences. Hence, Lea et al., (1990) provided the first empirical evidence that the model's predictions apply to every-day situations such as text processing.

The next objective of Lea and his colleagues was to detect whether the logical inferences are drawn on-line. Keenan et al. (1990) give some methodological suggestions on how to approach this problem. The typical experiments in this category are based on the assumption that the result of an inference is stored in the working memory. Working memory is engaged in any conscious cognitive processes, storing the products of those processes (Baddeley, 1986; Baddeley & Hitch, 1974); therefore, it is involved in text comprehension (Just & Carpenter, 1992). During text comprehension the results of all the on-line inferences are

readily available in the working memory. Words stored in the fast working memory are more readily available then words retrieved from long-term memory. This fact is explored by the *word recognition* and the *naming task*.

In the naming task the subject has the task to read as quickly as possible a word that appears on the screen. If the word is already primed and active in the working memory, a shorter reaction time is necessary to read the word than for a word from the long-term memory.

Lea (1995) constructed an experiment, where the subjects were reading 24 short stories with one of the propositional logic schemes included in its structure. In one set of experiments Lea (1995) used the propositional version of the *or-elimination* schema (schema no.1 on page 17), and in a second experiment the *modus ponens* schema was used (schema no.2 on page 18). Each story was presented on the screen of the computer one sentences at a time. The stories were five sentences long and had the following structure: first came an introduction sentence, then a sentence containing the first premise of the logical scheme followed by a filler sentence. The fourth sentence introduced the second premise of the schema. Immediately after the introduction of the second premise a word appeared on the screen. The subjects were instructed and trained to read out loud such a word as soon as it appeared on the screen (the naming task). The last sentence presented either a valid or an invalid conclusion of the logical inference. After reading the last sentence the subjects had to answer a comprehension question, as they were led to believe that the purpose of the experiment is to test aspects of text comprehension.

In the experimental versions of the stories, the word probe for the naming task was the result of the logical inference. The control versions of the stories were identical to the experimental ones except that the fourth sentence, located immediately before the naming task, did not contain the second logical premise. The experimental and control stories were inserted between filler stories, which either or didn't contain any logical particles at all, or the naming task probe repeated some other word from the story.

The model states that the logical inference will be made at the moment both premises are simultaneously available. Therefore, naming task targets that follow the experimental inference stories should be identified significantly faster than those that follow the control no-inference stores, and that is exactly what Lea (1995) found; when targets followed inference versions of the passages the average lexical decision latency was 19ms faster than when they followed the control versions. This difference (reliable at $p < .01$) demonstrates that

the inference targets were primed and strongly indicates that the participants were making the logical inference on-line at the moment both premises were available, as the model predicts.

A similar result was obtained by the lexical decision task, where subjects were presented with a string of letters on the screen and had to decide as quickly as possible whether the target represents an English word. The targets were identified more quickly as words when they followed stories that permitted the reader to infer a semantic associate of that word than when they followed an otherwise identical story that did not sanction that inference.

Lea et al. obtained a similar inference effect for the modus ponens schema, where the lexical decision task provided evidence that readers draw forward inferences according to this schema.

The experiments for both or-elimination and modus ponens schema were replicated with stories without titles investigating the effect of thematic focus on forward inference drawing. Also data from the no-title experiments produced significant effects for the forward inferences, although in the absence of titles the difference between the inference and no-inference versions of the stories was somewhat smaller.

These experiments show that the mental propositional logic model is suitable for the description of logical inferences made during text comprehension. The question remains, whether the second part of the model, the mental predicate logic, will prove to have the same utility in everyday reasoning. Up till today, the mental predicate logic schemas have not been subjected to examination of their use in text and discourse comprehension.

Objective of the Study

The study examines the predicate mental logic model's ability to explain logical inferences in text comprehension. The objective of the study is to test whether:

1. Mental logic inferences are used effortlessly and errorlessly during text comprehension,
2. The use of propositional logic schemas follows the Direct Reasoning Routine,
3. Mental logic inferences are drawn automatically, i.e., on-line.

Three experiments were designed in order to reach the above listed three objectives.

CHAPTER 2: EXPERIMENT 1A – SINGLE SCHEMA PROBLEMS - VALIDITY TASK

Method

Participants

Twenty undergraduate university students of psychology and phono-audiology from Universidade Federal de Pernambuco and Universidade Católica participated on a voluntary basis. All subjects were native speakers of Portuguese, between 19 and 24 years old (mean age 20.4, *SD* = 1.4), 5 men and 15 women.

Materials and Experimental Design

Two sets of 8 short stories in Portuguese were constructed. In each of the stories one of the inference schemas of the mental predicate logic model was embedded. The schemas used in this experiment are the 8 core schemas of the predicate mental logic listed in chapter Predicate and Propositional Mental Logic. The schemas are described in detail on pages 17 to 19 and summarized in Table 3. Two parallel sets of stories were constructed, so that each schema was used in two different contexts. Therefore, any two parallel stories had different content but an identical logical form.

Each story was five to ten sentences long. The text was followed by two possible final sentences – Ending 1 and Ending 2, where one was a valid and other was an invalid conclusion of the logical inference. On order to make appropriate sensibility judgments the subjects had to integrate the logical information in the paragraphs in a way that is consistent with the Mental Logic theory, judging the logically valid conclusions as sensible, and the logically not valid sentences as not making sense.

All the experimental texts have been constructed and tested in a pilot study conducted with 12 subjects. About 50% of the stories had to be reformulated in order to eliminate text comprehension errors and wording allowing additional premises invited by “conversational implicatures” (Grice, 1975). The 16 experimental stories and an example of the experimental protocol can be found in Appendix A.

The stories were presented in two random orders and the subjects were randomly assigned to one of the two groups. The order of the presentation of the valid and invalid endings was randomly assigned to the two ending sentences.

Table 3

Description of the Core Mental Predicate Logic Schemas Used in the Experiment

Schema	Notation and example
1a	S1[All X] OR S2[PRO-All X]; NEG S2[α]; α ⊇ [X] /∴ S1[α] The boys either played with girls or fought with girls; Tom and Dick did not play with girls/ Tom and Dick fought with girls.
1b	S1[All X] OR S2[PRO-All X] /∴ S2[All X: NEG S1[PRO]] The boys either played with girls or fought with girls/ The boys who did not play with girls fought with girls.
2a	S[All X]; α ⊇ [X] /∴ S[α] The girls all wore red jeans / The girls in sneakers wore red jeans.
2b	NEG S[~Some X~]; α ⊇ [X] /∴ NEG S[α] None of the boys wore striped shirts / Sam and Henry did not wear striped shirts.
3a	NEG E[~Some X: S1[PRO-All X] & S2[Pro]~]; S2[α]; α ⊇ [X] /∴ NEG S1[α] There were no boys who wore sandals and blue jeans; The boys that plays with Mary wore blue jeans / The boys that plays with Mary did not wear sandals.
3b	NEG (S1[All X] & S2 [PRO-All X]) /∴ NEG S2[All X: S1[PRO]] There were no boys who wore sandals and blue jeans/ The boys that wore blue jeans did not wear sandals.
4	S1[All X] OR S2[PRO-All X]; S3[All X: S1[PRO]]; S3[All X: S2[PRO]] /∴ S3[All X] All the cars in the lot have stickers or the guards tow them away. The cars that have stickers are Toyotas. The cars that the guards tow away are Toyotas / All the cars in the lot are Toyotas.
5	S1[All X] OR S2[PRO-All X]; S3[All X: S1[PRO]]; S4[All X: S2[PRO]] /∴ S3[All X] OR S4[PRO-All X] All the cars in the lot have stickers or the guards tow them away. The cars that have stickers are Datsuns. The cars that the guards tow away are Toyotas / The cars in the lot are all Toyotas or Datsuns.

Procedure

The experiment was presented in group sessions. The subjects were informed that the experiment is investigating text comprehension and memory. Each subject had to work through the set of 16 stories. The experiment was introduced by the instruction: "Read the following stories and indicate the ending which is more appropriate." After each story the two endings were offered, Ending 1 and Ending 2 and the subjects were supposed to mark the more appropriate one (the *validity task*). Each page contained two stories.

Results

In average the subjects made correct judgment in 93% of all the stories. Ten subjects (50% of the sample) did not make any erroneous judgment and 7 (35% of the sample) made one error across the 16 stories. The median error for a subject across the set of 16 stories was 0.5 (M = 1.2, SD = 1.9). Three subjects achieved an unusually high error rate – four, five and seven errors over the 16 stories (see Figure 1).

It has been suspected that the errors of these three participants were of the processing type as described by Braine and O'Brien (1998a), caused by lack of attention or motivation. These three subjects whose error rate values exceeded three medians of the whole sample were eliminated as outliers. After this amendment the total error rate in the reduced sample of 17 subjects dropped to 3%, so the average percentage of correct judgments was 97%. Median error was zero (M = .4, SD = .5). Table 4 describes the proportion and frequency of correct judgments per story and per story pair in the original sample as well as in the reduced sample.

Without making logical inferences the subjects would have no basis to prefer one ending over the other and they would perform on the validity task at chance (p = .50). The range of correct responses was 17 to 20 in the whole sample (N = 20) and 15 to 17 in the reduced sample (N = 17). A binomial distribution analysis indicated that in both the whole and the reduced sample the subjects were performing above chance on all 16 stories. The worst score in the full sample was 17 correct responses out of 20, which is significantly different from the proportion of correct answers given by chance (Z = 3.13; p < .001). In the reduced sample, the worst score was 15 correct answers out of 17, which is also significantly different from answers given by chance (Z = 2.90; p < .002).

Table 4

Proportion and Frequency of Correct Judgments Among All Participants and After Eliminating Outliers

		Correct answers			
		In whole sample (N = 20)		In reduced sample (N = 17)	
Used schema	Story	Frequency	Percentage	Frequency	Percentage
1a	A	19	95%	17	100%
	B	18	90%	16	94%
	A and B		93%		97%
1b	A	19	95%	17	100%
	B	20	100%	17	100%
	A and B		98%		100%
2a	A	19	95%	17	100%
	B	18	90%	16	94%
	A and B		93%		97%
2b	A	18	90%	16	94%
	B	18	90%	17	100%
	A and B		90%		97%
3a	A	20	100%	17	100%
	B	18	90%	16	94%
	A and B		95%		97%
3b	A	18	90%	17	100%
	B	18	90%	17	100%
	A and B		90%		100%
4	A	20	100%	17	100%
	B	17	85%	16	94%
	A and B		93%		97%
5	A	20	100%	17	100%
	B	17	85%	15	89%
	A and B		93%		93%
Across all stories			93%		97%

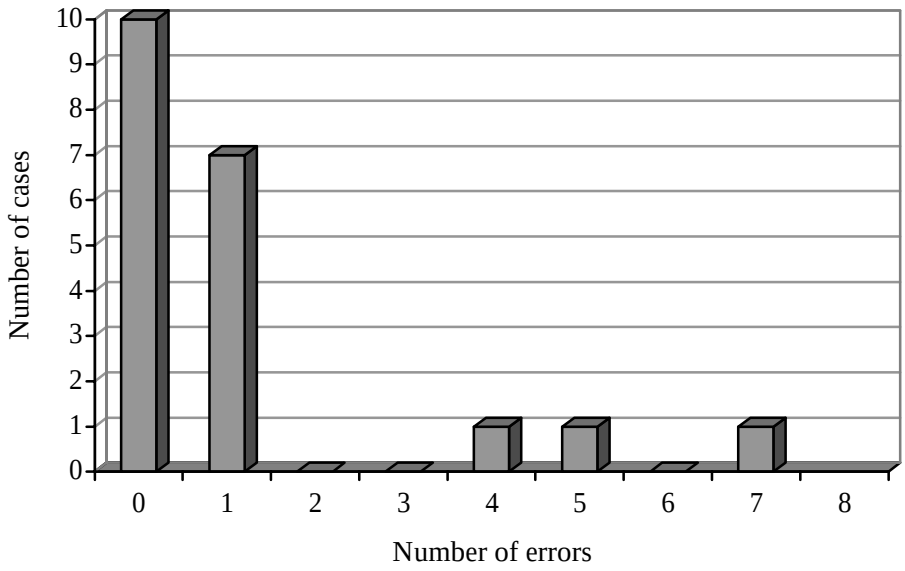

Figure 1: Frequency of errors in the validity task (N = 20)

Discussion

The validity task was designed to measure the level of errors in logical inferences in text comprehension. The results demonstrate that the subjects were integrating logical information in the stories correctly and drawing inferences predicted by the predicate ML theory achieving a very high percentage of correct responses (93 to 97%). Short texts that contain premises for a single mental propositional logic schema seem to be very easy to comprehend.

Each schema was used in two different contexts in order to diminish the influence of content of the text on the logical inference. Contents resembling permission or obligation schemas and other social contract schemas were avoided so the inferences could not be explained by content-bound theories, like those suggested by Chang and Holyoak (1985), or Cosmides (1989). Nevertheless, in our design it was not possible to eliminate completely the influence of the content and wording of the story on the probability of drawing the logical inference. Should we want to do so, the possible strategy would be to present the task in an abstract version (like, for example Lea et al., 1990, in the third experiment of his

study), or use the same schema in more than two different contexts. Data from such a set would permit, for example, to compare the relative difficulty of the schemas.

The design of this experiment was inspired by a similar study conducted by Lea et al. (1990). Lea et al. embedded mental propositional logic schemas in short stories and asked the participants to judge the last sentence on whether it made sense in the context of the story or not. In order to make a correct judgment about the last sentence the subjects had to integrate correctly the logical information in the story. Alternating various instructional conditions Lea et al. reported a rate of correct responses from 87 to 94%, which resembles the result of this study.

There are two differences between the experiment of Lea et al. (1990) and this study: First, the logical schemas used in Lea et al. were propositional, whereas our objective was to test the predicate logic, introducing quantifiers and their scope in the texts. Second, as the use of predicate logic in text comprehension has not been tested yet, the texts of this study contained only one of the mental predicate logic schemas as opposed to the experiment of Lea et al. (1990), where several schemas have to be applied in order to come to the conclusion of the story. A design using stories with several schemas of the mental predicate logic was used in Experiment 2 of this study.

The objective of this experiment was to make a preliminary test of the errorless use of mental propositional logic schemas suggested by Braine and O'Brien (1998c). The participants performed above chance on all stories and achieved a relatively high percentage of correct responses confirming the prediction of Braine and O'Brien for all the core propositional mental logic schemas.

CHAPTER 3: EXPERIMENT 1B – SINGLE SCHEMA PROBLEMS – RECOGNITION TASK

Method

Participants

Twenty-nine undergraduate university students of psychology from Universidade Federal de Pernambuco participated in the experiment on a voluntary basis. All subjects were native speakers of Portuguese, between 19 and 38 years old, median age 21 (mean 22.1, *SD* = 3.4), three men and 26 women.

Materials and Experimental Design

Two sets of 8 short stories identical to those that were used in Experiment 1a – Validity Task were used also for the Experiment 1b – Recognition Task: Short texts five to ten sentences long, in which premises related to one of the 8 core schemas of the mental predicate logic model were embedded (see Table 3), in Portuguese language. The stories were presented without the conclusion of the mental logic schema inference (which in Experiment 1a was presented in the last sentence).

The task of the participants was to read the story and, after turning the page, consider three sentences related to the story. One sentence was a paraphrase of information presented in the story (the *paraphrase item*), another sentence represented an output of the mental logic schema (the *model-predicted item*) and the third sentence was a conclusion that is valid in standard logic but does not make part of the ML theory schemas (the *foil item)*. This task, referred to as the *recognition task* required subjects to judge each of the three sentences whether the information appeared in the story or had to be inferred from information presented in the story. An example of a story with the three recognition items is presented in Table 5.

The participants had three possibilities of answering whether the information in the recognition item was presented in the story: “Was presented in the story (although not word-for-word)”, “I am not sure”, or “Was not presented in the story (but might have been figured out)”.

The 3-point scale was developed during a pilot study conducted with 15 participants. The option "I am not sure" was included to avoid guessing behavior.

Table 5

Two Sample Stories from Experiment 1b and their Recognition Items

The Luncheon (from story pair 1, using schema 1a)

The boss invited some of the people who work for him to a BBQ. All the guests at the party were from the same department. The boss was famous for always offering delicious food. This time everybody could choose either rib steak or grilled fish.

Márcia, the secretary, also came to the BBQ. When the boss saw that Márcia and her husband already had finished eating, he was curious about what they had chosen.

"I had rib steak yesterday for dinner," said Márcia, "so I didn't have it again today."

Model-Predicted item: Marcia had grilled fish on the BBQ offered by the boss.

Paraphrase item: Marcia told the boss that she didn't eat rib steak.

Foil item: A person from another department didn't come to the BBQ.

School (from Story Pair 5, using schema 3a)

Some women from Casa Forte were speaking about their children when they were having lunch. Angela mentioned that she wanted to her oldest son go to a school that would give the curriculum both in English so that he would learn to speak English properly and she complained:

"But there is no school in Recife that is cheap enough for us to afford and gives the curriculum in French. The American school of Recife in Boa Viagem is one of the schools I am speaking about, because it, in fact, does give the curriculum in English.

Model-Predicted item: The American school of Recife isn't cheap enough for Angela to afford.

Paraphrase item: Angela wanted her son to learn to speak English properly.

Foil item: The women were from Casa Forte or Boa Viagem.

Pilot study was also necessary to find a wording of the text and of the three recognition items that would avoid invited inferences.

None of the three sentences appeared word by word in the story. The participants were required to make a rather subtle judgment about whether the sentence was a paraphrase or it had to be inferred from information in the story. It was thus necessary to control for other aspects of the recognition items that could influence the recall of the information from the text: The use of negatives, average length of the sentences, distance of the premises in the text, and general linguistic similarities.

The mean number of words in the three recognition items was 12.4, 12.0, and 11.6 for the model predicted, paraphrase, and foil item, respectively. From the 16 stories, negatives were used in 9 of the model-predicted items, 9 of the paraphrase items, and 8 of the foil items.

The general linguistic similarity was counted as the proportion of words in the recognition item that are the same and presented in the same order as they appear in the most similar sentence in the text of the story. The mean proportions of corresponding words were .30 (*SD* = .08), .31 (*SD* = .14), and .29 (*SD* = .11) for the model predicted, paraphrase, and foil item, respectively.

The information used in the three recognition items had to be integrated from in different locations of the story, sometimes from the beginning, sometimes from the middle and sometimes from the end of the story. This, too, can influence the recall of the information; therefore, the relative position of relevant information in the story for paraphrase and foil item was counterbalanced across the stories. Paraphrase items referred to information from the beginning of the story in stories no. 4, 9, 11, 12, and 16, and to information from the end of the story in stories 1, 3, 6, 7, and 8. In stories 2, 5, 10, 13, 14, and 15 the relevant information was in the middle of the story. Foil information integrated information from the beginning of the story in stories 1, 3, 8, 9, 10, and 14, and from the middle in stories 5, 7, 13, 15, and 16, from the end of the story in stories 2, 4, 6, 11, and 12.

In case of logical inferences the distance between the premises in the text could influence the relative difficulty of drawing the inference. For each story pair, the premises in one story were introduced next to each other and at the very end of the story. In the other story of the same pair, the premises were separated by one or two sentences from each other and from the end of the story.

The stories were presented in two random orders and the subjects were randomly assigned to one of the two groups. The order of the presentation of the three recognition items was also randomized.

Procedure

The experiment was run in one group session of 29 subjects. The participants were informed that the experiment is investigating text comprehension and memory. Each subject had to work on one practice story and 16 experimental stories. The stories were presented one on each page. The experiment was introduced by the instruction to read the story. After finishing reading the subjects had to turn the page and respond three questions about the story. Turning the page was necessary so that the text of the story was no longer available for checking. Participants were reminded on both pages that once they passed to judging the three sentences they were not allowed to turn the page back to check the text of the story.

The next page contained the three recognition items and the subjects were asked to indicate whether the information had been presented in the stories or they had to figure it out. The instructions explained that all three phrases contain either information that was presented in the story (although not word-for-word) or was not presented in the story but might be figured out from the story. For each of the three sentences, the subjects had to mark one of the three responses: “Was presented in the story (although not word-for-word)”, “Was not presented in the story (but could be figured out), or “I am not sure”. The participants were also asked not to give the same answer to all three items (like, for example, three times “Was presented in the story (although not word-for-word)”). The experimental material and an example of the experimental protocol can be found in Appendix A.

Results

The subjects were randomly assigned to two groups that differed in the order of presentation of the stories and of the recognition items. An initial analysis was computed in order to assess possible differences between the two orders of presentation.

A Whitney-Mann U test showed that across all the 16 stories except for one item[1] there was no difference in judgments of the recognition items between the two orders, so the two data sets were combined.

A comparison between the foil and paraphrase items drawn from the beginning of the story with the ones related to the information presented in the end of the story did not show any significant difference. There was also no significant difference between the stories where the logical premises were one sentence distant from each other and from the end of the story, and stories where logical premises were following each other at the very end of the story.

The task of the subjects was to provide judgments whether the recognition item information was presented in the story or not (- it had to be inferred from the information in the story). The responses of the subjects were coded as follows: An answer "Was presented in the story (although not word-for-word)" was coded as 1, "I am not sure" was scored as 0.5, and "Was not presented in the story (but could be figured out)" was assigned 0. The response "Was presented in the story (although not word-for-word)" will be referred to as an "Yes" answer, and the answer "Was not presented in the story (but could be figured out)" will be referred to as an "No" answer.

There are two possible interpretations of such a scale: Should one assume that the assertion "I am not sure" is expressing half of the probability between "Yes" and "No", then the ratings 0, 0.5, and 1 would express the probability that the information was presented in the story. The mean score on such a three-point scale would indicate the average probability of "Yes" responses that the subjects assigned to the item. A second possibility is to assume that a response "I am not sure" means that the subject considered a certain part of the information in the sentence originating from the story and another part not, having to be figured out. In this case, the sentence as a whole should properly be judged as not being presented in the story. Therefore, the judgments "I am not sure" were recoded from .5 to 0 (zero). This re-codification dichotomized the answers of the subjects into zeros and ones reducing the 3-point scale to a 2-point one.

Table 6 illustrates the mean answers for the three recognition items across all 16 stories. On the original 3-point scale this proportion varies from .09 to .97, reaching the mean answers of .27 for the foil item, .65 for the model predicted item, and .81 for the paraphrase item.

[1] Foil Item for the story nr 10 – "School" $Z = 2.73$ ($p < .005$)

Table 6

Mean Responses for the Recognition Items Across all 16 Stories on a 2-Point and 3-Point Response Scale.

		Mean Responses					
		Foil		Model predicted		Paraphrase	
	Story Title	3.-point scale	2.-point scale	3.-point scale	2.-point scale	3.-point scale	2.-point scale
1	Lunch	.22	.17	.50	.48	.91	.90
2	Legs	.26	.00	.29	.28	.97	.93
3	Conference	.31	.21	.64	.62	.78	.76
4	Concert	.38	.38	.62	.55	.57	.48
5	Friends	.29	.10	.53	.48	.78	.72
6	Sandwich	.34	.21	.81	.79	.72	.69
7	Fruit	.34	.10	.64	.62	.84	.79
8	Restaurant	.31	.18	.45	.46	.83	.86
9	Guide	.26	.14	.66	.66	.79	.76
10	School	.26	.00	.57	.55	.98	.97
11	Stealing	.24	.14	.91	.90	.84	.79
12	Mini	.09	.00	.81	.79	.84	.83
13	Exam	.29	.10	.74	.76	.78	.66
14	Chairs	.17	.24	.76	.72	.71	.72
15	Exhibition	.19	.21	.71	.66	.88	.69
16	Dance	.28	.03	.72	.69	.74	.86
	Mean	.27	.14	.65	.63	.81	.78
	SD	.07	.10	.16	.16	.10	.12

Note: The range of responses is between .00 and 1.00.

The re-codification of the data to a 2-point scale lowered the scores to values ranging from .00 to .93. The dichotomization affected mostly the foil item, where the mean responses dropped from .27 to .14. The difference of .13 between the score before and after dichotomizing was not statistically significant. The re-codification had practically no impact on the score for the model predicted item, which achieved a mean response of .63 (difference between the 3-point and 2-point scale was .02). The score for the paraphrase item also did not suffer any significant change, the mean dropped from .81 to .77 (difference of .04).

These results confirm that between the recognition items the response "I am not sure" was most frequent for the foils. Several foil items contained a disjunction, where one of the predicates repeated information from the story and the other predicate represented falsification of some information from the story. (Such a disjunction is a valid inference in formal logic.) A comparison of number of "I am not sure" answers for foil items containing such a disjunction with the rest of the foil items revealed that the foils containing the disjunction obtained significantly more "I am not sure" answers than the rest (t (14) = 4.07; $p < .001$). These results suggest that, at least for the foil items, most of the answers "I am not sure" did not mean that the subject did not remember whether the information was presented in the story or not, but instead that he/she understood that half of the information comes from the story and half not. Such an item should in fact be marked as not being presented in the story and coded as 0. It was decided to continue analysis with the recoded dichotomous scale.

The results summarized in Table 7 show that on the new 2-point scale the mean responses over all stories and all subjects for the foil, model predicted and paraphrase item were .14, .63. and .78, respectively. The standard deviations were .10 for the foil item, .16 for the model predicted item, and .12 for the paraphrase item. The foil recognition item in the story *Concert* reached the highest mean answers (.38), contrasted with foils in stories *Legs, School*, and *Dance*, that were left after re-codification with score .00. Between the model predicted items the item of the story *Legs* reached the lowest mean answers (.28). This value is more that two standard deviations distant from the mean value for all model predicted items (.63). Model predicted item of the story *Stealing* reached the highest mean answers (.90), which is higher that the paraphrase recognition item of the same story (.79). The paraphrase item of the story *Concert* reached the lowest mean (.48) as compared with paraphrase items of stories *Legs*, or *Guide*, which reached the highest proportions of the set (.93, and .97, respectively).

Possible Differences Between Parallel Stories of Each Story Pair. For each two stories containing the same logical schema the score for corresponding recognition items were

compared. A binomial distribution analysis showed that between the two sets of 24 recognition items (two sets of eight stories with three items each) there were no significant differences beyond what could be expected by chance alone[2]. Therefore, the data from the two stories containing the same logical schema were combined.

Table 7

Mean Responses for the Recognition Task on the 3-Point and the 2-Point Rating Scale per Story Pair

Schema Used in Story	Mean Responses per Story Pair (N = 29)					
	Foil		Model Predicted		Paraphrase	
	3-p.scale	2-p.scale	3-p.scale	2-p.scale	3-p.scale	2-p.scale
1a	.24	.09	.40	.38	.94	.91
1b	.34	.29	.63	.59	.67	.62
2a	.32	.16	.67	.64	.75	.71
2b	.33	.14	.54	.53	.84	.81
3a	.26	.07	.61	.60	.89	.86
3b	.16	.07	.86	.84	.84	.81
4	.23	.17	.75	.74	.74	.69
5	.23	.12	.72	.67	.81	.78
Total	.27	.14	.65	.63	.81	.77

Note: The range of responses is between .00 and 1.00.

The Mean Responses per Each Story Pair. Table 7 illustrates the mean answers per story pair on the three recognition items using both the three-point as well as the two-point scale. On the 3-point scale the mean responses, indicating the subjects' judgment of probability that

[2] Two story pairs obtained significantly different judgments on a certain item: story pair using schema 2a on the Model Predicted Items of ($p < .05$) and the story pair using schema 3b on the Paraphrase Item ($p < .05$).

the information was presented in the story, varied between .16 (foil item in story pair using schema 3b) and .94 (paraphrase item in the story pair using schema 1a). The foil item reached maximum probability of .34 for the story pair using schema 1b, the model predicted item obtained scores between .40 for the story pair 1a and .86 for the story pair 3b. The paraphrase items reached a minimum mean response of .67 (story pair 1b) and maximum of .94 (story pair 1a). The mean responses for the three recognition items were .27, .65, and .81 for the foil, model predicted and paraphrase item, respectively. Figure 2 illustrates the results from Table 7 showing the mean answers on the three recognition items across the eight story pairs. The graph shows that for all the story pairs, the foil item ratings were lower than the other two recognition items. The model predicted items scored in most of the stories relatively closer to the paraphrase item than to the foil item. The scores for the individual recognition items varied across the eight story pairs. Model predicted items reached relatively lower mean answers in story pairs containing schema 1a and 2b, and high scores in story pairs 3b, 4, and 5.

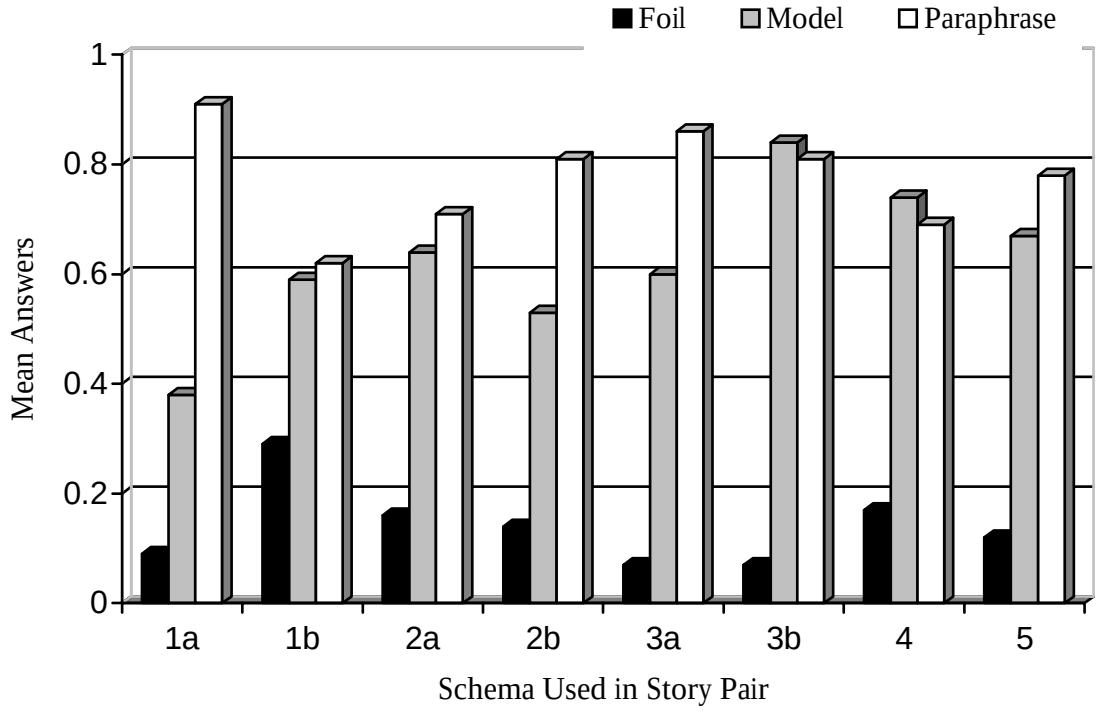

Figure 2: Comparison of the mean answers on the three recognition items across story pairs.

Difference between Recognition Items Across All Stories. The mean answers on the three recognition items across all the subjects and all the story pairs were compared in a Friedman's ANOVA that indicated a significant difference ($\chi^2(2) = 42.11$; $p < .000$). A Wilcoxon's matched pairs test showed that predicted items received significantly higher responses then foil items ($Z = 4.63$; $p < .000$), and paraphrase items received significantly higher rating than model predicted items ($Z = 2.73$; $p < .01$).

The differences within the recognition items across the eight story pairs, analyzed by Friedman's ANOVA, were significant for all three items: foils ($\chi^2(7) = 21.34$; $p < .01$), model predicted items ($\chi^2(7) = 35.05$; $p < .000$), and for paraphrase items ($\chi^2(7) = 26.19$; $p < .000$). As the mental logic inferences were tested in the model predicted recognition sentence, the post-hoc analysis was carried out only for this item. The proportions test showed significant differences between stories containing schema 1a and 3b ($Z = 2.47$; $p < .007$), and between stories containing schemas 1a and 4 ($Z = 1.95$; $p < .03$). The difference between stories 2b and 3b was also relatively big, but did not reach significance ($Z = 1.48$; $p < .07$). Schemas 3b and 4 reached relatively high scores being rated mostly as paraphrases of the information in the story. The scores on schemas 1a and 2b were relatively low, rated by the subjects as requiring an inference.

Table 8 shows the results of the proportions tests indicating that the mean answers for the model predicted items were more close to the scores for the paraphrase items, compared with the scores for the foil items. In five of the eight story pairs the mean answers for model predicted items were not significantly different from the scores for the paraphrase item. The difference in rating of model predicted and paraphrase item was significant for story pairs containing schemas 1a, 2b, and 3a. On the other hand, foil items received significantly lower scores than the model predicted items in all eight story pairs.

Table 8

Comparison of Mean Answers Between the Three Recognition Items.

Schema Used in Story Pair	Mean Answers			
	For Foil less than for Model Predicted?		For Model Predicted less than for Paraphrase?	
	Z	*p* <	*Z*	*p* <
Schema 1a	3.53	.000	3.86	.000
Schema 1b	2.56	.05	.15	*ns*
Schema 2a	4.12	.000	.33	*ns*
Schema 2b	3.45	.001	2.38	.05
Schema 3a	3.96	.000	2.50	.05
Schema 3b	4.69	.000	.13	*ns*
Schema 4	4.31	.000	.23	*ns*
Schema 5	4.24	.000	.49	*ns*
Across all Stories	4.63	.000	2.71	.01

Discussion

The analysis of the recognition task data revealed that in 63% of the time items based on the mental predicate mental model were judged as not requiring any inference. This score is similar to the subjects' judgments of the paraphrase items (77%), and quite distant from the evaluation of the valid inferences of formal logic presented in the foil item (14%). In most of the stories the judgment of the mental logic inference was not significantly different from the judgment of the paraphrase of the story. On the other hand, in all story pairs the participants judged the mental logic item as being part of the story significantly more often than the formal logic item. These results give support to the prediction of the mental predicate logic model that during text comprehension the inferences of the mental logic are drawn with such ease that the

subjects are not aware of drawing any inference at all.

The objective of the experiment was to access the relative difficulty of drawing logical inferences based on the mental predicate logic model. The results suggest that logical inferences that do not make part of the mental logic model require more cognitive resources and attention of the subjects than the schemas of the mental logic. In this way, during recall of the story, the readers are aware of having drawn an inference and judge the formal logic inferences as not being presented in the story. Mental logic inferences do not require such a conscious effort and therefore are confounded with paraphrases of the text.

The results of this experiment strongly resemble those achieved by the experiment of Lea et al. (1990). Lea et al. applied the validity task in a similar experiment in order to access the relative difficulty of drawing mental propositional logic inferences. The text of the stories contained premises for two to five schemas of the model. Mental logic inferences were judged as being presented in the story in 69% of the cases, compared with paraphrase items in 89%, and foil items in 17% of the cases. Lea et al. concluded that the subjects had little or no difficulty correctly executing inferences predicted by the mental propositional model while reading the text.

Another similar study was run by Zimny (1989), in Kintsch et al. (1990). Zimny studied general aspects of sentence memory. She let the readers read texts 150 to 200 words long and compared the memory for verbatim copies of the sentences from the text, paraphrases, inferences, and entirely new sentences. Paraphrases involving minimal word and order changes were judged as being present in the text in 70 to 75% of the cases. Inferences "…that could be inferred by the readers from the surrounding context with high reliability" (p.138) were judged as being presented in the story in approximately 20% of the cases. Our study offered two types of inferences as recognition sentences: Those proceeding from mental logic reached 63% of recognition score, which is close to the score of paraphrases in Zimny's study. The score on formal logic inferences, mentioned in the foil items of the present study, was 14%, resembling Zimny's scores for inferences.

Analyzing the recognition task, it has to be noted that there are factors other that relative difficulty of the inference that could influence the recall whether certain information was presented in the story or had to be figured out. As none of the recognition items had the same wording as the text of the story, the readers had to make a subtle judgment on whether the information was presented it the story or not, trying to remember the exact wording of the story. For example, should the information used in the recognition item come from

the beginning of the story, the reader might have difficulty in remembering the exact wording and he/she might have more trouble to distinguish between an inference and a paraphrase. The longer the delay between reading the text and recognition, the more inferences are accepted as paraphrases of the text. (Kintsch et al., 1990). For recognition items dealing with information from the very end of the story it would be easier to remember whether an inference had to be drawn or not, simply because of the recency of the information.

In the present experiment the location of the information relevant for the three recognition items in the story was balanced, situated equally in the beginning, middle and end of the story, so this factor should not have influenced the results of the validity task in this experiment.

A potential explanation of the results of the experiment could also lay in possible superficial linguistic similarities between the recognition items and the sentences of the text of the story. The judgment whether the information offered in a recognition item was presented in the story or not could have been influenced by surface structure and semantic similarity instead of relative difficulty of drawing an inference (Kintsch et al., 1990). Some recognition items could have been judged as being presented in the story simply because they resembled a certain sentence in the story more closely than another recognition item. Should the model predicted items systematically contain more words in common with a certain sentence from the text that did the foil items, they might lead the subjects to prefer model predicted items to foil items. The percentage of words of the recognition item that is the same and in the same order as presented in the most similar sentence of the text linguistic similarities was counted for all the recognition items. The mean percentages of linguistic similarity were almost equal for all three recognition items. Also, there was no difference between the mean length (number of words) of the recognition items. Thus, simple linguistic similarities could also not account for the results obtained in the experiment.

Another factor influencing the judgment on the recognition task has to do with the text comprehension mechanisms. Singer (1990) suggests that inference processing is concentrated on the ideas related with the thematic ideas in a discourse. Important sentences related to the main topic of the text are remembered more accurately than unimportant details (Walker & Yekovich, 1984). According to the sentence recognition theory of Singer and Kintsch (2001), recognition sentences that are connected to the strongly activated nodes of the text representation receive more activation and are recognized more easily that those related to weaker nodes. Strong nodes are obviously related to the thematic ideas or the gist of the text

(Kintsch, 1998).

In order to provide equal conditions for recall of the recognition items it would be therefore desirable to construct the three recognition items in such way that all three would refer to information from the theme of the story. This condition could not always be satisfied because it was difficult to incorporate premises of both the mental as well as the formal logic inferences as being related to the theme of the story, when the stories were only 5 to 10 sentences long.

The inferences of formal logic presented in the foil items were mostly judged as not being presented in the story. This means, that the subjects are conscious of the cognitive effort necessary to draw such an inference. One of the explanations of why some of the formal logic schemas do not make part of our everyday reasoning rules can be found in the theory of Grice (1975, 1979). Grice's theory gives explanation on the logic of our everyday communication. His basic assumption, the Cooperative Principle, says that every speaker intends to meet the conversational demand of the moment. For example, the contribution should be as informative as required; the speaker should not provide more information than necessary, nor should information be missing. It should also be true – we do not say what we believe is false, or for what we lack evidence. The utterance should be relevant and perspicuous, avoiding obscurity of expression, ambiguity; it should be brief and orderly.

Formal logic studies the most general features of the reality and the set of it's schemas have to be complete not leaving any space for doubt. Formal logic "…is concerned with absolutely everything, whether it is merely possible or actually exists" (Kearns, 1988, p.1). By mentioning or considering everything, or assuming that the reasoner will consider everything, even that that is not mentioned in the premises, formal logic flouts the conversational maxims defined by Grice. Everyday reasoning and the conversational logic proceed from what actually exists or we believe that exists.

Consider the formal logic inferences offered in the foil items. Several of them had the form of logical disjunction between two or three propositions, where one of the propositions was a negation of some information from the story. As the other part of the disjunction was true the whole argument was logically valid. Nevertheless, most participants of the experiment either judged this recognition item as not being presented in the story or expressed their confusion by marking the answer "I am not sure". For example, in the text of the story *School* it was stated that some women having lunch were from Casa Forte. The foil recognition item suggested that the women were from Casa Forte or from Boa Viagem. Most of the

subjects indicated that this information was not presented in the story or that they are not sure about the answer. Indeed, in everyday communication, when we believe that somebody is from Casa Forte, and we are quite sure that he/she is not from Boa Viagem, there is no reason to say that he/she is from Casa Forte *or* Boa Viagem. We do not mention information we believe is false in our discourse.

Gricean rules also help to determine the occurrence of the inferences added to the premises through some pragmatic principles. This can be illustrated by the example of story *Guide*. The story is about a hotel owner who is looking for a guide who speaks both Japanese and Korean. At the end of the story the owner of the hotel complains that nobody from his staff speaks both of these languages, which constitutes the first premise of the *not-both* logical schema (schema no.3 on page 17) . In the pilot version of the study this statement was followed by the owner's utterance *"Isabel speaks well Japanese."*, which represents the second premise of the logical schema. The presence of this sentence according to the Conversational Maximums means that it is relevant to the topic, and indeed, the readers did infer that Isabel is one of the members of his staff, which is an information not presented in the text. Nevertheless, they often did not draw the inference that Isabel does not speak Korean, because they might have expected that should this be the case, the information would already be present in the text. The owner of the hotel was expected to make his contribution as informative as required for the purpose of the conversation. The purpose of saying *"Isabel speaks well Japanese"* might have been confusing – is it a continuation of the complaint that nobody speaks both languages, or is it a beginning of an idea on how to solve the problem? Enriching the last sentence to the form *"It's a pity because Isabel speaks well Japanese"* increased the occurrence of inferences that Isabel does not speak Korean. The connective expression *"It's a pity because…"* explained to the readers the purpose of this sentence: It was a continuation of the complain of the hotel owner, that there were no members of his staff that speak both the languages. The trustworthiness of the first premise of the logical schema (- there is nobody in the staff that would speak both languages) was strengthened, which lead to a higher frequency of drawing the inference.

The occurrence of the inference that Isabel does not speak Korean increased even more, when another sentence was put in the mouth of the hotel owner: *"I think we will have to look for an interpreter somewhere else."* This sentence increases the trustworthiness of the premises even more, and supports the inference that Isabel is not a person that would speak both Japanese and Korean.

The example of the story *Guide* and the effects of relevance of the utterances could be analyzed also within another perspective of the logic of conversation. Cooren and Sanders (2002) propose a general conversational schema (or structure of exchange) in which the Gricean principles would be embedded. This schema consists of five phases: Disorder (what is wrong), Manipulation (having or wanting to do something about it), Competence (means to accomplish the objective), Performance (doing or delegating), and Sanction (the result of Performance). Consider, for example, that the story *Guide* would present the second premise by the hotel owner's utterance *"Isabel speaks well Japanese"*. According to Cooren and Sanders, the text provides the description of the Disorder, and the cited utterance should explain the phases of Manipulation, and Competence: The reader would have to infer that the owner wants to do something about the problem, and that Isabel is considered a means to solve her problem. Nevertheless, it is not clear whether Isabel has the competence to solve his problem. Only adding *"It's a pity, because Isabel speaks well Japanese*" clarifies that she has not. The utterance *"I think we will have to look for an interpreter somewhere else"* defines the Performance phase of Cooren and Sanders' model, and induces the reader to draw the inference that Isabel does not speak Korean, as predicted by schema 2b.

Grice's transgression of quantity, quality, relation, and manner could be seen as negligence of any of the phases of conversational exchange proposed by Cooren and Sanders. Both Grice's and Cooren and Sanders' theories can predict the probability of drawing logical inferences in the texts of the present experiment.

In most of the stories the foil item scored lower than the other two recognition items and the model predicted item's mean was lower or similar to the score of the paraphrase item. Only few recognition items did not follow this general tendency. For example, the model predicted item of the story *Legs* reached a score substantially lower (.28) than the mean of all model predicted items across all the stories (.63). What could explain such a result? The mental logic schema used in the story *Legs* was the same as the schema used in the story *Lunch*, where the model predicted item reached a score of .48, therefore higher than in *Legs*, although still below the mean. One possible explanation of the low score of the model predicted item in *Legs* could be that the schema 1a is relatively more difficult than the other seven predicate mental logic schemas tested in the experiment. This schema uses the particle *or*, which is, according to Yang et al. (1998) a schema of medium difficulty. Should this be true, the stories of schema 1b should have similarly low scores because schema 1b also contains the single logical particle *or*. Schema 1b appears in the stories *Conference* and *Concert* and the model predicted item in these

stores obtained mean scores of .62 and .55 respectively, which are close to the mean score for all the model predicted items. Therefore, the difficulty of the schema alone could not account for the extremely low score on the model predicted item of the story *Legs*.

A possible factor that could contribute to a lower score in the model predicted item of the story *Legs* had to do with the location of the logical premises in the story: the two premises were both mentioned in the last sentence of the story. Therefore, the exact wording of the last sentence might still be available in the memory of the subjects when they were judging the model predicted item. The recall of the exact wording could eliminate recall intrusions, which are the basis of the recognition task in this experiment. This fact might have contributed to the judgments of this item as requiring an inference. Nevertheless, some other stories that achieved a higher score for the model predicted item had also a similar location of the logical premises.

Another relevant factor could be the issue of trustworthiness of the premises. Should the readers had not assumed the premises true, they would hesitate to draw a mental logic inference and mark the recognition item as needed to be figured out. The story *Legs* is about a girl named Roberta who is going to go dancing with her boyfriend and she is choosing what to wear. Roberta claims that she always wears pants or long skirts and this time she will not wear long skirt, because it would annoy her during dancing. In the model predicted recognition item the readers had to judge whether the inference that Roberta decided to wear pants was presented in the story or not. The trustworthiness of the premises could be weakened by the fact that such a conclusion points to a possible action in the future (intention to wear pants), as compared to most of the other stories where the model predicted items were about events in the past or present.

The believability of the premises could be also undermined by some pragmatic principles adding invited premises to the logical premises explicated in the text (Braine & O'Brien, 1991). For example, the reader could have the practical experience that girls often change their mind in relation to what they will wear or not wear. Such an assumption would add a premise like, for example: "*Roberta might wear pants or skirt or basically anything*", that would discourage the subjects to draw the predicted mental logic inference. As Evans (2002) confirms – "…people respond to and express degrees of belief rather than making absolute deductions about truth and falsity" (p.984).

In sum, all of the mentioned factors together could influence the recognition score of a specific item of a specific story. The case of the story *Legs* shows that it is impossible to

distinguish between all the factors influencing inference drawing during text comprehension.

Another factor that could have influenced the occurrence of logical premises was detected during the construction of the stories and their test in pilot studies. For example, in the story *Exam* one of the actors (Stefano) presents the premises of the logical schema, telling his friends about an exam he had just taken. The premises introduced in Stefano's discourse were: *All the questions of the exam were about memory and perception; all the questions about memory Stefano had answered; all the questions about perception Stefano had answered*. According to schema 4 these premises lead to the conclusion that *"Stefano answered all the questions of the exam."*(see Table 3). Nevertheless, when this sentence was offered as the model predicted item, the participants of the experiment often hesitated to confirm this conclusion. The rate of confirmation increased when the conclusion was formulated as *"Stefano told his friends that he answered all the questions"*. It appeared that the readers made a distinction between what Stefano had said and what "really happened" during the exam. Readers do track "who knows what" over the course of longer texts (Lea, Mason, Albrecht, Birch, & Myers, 1998). Gerrig, Brennan, and Ohaeri (2001) introduce the concepts of *projected knowledge* and *projected co-presence* to describe the above-described situation, in which readers infer that characters possess certain information presented only in narration. Readers use evidence from the text to project their own knowledge to the characters: they infer that the character knows what they know (projected knowledge). Other times reader infer that two or more characters have mutual knowledge of the information, for example, when several characters have the possibility to witness a certain perceptual event (projected co-presence). Gerrig et al. also states that it is frequently the case that readers project co-presence among characters for knowledge they themselves do not know: "Readers are able to make subtle distinctions between, for example, what they know and what the characters (e.g., narrative speakers and addressees) know" (p.94). These distinctions seem to determine the probability of drawing a logical inference. In the present experiment, when premises were introduced in the story as utterances or ideas of some of the characters of the story, the mental logic schema conclusion in the recognition item had to be presented also as an inference drawn by the same characters. When this condition was not met and the conclusion was presented as a general fact, the conclusion was judged as not being presented in the story.

In our experiment, eight of the core schemas of the mental predicate logic were tested. Each schema was used in two different stories in order to reduce the influence of a specific content and context on drawing the inference. The function of the validity task in this experiment was to offer information about the relative cognitive effort of drawing an inference.

Assuming that the recognition task does provide reliable information about the difficulty of drawing the inference from the text, we could make certain assertions concerning the relative difficulty of the individual schemas of mental predicate logic. The response means for the mental logic inferences varied substantially across the eight story pairs. Each story pair contained one schema of the model. A high score achieved for a certain story pair would mean that the mental logic schema used in the story pair is relatively easy, as compared with story pairs where the mental logic recognition item reached a low response mean. Story pairs that reached relatively higher scores are those using schemas 3b and 4 and story pairs that achieved the lowest scores are those using schemas 1a and 2b. Precaution should be taken evaluating the mean score of story pair 1a, as the mental logic recognition item of one of the stories of the pair scored more than two standard deviations lower on the model predicted item than the mean of the all the stories. The low score of this story might be caused by the influence of factors of text comprehension other that relative difficulty of drawing an inference, as has been discussed in relation to the story *Legs*. Without considering the story pair 1a it can be suggested that the mental logic schemas 3b and 4 are relatively easier to use than the schema 2b. Nevertheless, this proposal needs to be evaluated in an experimental design that would allow testing each schema in a bigger number of stories.

Yang et al. (1998) assessed the relative difficulty of individual schemas of mental propositional logic model. Not all the schemas used by these authors overlap with those applied in the present experiment. Yang et al. included feeder and incompatibility schemas in their evaluation and, based on intuition and preliminary work, assigned the same weights to some of the schemas, as, for example, 2a and 2b. The authors conclude that schemas containing one of the logical particles *and*, *if*, *some* and *not* are relatively easy, schemas that use the particle *or* is of medium difficulty, and schemas using more than one logical particle are relatively more difficult. The results of the experiment of Yang et al. are surprisingly contradictory to the suggestions mentioned above: the authors classify schemas 3b and 4 as some of the more difficult ones, while the present experiment suggests that these schemas be relatively easy to apply. Schema 2b seems relatively more difficult in the present study, while according to Yang et al. it should be an easy one.

Cordeiro (2003) tested the relative difficulty of single predicate mental logic schemas in short narratives presented to deaf and hearing children and adolescents with different level of schooling. Cordeiro presented the subjects with four different stories for each schema and applied the validity task. All groups committed less error in stories containing schema 4, and

had more difficulty with stories based on schemas 2b, and 3a, 3b, and 1b. These results agree with the results of the present study in relation to schemas 4 and 2b, and disagree in relation to schema 3b.

What could explain these differences in relative difficulty of the schemas? Graesser et al. (1994) noted that readers draw more often inferences that are highly activated from multiple information sources. This observation is related to the text comprehension process: During reading, readers are connecting propositions in an integrated network (Kintsch, 1998). When a certain proposition is connected to a large number of other propositions, it receives more activation and during integration process it is maintained as part of the macro-proposition of the text. Schema 4 requires three premises to be triggered. For example, the text of the story *Exam* contained premises: *All the questions of the exam were about memory and perception. Stefano answered all the questions about memory. Stefano answered all the questions about perception.* Most of the subjects judged the conclusion that *Stefano answered all the questions of the exam* as having been presented in the text.

The arguments *questions about memory*, or *questions about perception* are repeated in all three premises, and the argument that the actor who answered them was *Stefano* was mentioned in two of the premises. According to the Construction-Integration model of text comprehension of Kitsch (1998) it can be predicted that the concepts *Stefano, questions about memory*, and *questions about perception* would receive more activation than other concepts from the text, mentioned only once. After reading the text the reader is confronted with the recognition sentence *Stefano answered all the questions of the exam*. Singer and Kintsch (2001) propose a model of sentence recognition, which is based on the assumption that items that probe people's memory for text, such as test recognition sentences, themselves constitute text. Therefore, general principles of text comprehension apply to such test items. Propositions from the test sentence are extracted and arranged into a network linked to the network of the text representation in memory. From the activated portion of the memory structure, activation flows into the test sentence. How much activation the sentence receives ultimately contributes to how well it is recognized.

In our example, the part of the network concerning Stefano and the questions of the exam would be highly activated because it was repetitively mentioned in the text. It can be expected that this activation would flow into the test sentence and thus facilitate the recognition of this sentence.

This hypothesis suggests that the higher the number of logical premises presented in the text, the easier the recognition of the test sentence would be. It will be shown that

this relation is not as straightforward. Schemas that contain three premises and could expect such a facilitation effect, are schemas 4, 5, 1a, and 3a. The facilitation of the larger number of premises was suggested only in relation to schema 4. (As mentioned earlier, story pair containing schema 1a behaved somewhat strange and was excluded from this analysis.)

Why wouldn't, for example, schema 5 be facilitated by its three premises? Schema 5 was used in story *Dance*. The text contained the following premises: *All the bands at the bar are from Recife or Caruaru. All the bands from Recife will play forró. All the bands from Caruaru will play xote*. The recognition item presented the conclusion *All the bands at the bar will play forró or xote*. It can be noted that this structure of premises does not repeat the arguments as often as in the schema 4. The argument *forró* and *xote* are mentioned only once in the premises, so the respective nodes do not receive more activation than other concepts from the text. Therefore, the recognition of the sentence *All the bands at the bar will play forró or xote* is not facilitated as much as the conclusions of schema 4.

Another schema requiring three premises was schema 3a. This schema was applied, for example, in the story *Guide*: *Nobody from the staff speaks both Japanese and Korean. Isabel is from the staff. Isabel does speak Japanese*. Conclusion: *Isabel does not speak Korean*.). In spite of requiring three premises the arguments *Isabel* is mentioned twice, and the argument *Korean* only once in the premises. Moreover, this schema presents a negation in the first premise and requires a negation for drawing the inference. MacDonald and Just (1989) propose that negation shifts discourse focus from the negated item to one that has not been negated. Lea and Mulligan (2001) confirm, that negated concepts are less accessible that comparable concepts that are not negated. The presence of negatives both in the premises and in the conclusion could also have increased the relative difficulty of schemas 2b and 3a.

In sum, schema 4 could be perceived as relatively easy because it's premises prime repetitively the arguments necessary for the conclusion, facilitating the recognition of the conclusion presented in the test sentence. Moreover, there are no negatives that would shift the focus of the discourse from the topic of the conclusion.

The facilitation effect of repetition of the arguments in premises can be supported or weakened by the surface structure of the text. For example, schema 2b in the story *Restaurant* was introduced by the premise *None of the dishes on the menu contain red meat*, and the logical conclusion offered in the recognition item was *The 'Specialty of the Chef' does not contain red meat*. In fact, this schema requires two premises: *None of the dishes on the menu contain red meat*, and *The 'Specialty of the Chef' is a dish on the menu.* The second premise was implicit in the text, the

readers had to infer it. Should the hypothesis of facilitating the conclusion by repeating the arguments in the text be valid, omitting to mention explicitly the second premise could have lowered the recognition score on this item.

The crucial condition for facilitating the recognition of the conclusion of the schema in the test sentence does not seem to be simply how many premises a schema requires, but rather how many times the arguments necessary for the conclusion are explicitly mentioned in the text.

The present experiment identified also the schema 3b as relatively easy. What could account for this result? Schema 3b requires only one premise, so the hypothesis of a facilitation effect of multiple premises does not work in this case. Schema 3b was used, for example, in the story *Stealing.* The premise was "*The employees do not work in the warehouse having a criminal record,*" and application of the schema leads to the conclusion that "*The employees who work in the warehouse do not have a criminal record.*". Even though the conclusion has a predicate-argument structure different from the premise, the two sentences are linguistically quite similar. According to Kintsch (1998), a text is represented on three levels: surface representation, semantic (or predicate-argument) representation, and situational model. It could be argued that the predicate-argument structure of the two sentences mentioned above is different, but the situational model constructed from them could be identical. This could explain why schema the inference of the schema 3b was so often judged as a paraphrase of the text.

In sum, this experiment showed that mental predicate mental logic schemas suggested by Braine (1998) are readily available during text comprehension. Readers apply these schemas basically without any conscious effort. Triggering the schemas is tightly interwoven with cognitive processes of text comprehension (Kintsch, 1998). The main factors interfering with logical inferences in the text seem to be the additional pragmatic inferences that strengthen or weaken the trustworthiness of the premises. Grice's (1975) conversational implicatures and Cooren and Sanders's (2002) models provide useful frameworks to predict these additional pragmatic inferences. The probability of drawing a logical inference is also influenced by "who knows what" in the narrative (Gerrig et al., 2001; Lea et al., 1998).

In spite of the core mental predicate logic schemas being very easy to apply, there are some differences in their relative difficulty. Schemas 3b and 4 seem to be relatively easier that schema 2b. The relative difficulty of a logical schema seems to be related to the number of premises required by the schema, and to whether all arguments of these premises are explicitly mentioned in the text. The more often the reader encounters propositions related to

the schema in the text, the more reliable is drawing the inference. On the other hand, negatives seem to boost the relative difficulty of the schema.

The experiment showed that the use of validity task for assessing relative difficulty of drawing the inference is limited as the recall of the recognition item is influenced by numerous characteristics of the message and of the reader. The sentence recognition process has to be seen within the framework of the sentence recognition model of Singer & Kintsch (2001).

The participants of the present experiment had to draw inferences related to a single mental predicate logic schema in each story. The results provided evidence that the individual mental predicate logic schemas are drawn errorlessly and effortlessly during reading. The mental logic theory predicts that this would be true also for a text where application of several logical schemas is necessary to comprehend the story. Experiment 2 of this study tested this prediction.

CHAPTER 4: EXPERIMENT 2 – MULTIPLE SCHEMA PROBLEMS

Method

Participants.

Thirty eight undergraduate university students of psychology from Federal University of Pernambuco on a voluntary basis. All subjects were native speakers of Portuguese, between 18 and 29 years old, median age 20 (mean 21.0, *SD* = 2.8). Nine subjects were men and 29 women.

Materials and Experimental Design

The objective of this experiment was to check whether the subjects reason according to the Direct Reasoning Routine proposed by the mental logic theory. If yes, they would apply a sequence of logical inference schemas where the output of one schema is fed as a premise for the following schema, until this program arrives to the correct logical conclusion. The participants would draw the mental logic inferences errorlessly and with such ease that they would confound the inferred information with information presented explicitly in the text. The errorless use of mental logic schemas was accessed by the validity task and the effortless prediction of the theory was tested by the recognition task.

The participants were presented with 9 short narratives 5 to 10 sentences long, in Portuguese language. Each story required application of two to three of the 8 core predicate mental predicate logic schemas (see Table 3). A detailed description of these schemas can be found on pages 17 to 19. The stories were followed by two sentences, where one was a valid and the other invalid conclusion of the sequence of logical inferences in the text. Similarly to Experiment 1, the subjects had to fulfill the validity task - judge which of the two sentences was a more appropriate ending of the story.

After concluding the validity task the subjects had to carry out the recognition task: They were offered three sentences that contained information from the text and had to judge whether the information was presented in the story or had to be figured out. One sentence was a paraphrase of some information from the text (paraphrase recognition item),

another sentence was a valid conclusion of one of the mental predicate logic schemas embedded in the story (model predicted recognition item), and the third sentence was a conclusion that is valid in formal predicate logic but does not make part of the schemas included in mental logic. The experimental materials and an example of the experimental protocol can be found in Appendix B.

An example of an experimental story with the two endings and three recognition items is presented in Table 9.The story started with a title that introduced the theme of the story. The first three sentences explained the plot of the story – Peter worried whether his friends were lost or not. The first logical premise is introduced in sentence (5): Peter's friends are German. The next premise relevant to one of the schemas of mental predicate logic can be found in sentence (7): All Germans took a canoe trip. Applying the schema 2b one can conclude that Peter's friends took a canoe trip. Continuing the story, the premise in sentence (9) says that there were no tourists that took the canoe trip and got lost. This information together with the previous conclusion that Peter's friends took a canoe trip present the premises necessary to apply schema 3a, leading to the conclusion that Peter's friends, who took the canoe trip, did not get lost.

The paraphrase item refers to the information in sentence (2) and (3) of the story. The model predicted recognition item corresponds to the conclusion of the first logical inference using schema 2b based on premises in sentence (5) and (7). The foil item is a valid formal logic conclusion from premises in sentence (6) and (7) in the story.

Each story contained a different combination of premises relevant to two to three of the eight core mental predicate logic model. Each schema was used at least two times in the whole set of nine stories.

All the task features that had to be controlled for in the set of stories for Experiment 1 are relevant for this experiment as well: The mean number of words of the three recognition items, their general linguistic similarity, and the use of negatives. Also, the relative position of information relevant to the recognition items had to be equally balanced.

The mean number of words in the three recognition items was 12.4, 13.4, and 12.3 for the model predicted, paraphrase and foil item, respectively. From the 9 stories, negatives were used in each of the recognition items 4 times.

Table 9

Sample Story from Experiment 2 and its Recognition Items

	Peter's Friends
(1)	Peter was worried because two of his friends went with a group of tourists on a trip to the Amazons, and he heard that two tourists from that group got lost.
(2)	He phoned to the lodge and spoke with the receptionist.
(3)	But the receptionist did not know the names of the two lost tourists.
(4)	"Can you check what was the nationality of the lost tourists?
(5)	My two friends are German," asked Peter.
(6)	"On the day the tourists got lost we only had American and German guests in the hotel," the receptionist explained.
(7)	"And all the German tourists took a canoe trip", she added.
(8)	"OK, but do you know anything more?" Peter asked impatiently.
(9)	"Well, sir, it is confirmed that there were no tourists that took the canoe trip and got lost", the receptionist remembered.
Valid Ending:	"That's a relief, so my friends did not get lost", Peter thought.
Invalid Ending:	"So that means that my friends could be lost", Peter worried.
Paraphrase Item:	The receptionist of the hotel did not know the names of the lost tourists.
Model Predicted Item:	Peter concluded that his friends took a canoe trip.
Foil Item:	A tourist who took the canoe trip was not German or was not American.

The general linguistic similarity was counted as the proportion of words in the

recognition item that are the same and presented in the same order as they appear in the most similar sentence in the text of the story. The mean proportions of corresponding words were .28 (*SD* = .07), .28 (*SD* = .14), and .29 (*SD* = .11) for the model predicted, paraphrase and foil item, respectively.

Paraphrase items referred to information from the beginning of the story in stories number 3, 4, and 5, and to information from the end of the story in stories 1, 2, and 7. In stories 6, 8, and 9 the relevant information was in the middle of the story.

Foil items integrated information from the beginning of the story in stories 3, 4, 7, and 9; from the middle of the text in stories 1, 6, and 8; and from the end of the story in stories 2, and 5.

An additional issue in this experiment was the order of presenting the premises in the text and the distance between them. Presenting the premises in a mixed order or in distant locations within the text would put additional load on the working memory and could lead to processing errors. The objective of this study was to eliminate the comprehension and processing errors as much as possible in order to be able to measure the errors in the logical inferences. The stories were constructed so that the distance and order of the premises would facilitate the drawing of appropriate inferences (for example, presenting premises relevant to a certain schema close to each other, or not mixing premises relevant to one schema with premises relevant to another schema).

A pilot study with 8 subjects served to organize the premises in the story in a way facilitating drawing the inferences, as well as to avoid wording inviting comprehension errors and additional inferences related to conversational implicatures (Grice, 1975).

The stories were presented in two random orders and subjects were randomly assigned to one of the orders. The order of presentation of the three recognition items was randomized.

Procedure

The experiment was concluded in-group sessions. The participants were informed that the objective of the experiment was to study aspects of text comprehension. Each subject concluded both the validity and recognition task for each story. The experimental protocol contained two pages for each story; on the first page the subjects concluded the validity task and on the second page the recognition task.

The story on the first page was introduced by the instructions: “Read the following story and indicate the more appropriate ending.” Then the title and the text of the story were presented followed by two sentences. These two sentences were marked *Ending 1* and *Ending 2*, and one of them was a valid and the other an invalid conclusion of the sequence of logical inferences from the story. At the bottom of the page were instructions to turn the page in order to answer three questions related to the story. The subjects were asked not to turn back to the first page once they had passed to the second page. This was necessary as the following recognition task had to be concluded from memory without visual reference to the text.

The instructions on the second page explained that the three sentences contain information that either was presented in the story the participant just read, (although not word-by-word), or had to be figured out. For each of the three sentences, the subjects had to mark one of the three responses: “Was presented in the story (although not word-for-word)”, “Was not presented in the story (but could be figured out), or “I am not sure”. The participants were also asked not to give the same answer to all three items (like, for example, three times “Was presented in the story (although not word-for-word)”).

Results

Validity Task

The subject of analysis of the validity task was the proportion of correct answers for each story. An initial analysis revealed that there was no significant difference between the proportions of correct answers per story between the two orders of presentation, so the data sets were joined. Table 10 illustrates that the mean proportion of correct answers varied between .66 (story number 9 – *Where is the Musical Mouse?)* and 1.00 (story 7 - *Languages*), reaching the total mean of .92 (SD = .10). The mean score for story 9 was more than two standard deviations distant from the mean across all stories.

Should the participants distribute their answers randomly between “yes” and “no”, the proportion of correct answers would be close to .50. A binomial distribution analysis revealed that giving 25 or more correct responses indicates an above chance performance (Z = 1.95; p < .03). Story 9 received exactly 25 correct responses. All other stories received 34 or more correct answers, so the above chance performance was even more significant (p < .000).

The subjects made minimum zero and maximum 3 errors over the set of nine stores.

Fifty per cent of the subjects did not make any error at all, 32% made one erroneous judgment. Median error was .5.

Recognition Task

Data from stories that received an incorrect answer on the validity task were excluded from the recognition task analysis because in these cases the comprehension and inferencing processes were somewhat disturbed.

Table 10

Frequency and Proportions of Correct Answers per Story.

			Correct Answers (N = 38)	
	Story number and title	Used Schemas	Frequency	Proportion
1	A Fancy Present	3a, 4	37	.97
2	Peter's Friends	2a, 3a	34	.85
3	Marina's Exercise	1b, 3b	37	.97
4	Who will go to the Game?	1a, 4	36	.95
5	The Lunch Specials	1b, 5	36	.95
6	Experimental Drug	1a, 2b	36	.95
7	Languages	1a, 3b, 5	38	1.00
8	The House Mom Wants	3a, 4	36	.95
9	Where is the Musical Mouse?	1a, 2a, 2b	25	.66
	Mean			.92
	Standard deviation			.10

The subjects could choose between three possible answers on the recognition task. These answers were coded in a following way: An answer "Was presented in the story (although not word-for-word)" was coded as 1, "I am not sure" was scored as 0.5, and "Was not presented in the story (but could be figured out)" was assigned 0. The response "Was presented in the story (although not word-for-word)" will be referred to as an "Yes" answer, and the answer "Was not presented in the story (but could be figured out)" will be referred to as an "No" answer. For the same reasons described in Experiment 1b, this three-point scale

was recalculated to a two-point scale, collapsing the "No" answer together with the "I am not sure" answer into a response coded 0 ("No").

Mean answers on both the 3-point and 2-point scale are presented in Table 11. The mean scores on the original 3-point scale were .29, .62, and .81 for the foil, model predicted and paraphrase items, respectively. These means dropped to .17, .58, and .77 respectively on the 2-point scale. Likewise Experiment 1b, dichotomizing the scale affected mostly the foil item, which lowered by .12, as compared with the model predicted item (difference of .04) and paraphrase item (difference of .05). On the recalculated 2-point scale, the foil item ranged between .00 and .36, the model predicted item between .44 and .86, and the paraphrase item between .53 and .97.

Table 11

Mean Responses for the Recognition Task on the 3-Point and 2-Point Scale

	Mean Responses (N = 38)					
	Foil		Model Predicted		Paraphrase	
Story	3-p. scale	2-p. scale	3-p. scale	2-p. scale	3-p. scale	2-p.scale
1	.15	.14	.83	.81	.71	.64
2	.37	.18	.60	.56	.79	.79
3	.32	.16	.59	.57	.97	.97
4	.40	.34	.49	.49	.83	.74
5	.40	.22	.88	.86	.57	.53
6	.44	.36	.56	.50	.99	.97
7	.25	.08	.54	.45	.80	.76
8	.07	.00	.61	.58	.90	.83
9	.20	.08	.48	.44	.76	.72
Total	.29	.17	.62	.58	.81	.77

Note: The range of responses is between .00 and 1.00.

Figure 3 illustrates the relations between the recognition items across the nine stores. It can be noted that the foil item score was lower than the model predicted item for all stories. The highest scores for foil items were in story 6 (.36) and story 4 (.34). Foil in story 8 scored zero. The model predicted item mean was lower than paraphrase mean for seven out of nine stories. In story 1 the model predicted item reached a score .17 higher than the paraphrase item and in story 5 the model predicted item scored .33 higher that the paraphrase. On the other hand, the model predicted item in story 4, 6, 7, and 9 reached a relatively low rating (between .44 and .50). The paraphrase recognition items got to the highest scores in stories 3 and 6 (both .97). These means contrast with the paraphrase of story 5, which attained a mean answer of .53.

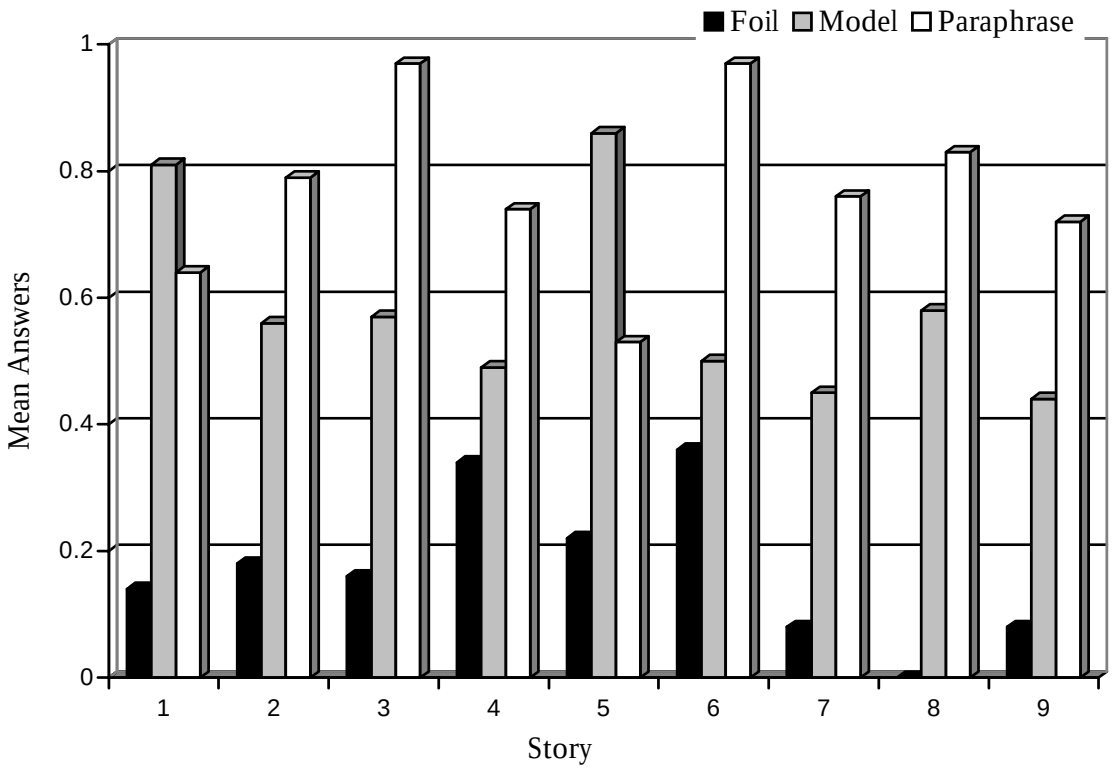

Figure 3: Mean responses for the three recognition items per story

The means of the three recognition items across all subjects were compared by

Friedman's ANOVA, which showed a significant difference ($\chi^2(2) = 65.51$; $p < .000$). Wilcoxon's matched pairs test specified that model predicted items received significantly higher mean responses then foil items ($Z = 5.12$; $p < .000$) and paraphrase items received significantly higher answers than model predicted items ($Z = 4.50$; $p < .000$). The differences within the recognition items across the nine stories were tested by Cochran Q test[3] and were significant for all three items: foils ($Q(8) = 28.07$; $p < .000$), model predicted items ($Q(8) = 30.92$ $p > .000$), and for paraphrase items ($Q(8) = 37.66$; $p < .000$).

Table 12

Comparisons of Mean Responses Between the Recognition Items

Story number	Used schemas	Response for Foil lower than for Model Predicted?		Response for Model Predicted lower than for Paraphrase?	
		Z	*p* <	*Z*	*p* <
1	3a, 4	4.24	.000	.88	*ns*
2	2a, 3a	3.41	.001	1.25	*ns*
3	1b, 3b	3.27	.001	3.77	.000
4	1a, 4	1.03	*ns*	2.67	.008
5	1b, 5	4.71	.000	-2.56	.05
6	1a, 2b	.94	*ns*	4.24	.000
7	1a, 3b, 5	3.13	.002	2.83	.005
8	3a, 4	4.69	.000	2.67	.008
9	1a, 2a, 2b	2.67	.008	2.86	.004
Across all Stories		5.13	.000	4.50	.000

Table 12 shows a comparison of mean scores on the three recognition items by stories. The model predicted items reached a higher score than paraphrase items in seven out of nine stories. The stories where these two items were not significantly different are stories 4 and 6. In stories 3, 4, 6, 7, 8, and 9 the model predicted item had a lower mean score than paraphrase

[3] Cochran Q test was chosen for comparison of dependent samples with dichotomous variables (Siegel, 1956).

item, in story 5 the model predicted item reached a higher score than the paraphrase item and in stories 1 and 2 the two items were not significantly different. Across all the stories the foil item had a lower score than model predicted item and the model predicted recognition item had a lower score than the paraphrase item.

Discussion

The objective of this experiment was to test whether readers apply the Direct Reasoning Routine during comprehension of texts containing premises of several schemas of the mental predicate logic model. The validity task showed that the subjects reached a correct conclusion in such stories in 92% of the times. This means that readers correctly apply the predicted logic schemas at the moment when premises are presented in the text, feeding conclusions of one schema as premise into the next schema, until they come to the correct conclusion of such a chain of inferences. The high level of correct judgments indicates that this process does not require any special effort from the readers.

The second task of the experiment, the recognition task, further confirmed the effortless use of mental predicate logic schemas: The inferences predicted by mental logic were judged as being presented in the story significantly more often than inferences based on formal logic in the foil recognition item.

Often the mental logic recognition items scored close to the paraphrase item, in one story even higher that the paraphrase. This means that the mental logic inferences are erroneously confounded with paraphrases of the text, applied without effort noticeable to the subjects. Thus, the results of the recognition task confirms hypothesis that the mental logic inferences are effortless.

Not all the stories behaved in the described manner. Story number 9 (entitled *Where is the Musical Mouse?*) seemed to create some problems to the readers as only 66% of them indicated the correct ending of the story. In order to come to the correct conclusion the reader had to apply three mental logic schemas: schema 1a, 2b, and 3a. Most of the stories required an application of only two schemas, nevertheless, the increased number of schemas does not explain the low score on this story because story 7 (*Languages*) was also based on three schemas and 100% of the subjects came to the correct conclusion. The choice of the applied schemas in

story 9 also cannot explain the low score, as all of the schemas used in this story were used also in at least one other story.

During development of the experimental stories the objective was to present the premises in an order that would facilitate feeding them into mental logic schemas. Ideally, the premises relevant for a certain schema were located close to each other in the text and not mixed between premises relevant for other schema. This objective might not have been achieved in story 9. This story is eight sentences long and logical premises are present in the 3rd, 5th, 6th and 8th sentence. Already the first inference that has to be drawn in order to reach the correct conclusion (applying schema 2b) takes premises from the 3rd and 8th sentence of the story. In order to draw this inference the subjects had to keep the premise from sentence 3 in the working memory all the time during reading and comprehending another five sentences until they had received the second premise in sentence 8. Such a distance between premises exerts a substantial load on the working memory capacity. Moreover, after drawing the first inference, its conclusion had to be fed as a premise in the second logical schema conjointly with premise in sentence 5 (applying schema 1a). The last logical inference (applying schema 3a) was drawn based on the output of the previous schema (1a) together with a premise presented in sentence 6 of the story.

Making such connections between premises located in different parts of the text could have negatively influenced the performance of the Direct Reasoning Routine, which links all the logical inferences in the story. This seems to be the most probable explanation of why the participants erred so often in this story.

The variance between the foil scores could be influenced by the relative difficulty of the logical schemas as well as the type of the information used in the schema. When the topic of the inference is closer to the central theme of the story the probability of recall is bigger than when it is relating to peripheral information (Singer, 1990). For example, the foil items in stories 4 (*Who will go to the Game?)* and 6 (*Experimental Drug*), which reached the highest score between the foils, refer to the central theme of the story. Foils in stories 7 (*Languages*) and 9 (*Where is the Musical Mouse?*) reached a very low score possibly because the information they treated was peripheral in the story. The foil item of story 8 (*The House Mom Wants*), which got the absolutely lowest recognition score was a dyadic conditional (i.e., a conditional that take two arguments) and included a negative. Such a logical schema might be difficult for the readers to understand.

It has been noticed during pilot studies that when participants tackled some difficult formal logic inferences in the foil item, they started to doubt whether this information is a valid inference at all. In spite of the instructions explaining that all the information in the three recognition items is valid, the subjects tended to wonder whether the conclusions of the foil items are valid or not, instead of judging whether they were presented in the story or not. As the validity of the foil in story 8 (*The House Mum Wants*)was rather difficult to evaluate, the participants could have marked this item as not presented in the story.

The model predicted recognition items also varied significantly between the stories. Again, the variation could be explained by a combination of factors. One factor could be, for example, the position of the relevant information in the text: Inferences referring to information from the beginning of the text could have been forgotten by the time the subjects were fulfilling the recognition task.

Another factor that could have influenced the judgments on whether the information was presented in the story or not, was the linguistic similarity. The number of words that are the same and in the same order as the most similar sentence in the text varied between 19% and 42%. For example, story 5 *(The Lunch Specials*), which had a very high score on the model predicted recognition item had also the highest linguistic similarity of the model predicted items (42%). Nevertheless, story 1 (*A Fancy Present*), which reached also a relatively high score, was much less similar to the relevant sentence in the text (20%). Linguistic similarity alone cannot explain the differences in recognition scores.

These two stories (number 1 and 5) had another interesting aspect: The model predicted recognition item scored even higher that the paraphrase item. This result confirms that a paraphrase and a inference of the mental predicate logic is perceived by readers as relatively the same difficult.

A relatively lower score of the paraphrase item of story 5 (*The Lunch Specials*) could be possibly explained by an influence of some invited inferences. The story is about two sisters having lunch. The second sentence states that they wanted to order some of the lunch specials. The rest of the text described how the sisters compared the different offers of the lunch specials. The paraphrase recognition item stated that the two sisters wanted to order some lunch specials during lunch. The reader might have not agreed with this item because he/she might have not recalled this information being presented in the second sentence of the text. Instead, he/she might have concluded that he sisters were just talking about the offers and, at the end, they could have changed their mind and not order any of them. In fact, the

story ended before the sisters made any order. Also in this case, the reader should have forgotten or not trusted the instructions stating that the information of the recognition items is valid information from the story, either a paraphrase or an inference. Instead, of judging whether the recognition item was presented in the story of «r not, the readers analyzed whether these sentences are valid conclusions from the story or not.

In sum, the variance between the recognition items across the nine stories can be caused by many different factors. The experiment was designed to illustrate the general tendency of the recall of the three types of recognition items. It did not permit to detect the variables that influenced the score on a specific recognition item in a specific story. Such an analysis would require, for example, a greater number of stories with a different content but the same logical structure. A qualitative analysis instrument, such as protocols registering all the inferences and justifications of the subjects during recognition task, would provide more insight on the reasoning process and on the influence of text comprehension factors.

The results of the experiment confirm the prediction of the mental logic theory that the schemas of the model are applied during reading errorlessly and effortlessly using the direct reasoning routine. A successful application of the Direct Reasoning Routine depends on the organization of the premises in the text. Texts where the distance between premises is relatively big, create additional load on the working memory capacity, which can lead to processing errors. The subjects seem to have the set of predicate mental logic schemas on disposition, but in cases of complicated texts with several premises they can fail to keep the track of the sequence of premises and inferences, which prevents them from reaching the correct conclusion.

CHAPTER 5: EXPERIMENT 3: ON-LINE INFERENCES

Method

Participants

Fifty four undergraduate students participated to fulfill a part of a course requirement in Introductory Psychology at the Baruch College of the City University of New York. Sixty five percent of the participants were native English speakers.

Materials and Experimental Design

The motivation of this experiment was to assess whether a particular logical inference is being drawn on-line, that is, whether it is being drawn as the requisite information enters working memory. It was predicted that during text comprehension the logical inference would be drawn automatically as soon as the logical premises were readily available in the working memory. The experiment was designed to reach this objective using a *naming task*.

The subjects were asked to read short stories in English that had the *or-elimination* predicate-logic schema from the mental-predicate-logic model included in their structure (schema 1a on page 17).:

S1[All X] OR S2[PRO-All X]; NEG S2[α]; α ⊇ [X] /∴ S1[α]

Stories were constructed in pairs, with each pair containing an Experimental version and a Control version. Each story was five sentences long. The Experimental version of each story contained the major logical premise for Schema 1a in the second sentence and the minor premise for the schema in the fourth sentence. Immediately after the introduction of the second (minor) premise, a word appeared on the screen. This word was the target item for the naming task. The participants were instructed (and trained) to read out-loud such a word immediately as it appeared on the screen (the *naming task*). In the Experimental version this word corresponded to the output of the predicted logical inference. The same word was used as a probe for the Control story versions, but the conditions for drawing the inference were absent, so the word should not be primed in this case.

The control version of the story differed from the experimental version in that it did not contain the second minor logical premise in the fourth sentence. Even thought no logical inference could be drawn, the participants were being asked to fulfill the naming task based on the same word probe as in the experimental story.

In the experimental versions of the stories, the two logical premises were separated only by one sentence, and therefore, according to currently accepted models of comprehension (e.g., Kintsch & van Dijk, 1978), should be simultaneously present in working memory. The mental-logic model predicts that in this situation the logical inference should be drawn automatically as a function of the information from the major and minor premises being considered conjointly in working memory. Should this assumption of mental-logic theory be true, then at the moment when the participants were receiving the naming task, the word probe should already be primed and present in working memory as a result of the logical inference. The control version of each story pair should not create the conditions necessary for priming the word probe. Faster reaction times in the experimental situation compared to the control one thus would suggest that in the experimental version the word probe was primed and in the control version it was not. It is likely that in the experimental version the word probe would be primed because subjects would have drawn the logical inference. This would indicate that the inferences were made as predicted by the mental-logic model: At the moment both logical premises are readily available in the working memory, the logical inference should be drawn on-line.

The last, or fifth, sentence of the experimental story presented either a valid or an invalid conclusion as judged by the logical inference, and the last sentence of the control version was either a consistent or an inconsistent elaboration of some of the facts from the story other than something that relied on an inference of the mental-logic theory. After reading this last sentence the participants were asked to answer whether this last sentence made sense in the context of the story. They were led to believe that the purpose of the experiment was to test aspects of text comprehension. In the context of the experiment, answering the last sentence served to ensure that the participants paid the necessary attention to the texts.

The experimental and control stories were embedded between another set of stories that were referred to as filler stories. A subset of the filler stories did not contain any logic particles at all, whereas another subset contained logic particles; In either case, the naming task word probes were not the result of a logical inference, but instead they simply repeated some other word from the story. Introducing the filler stories had the objective of disguising the

purpose of the experiment for the subjects by avoiding a prevalence of the same story structure across the whole set of stories being presented.

The detailed structure of the experimental and control versions of the stories is illustrated in the following example:

WHAT DOES THE BOSS DRINK?

(1) On her first day on the job the secretary was supposed to get refreshments at the meeting of the directors.

(2) Everyone at the meeting wanted either tea or juice.

(3) The secretary became very nervous because she was not sure which thing it was her boss wanted to drink

	EXPERIMENTAL VERSION:	**CONTROL VERSION:**
(4)	Suddenly she remembered about her boss' allergy to fruit acids: 'He definitely does not want juice.'	And, as if her troubles were not enough, she couldn't find enough clean glasses for the juice.
(5)	****** TEA ******	
(6)	Her boss wanted tea.	Her first day on the job was a success.

The title and Sentences (1), (2), and (3) and the naming task word probe (5) are identical for both the experimental and control version. After presenting the title, Sentence (1) introduces the theme of the story. Sentence (2) contains the first premise of the mental-predicate-logic schema. (*Everyone wants either tea or juice*.) The filler, Sentence (3), develops the "plot" of the story. Sentence (4) differs in the experimental and control version. On the left side of the example the experimental version of Sentence 4 presents the second premise (the minor premise) of the mental-logic schema (*The boss does not want juice*), but on the right side of the example the control version of Sentence 4 does not present a minor promise, leaving the

participant without sufficient information to apply the mental-logic schema. After sentence (4) the participants were asked to read out loud the word *TEA* (5). This is the naming task probe, and for the experimental story version, it is the result of the mental-logic inference.

Sentence 6 in both problem versions is a comprehension sentence. The comprehension sentence (6) in the experimental version (on the left side of the example) is a valid conclusion based on the predicted mental-logic inference. The subjects were required to indicate whether this sentence made sense as the ending of the story, which in this case required a response "*yes*". In the control version of the story (on the right side) sentence (6) is an elaboration of the facts introduced in the story and the subjects had to indicate whether or not this sentence makes sense as the ending of the story (which in this example requires a "*no*" answer). Participants were asked to indicate the consistency or inconsistency of the last sentence of the story by pressing a button on a box located on the table in front of the participant.

In both versions of the story the naming task was situated in the same instant of the text comprehension process: after the fourth sentence. This sentence had a key role in the experiment, as it determined whether or not the naming word probe was primed. In the Experimental version, the word should be primed as a result of the mental-logic inference, whether in the Control version not. It had to be avoided that the word would be activated for reasons other then being the result of the logical inference, for example, because of semantic priming. Therefore, the fourth sentence in the Control version always contained the key word of the second premise. It was assured that this sentence also ended with the same word as the experimental version. (In the example above the word *juice* is the key word of the second premise and in both versions the fourth sentence ends with this word.) Also, the average length of the fourth sentences had to be the same for Experimental and Control versions.

A special attention was given to the use of negatives. All the experimental versions of the stories necessarily contain a negative in the fourth sentence due to the structure of the logical or-elimination schema: the minor premise has to eliminate one of the arguments. The text comprehension research indicates that processing negatives requires some additional effort as compared with sentences without negatives (e.g., Lea & Mulligan, 2001). Use of negatives in the fourth sentence could therefore influence the reaction times of the naming task. To be able to control this influence, half of the control version stories also contained a negative.

The experimental material consisted of 66 stories: 36 filler stories and 30 trial stories. All trial stories were prepared both in their experimental (inference) version and

control (no-inference) version, and with a consistent and an inconsistent ending. Each individual subject was presented only one version of each trial story.

The filler stories were of three types (see Appendix C): Filler A were 10 stories with the same structure as the experimental stories, but the naming probe was a word from any part of the story except the fourth sentence. Filler B stories had the same structure as the control no-inference stories, but again, the naming task probe was a word from other parts of the story. The 16 Filler C stories didn't contain any logical premise at all, task probe was any word from the story. From the total 66 stories, 25 stories contained two logical premises, 25 stories contained just 1 premise, 16 stories had no premise at all. Naming word probe was taken from in the middle of the story 30 times, from the beginning of the story 18 times, and from the end of the story 18 times.

Each participant was presented with 33 stories with a consistent last sentence and 33 stories with inconsistent ending. The last sentences of the stories were designed in such manner that they should provoke clearly either a *yes* or a *no* answer. This measure allowed checking for English text comprehension abilities (35% of the subjects were not native English speakers). A high number of incorrect answers was likely to occur when either a participant's fluency in English was not sufficient for this experiment, or when he/she did not devote the necessary attentional resources to the experiment.

In order to control for the influence of negative in the sentence before the naming task, half of the control versions of the stories contained a negative in this sentence and half did not. (Items 1, 2, 3, 4, 5, 6, 10, 11, 13, 16, 20, 23, 26, 29 contained a negative, see Appendix C.)

The average length of the sentence before the naming task was 14 words for both the experimental and control versions of the stories. This sentence always ended with the same word for the experimental and control version of the same story.

From the point of view of text comprehension the stories for the experiment had to be carefully constructed avoiding equivocal interpretations or undesirable invited logical premises. A necessary part of the process of developing the experimental material was a pilot study during which 15 subjects were asked to work through the whole set of 66 stories. During the pilot study the experimenter closely monitored performance of the participants and checked their interpretations of the texts.

To summarize, each participant read 66 stories, 15 in the experimental version, 15 in control version, and 36 filler stories. Participants were randomly assigned to one of two groups.

Each participant in Group A received the stories 1 to 15 in experimental (inference) version and the stories 16 to 30 in control, no-inference version. Participants in Group B read stories 1 to 15 in no-inference version and stories 16 to 30 in the inference version. Thus, the within subjects factor (inference set) and the between subjects factor (group) yielded a split-plot, Group x Inference Set design. It was predicted that the main effect for the inference set would be significant, but that neither the main effect for group nor the interaction would achieve significance. This design is illustrated in Table 13.

Table 13

Experimental Conditions of the Four Groups of Subjects

Group	Version of items 1-15	Version of items 16-30
Group A	Inference	No-inference
Group B	No-inference	Inference

Procedure

Presentation of the experimental material and recording the reactions of the subjects was programmed in PsyScope 1.4 software. The program was run on a MacIntosh computer connected with a Button Box and a microphone. The participants were sitting in front of the screen of the computer with their fingers on the Button Box, speaking directly in a microphone.

The stories were presented to the participants in a random order. Each story was presented on the screen of the computer one sentences at a time.

A story trial began with the title presented in the center of the screen. When participants were ready to begin reading the story they had to press the yellow button on the Button Box. Pressing this button always replaced the current sentence with the next one on the screen. When participants read the sentence immediately preceding the naming task, their button press cleared the screen and activated the naming probe. An array of Xs appeared on the screen for 500ms followed by the naming word probe, which was presented in

capital letters in the middle of the screen between asterisks. Participants understood that they should read the word out loud immediately as it appeared on the screen. The reaction time (RT) between the moments the word appeared on the screen and the subject's utterance was registered. After the naming task the last sentence appeared on the screen and the participants had to indicate whether or not this sentence makes sense as the ending of the story. Pressing the green button on the Button Box indicated a "yes" answer, pressing the red button meant "no". Participant's response to the comprehension sentence advanced him/her to the next story.

The whole experimental session started with instructions and two practice stories. As the experiment demanded full attention of the subjects, the participants were encouraged to take a break between the stories whenever they felt tired. An obligatory break was introduced in the middle of the experiment. The whole experimental session took approximately 50 minutes.

Results

Outliers[4]

First, the natural logs of the reaction times were computed in order to normalize the data and minimize extreme points. For each participant, regression between the RT and the order of presentation of the story was calculated. The residuals from this regression were used to eliminate outliers. All data points further than three standard residuals were eliminated, which resulted in elimination of 43 RTs (2.6%). The regression analysis was then repeated without these outliers and 7 more RT further then three standard residuals were detected and eliminated. On total, 3.0% of the data was eliminated as outliers. The relation between the RT and order of presentation of the story was significant for 11 out of 54 subjects

[4] Analysis of RT are quite sensitive to outliers, yet there are no standard procedures for dealing with them. Some investigators identify outliers in terms of absolute values (e.g., <200ms or >2.500ms) on the basis of reasonable RTs for the process of interest. Others use standard deviations of some empirical distribution, often cutting at 2.5 or 3.0 SD. Once outliers are identified, some investigators drop these data points; others substitute the value of the cutoff for the observed RT (for overview, see Barnett & Lewis, 1994). For this experiment, the outliers' elimination procedure was equal to that adopted by Lea (1995) and Uleman, Hon, Roman, and Moskowitz (1996).

Incorrect answers

The subjects judged incorrectly the consistency/inconsistency of the last sentence of 13% of all the stories. The median number of errors per subject was 7. Five subjects with number of errors bigger than twice the median were eliminated (16, 17, 17, 18, and 18 errors, see histogram on *Figure 4*). The number of subjects therefore dropped from 54 to 49.

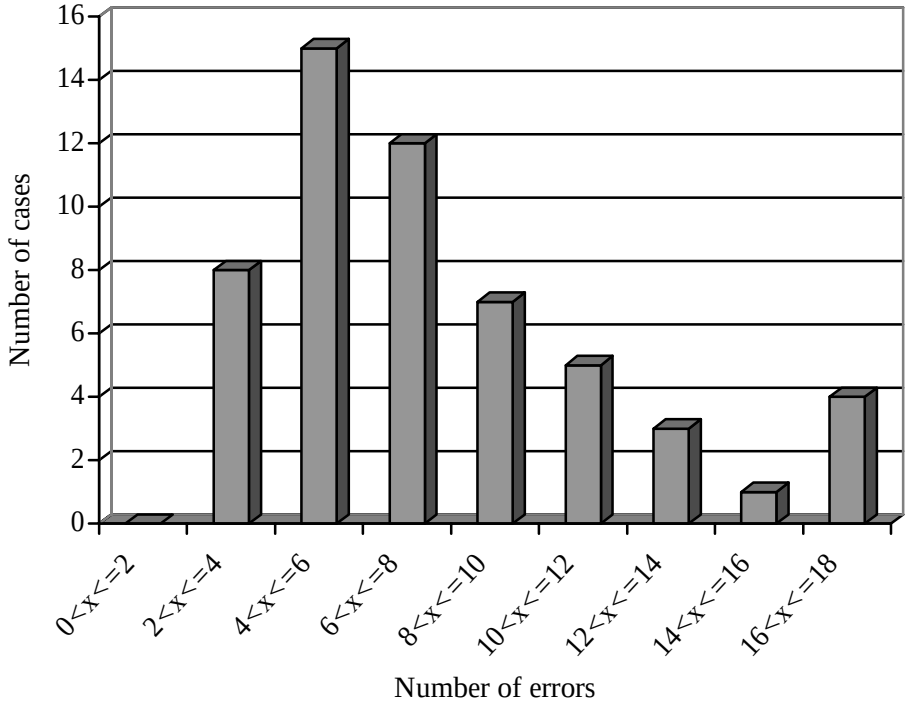

Figure 4: Histogram of errors by subjects (N = 54)

Each story was presented 49 times, to approximately half of the subjects in the experimental version and to the other half of the subjects in the control version. The median number of incorrect answers per story was 6, varying between 0 and 12. Story 7 was an exception, as it induced an incorrect answer in 16 cases. This result represents more then twice the median, so it was suspected that the story had not been constructed properly for this experiment, allowing equivocal interpretation. Therefore, all the RTs related to this story were excluded from the data.

The data set still contained 12% of RTs related to incorrectly answered stories, 10% in the experimental versions and 14% in the control versions. It was necessary to decide whether RT of the naming tasks of *all* incorrectly answered stories had to be eliminated from the data set or not. For the experimental versions of the stories, the mean RT for correct answers (535 ms) was 18 ms lower than the mean RT for incorrect answers (553 ms). For the control versions the RTs were almost identical for correctly and incorrectly answered stories (555 ms and 553 ms, respectively). It was concluded that the higher RT for an incorrect answer indicates that the word probe was not primed because the subject did not pay sufficient attention to the text of the story. Therefore, he/she did not draw the expected inferences necessary for text comprehension. All the RTs related to wrongly answered stories were therefore eliminated from the analysis (similarly to Lea, 1995).

Mean Reaction Times

Mean RT on the naming task of all the subjects reading the stories in the inference (experimental) version was 534ms (*SD* = 1ms), as compared with the mean of 554ms (*SD* = 1ms) of all the subjects reading the stories in a no-inference (control) version. These means represent the RTs after they were transformed to logs, averaged by participant and then transformed back. The variance due to order of presentation of the story was not eliminated from these means. As the means illustrate, the word probes were read more quickly in the inference version stories than in the control, no-inference versions of the stories. The difference between the two means is 20ms.

The experiment was conducted in a split-plot Group x Inference design (see Table 13). In order to verify the differences between the means, a 2 (Group: A, B) x 2 (Condition: Inference, No-inference) ANOVA was carried out. The main effect was significant for the inference set, ($F(1,47) = 25.89$, $p < .000$), but not for group. The interaction between group and inference set also reached significance ($F(1,1) = 4.26$, $p < .045$), see *Figure 5*. A Tukey's post-hoc test showed that the difference between the means of inference and no-inference stories was significant for both Group A ($p < .05$) and Group B ($p < .01$).

Another ANOVA computation was carried out in order to compare the means across the experimental and control items. The mean RT of all the stories in the inference version was 532ms (*SD* = 1), and the mean of the no-inference stories was 553ms (*SD* = 1). The inference version mean was 21ms lower than the no-inference mean. This difference was significant at $F(1,28) = 8.06$, $p < .008$.

Influence of negatives

The significant difference between RTs holds also when the analysis was run only with those stories in which both experimental and control versions contain negative in the fourth sentence before the naming task. This criterion reduced the number of analyzed items to 15 stories. The mean RTs of such stories was 560 ms and 540 ms for the corresponding no-inference and inference versions, which was significant both by subjects, $F(1,47) = 39.25$, $p < .000$, and by items, $F(1,14) = 8.06$, $p < .02$. Within only the control stories, the difference of mean RTs between the items with and without negative in the fourth sentence did not reach significance.

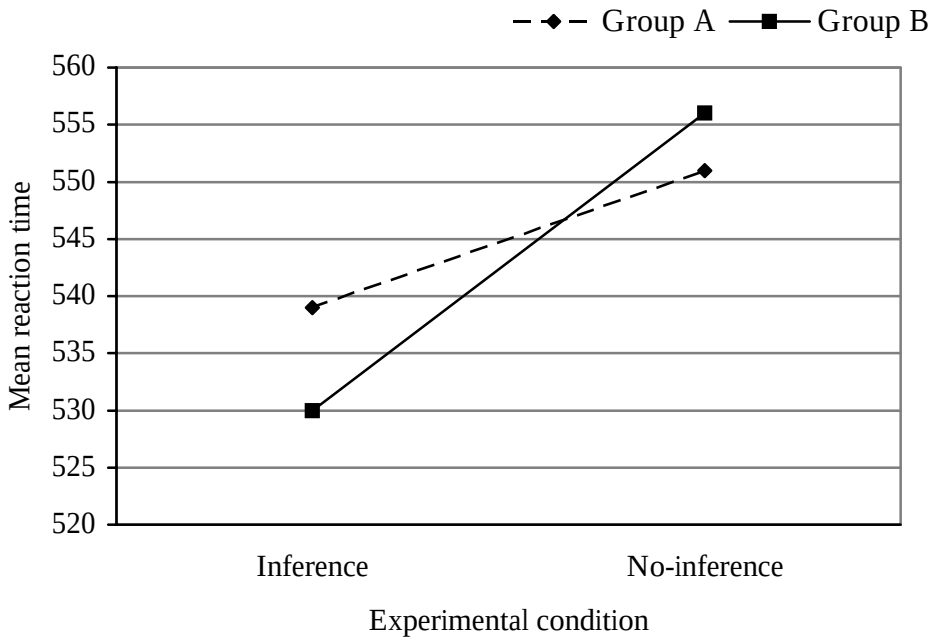

Figure 5: The mean RT on the naming task for the two groups and two inference conditions.

Discussion

The outcomes of the naming task showed that when readers were presented with the inference version of the stories, they reacted significantly faster on the naming task than in the control condition, where the story did not trigger such a logical inference. This difference in

reaction times can be explained by the priming effect that facilitated naming of the word probe in the inference condition. The logical inference schema required two premises. It can be assumed that when the subjects reached the second premise in the inference version story, they drew the logical inference, by which the word resulting from the inference was primed. When they were asked to say out loud such a word, the reaction time was shorter then in the control condition. In the control condition the story was identical except that the second premise was missing. Therefore, the logical inference could not be drawn, the word target was not primed, and the naming task reaction time was significantly higher than in the inference version of the story. The results indicate that the *or-elimination* schema inferences of mental predicate logic are drawn on-line during text comprehension at the moment when the necessary premises are conjointly held in the working memory. The logical inferences were not required for the local coherence of the texts.

During the experiment, all the stories were used in both inference and no-inference versions, although each story only in one version for each participant. The participants were divided into two groups that differed in which story was in the inference and which in the no-inference version. The pattern of mean reaction times was equal for both groups: the inference stories reached a lower mean RT on the naming task than the no-inference versions. Nevertheless, the difference between the two experimental conditions was not equally strong for both groups; as shown in *Figure 5*, one group of participants reached a higher difference than the other one. It has to be noted, that the experimental and control stories were inserted within the filler stories, and the order of presentation of the whole set of 66 stories as well as the assignment of the participants to one of the two groups was random. Therefore, the interaction between the group and inference condition could not be due to order of presentation of the stories.

The reaction time data were subsequently checked in search for another explanation for the interaction between the group and the inference set. Obviously, for most of the subjects the means of the inference stories were lower than for the no-inference stories. Nevertheless, RTs of some participants showed no difference between the two conditions, and in several cases (11 out of 49) the means revealed an inverse effect – the inference version mean reaction times were higher than the no-inference version RT. This inverse effect occurred in only 3 of the participants in Group B, but in 11 of the participants of Group A. Such an unequal division of these extreme values could account for the difference between Group A and Group B and was most probably caused by random influences.

The results of this experiment replicate closely the outcomes of the experiments of Lea (1995). Lea tested the on-line use of inferences related to two schemas of the mental propositional logic: the *or-elimination* and *modus ponens*. During reading short texts, both schemas were reliable drawn on-line. The naming task was used for testing the or-elimination schema. The mean reaction time in the inference and non-inference version of the stories was 507ms and 526ms respectively, resulting in a difference of 19ms. In our experiment the respective mean reaction times were about 30ms higher reaching 534ms and 554ms for the experimental and control version, respectively. The difference between the absolute RTs of Lea's experiment and the present experiment could have been caused by some technical aspects of the experiment, for example by the device that was used to measure the reaction times. Nevertheless, the difference between the inference and non-inference condition in the present experiment was almost identical to the result of Lea: 20ms.

The naming task reaction time is apparently the time necessary to decode and pronounce a word written on a screen in front of the subject. The interval of 20ms represents the difference between the reaction time when the word is already primed and presumably present in the working memory, and when it has to be drawn from the long-term memory.

It could be suspected that the results presented in this experiment were influenced by other factors than the presence of an on-line inference. For example, during the course of the experiment the subjects might have learned the basic structure of the stories and adopted strategies based on patterns of materials (- like, for example, looking for premises and anticipating an logical inference in the word probe). Each participant received a collection of stories where less than half of the texts contained two logical premises. From the total set of 66 stories, the naming task word probe was the result of a logical inference in only 15 stories. Such a pattern could hardly permit any kind of strategic processing.

The comprehension question served to motivate the readers to read the stores attentively. It was desirable to ensure that the cognitive processes involved in reading the stories were those typical for normal text comprehension, free of any unusual disturbances. The comprehension process could have been disturbed, for example, by insufficient proficiency in English. The participant pool was from an American university, nevertheless, only 65% of the participants were native English speakers. Language comprehension difficulties could cause distortion of reaction times in the naming task, as well as incorrect answers to the comprehension questions. The reading process would be also abnormal if the participants would not invest the necessary intentional resources and merely skim the text. In

this case, other cognitive processes would be involved, different from the common text comprehension processes. A consequence of such inattentive reading would be that the reader would not be able to answer correctly the comprehension question, and the naming task reaction times would be distorted, too.

The texts of the stories were rather simple, and most comprehension difficulties were eliminated during pilot studies. When the subjects did not answer the comprehension question correctly during the experiment, then they most probably either did not invest the necessary cognitive resources to comprehend, or were not sufficiently proficient in English, as argued above.

This assumption was confirmed by the pattern of reaction times. In the non-inference stories the reaction times of the naming task in correctly and incorrectly answered stories were practically the same. On the other hand, in the inference version of the stories, the correctly answered stories had mean reaction times 18ms lower than the incorrectly answered ones. Only during attentive reading the inferences that primed the naming task probe were drawn, and the comprehension question was answered correctly. When the text was merely skimmed, the inferences were not drawn, and the comprehension question was not answered correctly. This is why subjects with unusually high error rate on the comprehension question were eliminated from the study.

Individual reaction time measures in the naming task are sensitive to interruptions, caused by momentary lapses of attention or disturbances in the surrounding of the subject. Technical problems and speech disorders also impede that correct reaction time measures are taken. All these interferences reflect in isolated unusually long reaction times that have nothing in common with whether the word probe has been primed or not. Such extreme reaction times were classified as outliers and eliminated from the data.

It could be argued that during reading there are multiple other factors that can influence the probability of drawing an inference. It was desirable to reduce the possibility that the inference drawing in the experimental version of the story was caused by factors of discourse processing other that the availability of the schema of the mental logic. One of the possible arguments would be that the naming task word probe might have been activated because residual activation from the presentation of the concept in the text (semantic priming) and not due to an inference. Even though the naming task is not as sensitive to semantic priming as, for example, the lexical decision task, some measures were taken to avoid such accidental influence. Both the experimental and control version of the story contained

the same key concept in the sentence preceding the naming task. In the experimental version the concept was introduced by the second premise, in the control version it was simply mentioned. Therefore, semantic priming could not account for the results of the experiment.

The design of the experiment also ensured that the sentence preceding the word probe ended with the same word in both the experimental and control version, and that it was in average of the same length.

Another important factor that is known to influence text processing is the presence of negatives (e.g., Lea & Mulligan, 2002; MacDonald & Just, 1984). The nature of the or-elimination logical schema determines that the second premise contains a negative (For example: All X are a or b; a subset x is *not* a; therefore, a subset x is b). In all the inference version stories, the second premise, containing the negative, was present in the fourth sentence just before the naming task. In order to control for the influence of the negative, half of the no-inference versions of the stories also introduced a negative in the fourth sentence. The results show that the significant difference between the reaction times of the inference and no-inference stories holds even when only the stores containing the negatives are included in the analysis. Within only the no-inference stories there was no difference between the mean of the stories with and without negatives. This result confirms that the differences in reaction times on the naming task are not influenced by the presence of negatives.

Summarizing the preceding arguments, the conditions in the experimental and control version of the story were basically equal, except for the presence of the second logical premise that would trigger the logical inference. It can be argued that the difference in reaction times between the inference and no-inference versions of the stories was not influenced by other discourse processes and serves as an indicator that mental logic schemas, such as the or-elimination schema, are drawn on-line during text processing.

The example of the or-elimination schema implies that some forward logical inferences are drawn automatically on-line. Such result is contrary to most of the present models of inference drawing. The minimalist hypothesis of McKoon and Ratcliff (1992) stated that, in absence of special strategies, readers make inferences only under two conditions: (a) when the inferences are necessary for local coherence and (b) when they are drawn from information readily available. As has been noted earlier, the logical inferences tested in the Experiment 3 are not required for local coherence of the text, on the contrary, they can be seen as forward inferences which add information beyond that what was presented in the text, most often about the consequent actions of the actors of the narratives. One could agree that the

information in the premises of a logical inference is readily available, thus the mental logic inferences might be explained by the minimalist hypothesis. On the other hand, the condition related to readily available information is rather vague – innumerous information is readily available in the text as well as in the reader's memory, and still the subjects obviously do not draw all the possible inferences from it. Therefore, the minimalist hypothesis does not seem suitable for predicting the logical inferences found by this experiment.

Graesser et al. (1994) give different explanations on when and what kinds of inferences are drawn. Their constructionist theory is based on the assumption that the reader draws inferences that serve him/her in the search after meaning. The reader attempts to construct a meaningful representation of the text that (a) addresses reader's goals, (b) is coherent both locally and globally, and (c) explains why actions, events and states are mentioned in the text.

The logical inferences in this experiment were not required for local coherence of the text, so the condition (b) of the constructionist theory does not apply to logical inferences. It can be considered whether the logical inferences were drawn because they addressed the reader's goal as stated in the condition (a) of Graesser et al. (1994).

Several investigators pointed out the importance of the goal of the reader as well as the thematic focus of the text. McKoon and Ratcliff (1992) propose that when the surrounding text supports the inference concept, or when such a concept is in the focus of attention of the reader, the probability of drawing the inference is higher. Graesser et al. (1994) conclude that the goal of the reader plays a crucial role in drawing inferences on-line or off-line. Kintsch (1998) argues that when the instruction, task and material are properly tuned, many different types of inferences can be drawn. What role did the thematic focus and the goal of the reader play in the results of the present experiment?

The thematic focus is induced by the title of the text. All the stories in our experiment were introduced by a title, which could have stressed the theme of the story. As the experimental texts were rather short and with a simple plot, the logical inferences were indeed related to the theme of the story. Therefore, the drawing of the logical inference could have been supported by this factor. On the other hand, it can be argued that the title and the first three sentences were equal for both experimental and control version ensuring that the thematic focus was equally introduced. In spite of that, the naming task reaction times differed between the experimental and control version suggesting that in the experimental condition the logical inferences were drawn and in the control condition they were not.

The goal of the reader was determined by the instructions given to the participants. The participants in this experiment were asked to read the stories attentively to be able to answer the comprehension question, and to react on the word naming task immediately. Again, these instructions were common to both the experimental and the control version of the stories, so the goal of the reader was set equally for both conditions. In spite of that, logical inferences appeared only in the experimental version, triggered by the presence of the premises and not by any special goal.

The suggestion that the logical inferences are drawn independently on the thematic focus has already been confirmed by Lea (1995) who compared on-line logical inference processes in two conditions: In texts introduced by a title and in texts without a title. He found that evidence for on-line logical inferences in both title and no-title experiments though the effects in the stories without title was slightly smaller.

Condition (c) mentioned by Graesser states that inferences are drawn when they explain why actions, events and states are mentioned in the text. This prediction seems relevant to logical inferences as a similar suggestion was mentioned in Braine and O'Brien (1991) in relation to the theory of conversational implicatures suggested by Grice (1975). Grice deduced that in everyday discourse people make their contributions as relevant as required for the purpose of the current conversation and provide neither more nor less informative than required. This seems to be in agreement with Graesser's statement that when some actions or events are stated in the text the readers make inferences about them in order to explain why this information was mentioned. This explanation could account for logical inferences during text comprehension:

In sum, logical forward inferences do not seem to fulfill the condition (a) and (b) of the constructionist theory of Graesser et al. (1994), in the sense that they do not address the goals of the reader and are not required for the coherence of the text. On the other hand, they can be accommodated within the condition (c) as they can explain why actions, events and states are mentioned in the text.

There is another aspect of the logical inferences that seems to contradict even the position of Graesser et al.: Logical inferences are related to future events in the text and as such their occurrence is viewed skeptically by most of the researches on comprehension: "The constructionist theory predicts that readers do not normally construct inferences that forecast future episodes in the plot" (Graesser et al., 1994, p.372). The reasons often cited as to why forward inferences are not drawn on-line are that there is a risk that one will draw the

wrong inference or that the inference will not be necessary for understanding the text (e.g., Murray et al. 1993). Therefore, it is not efficient to draw such inferences. The difference between mental logic inferences and other forward inferences is that logical inferences are not probabilistic but deterministic: if the premises are true than the inference must be true. Logical inferences are more "cost effective" compared with other forward inferences, so they are drawn in spite of the fact that they forecast future episodes of the text.

A solid theoretical framework that would accommodate logical premises in text comprehension is offered by Kintsch (1998). According to Kintsch's construction-integration model, text and discourse comprehension proceeds in two stages. In the first, construction, stage the information from the text base as well from the knowledge base of the reader is encoded into a network of propositions. Many information elements are activated at this phase. In the second stage the propositional network is integrated by strengthening the contextually appropriate nodes and eliminating the inappropriate ones. Kintsch describes in detail the process of activating the relevant information in the long-term memory, which expands the working memory capacity to the extent necessary for comprehension of a text. The process of retrieval of information from the long-term memory Kintsch does not consider as inference drawing. Logical premises are according to Kintsch true premises, because they convey information neither present in the text, nor in the knowledge base of the subject. Nevertheless, Kitsch considers logical inferences as generated by controlled processes, whereas Experiment 3 suggests that they are drawn on-line, therefore automatically.

Murray et al. (1993), Klin, Murray, Levine, and Guzmán (1999), and others propose that there seems to be a variety of sub-types of forward inferences. Several authors (McKoon & Ratcliff, 1986, Potts et al, 1988, and Murray et al., 1993) have suggested that some inferences serve to repair a "causal coherence break", whereas others are "purely elaborative". The "causal coherence break" types of inferences represent a likely consequence of the action in the story, but also a cause or motivation of another action described in the story. Forming causal connections is central to building a coherent representation of the text (e.g., Klin, 1995; Trabasso & Sperry, 1985), so it is to be expected that this type of forward inferences should be reliably drawn. In contrast, purely elaborative inferences, serving only to embellish the text, are drawn under a far more restricted set of conditions (Klin, Guzman, & Levine, 1999).

The texts used in Experiment 3 can hardly be classified as serving to repair a causal coherence break. The or-elimination schema was applied in situations, where some of the characters have to decide their future action (for example, to buy a yellow dress or a blue dress,

to spend vacation on the beach or in the mountains, to sleep in beds or in hammocks). In other cases the schema served to specify a certain situation (patient has tumor or TB, the girls are doing judo or tennis, someone bought a table made of either glass or wood). Therefore, the investigated inferences seem to be of the purely elaborative type rather than explaining a cause or motivation of another action of the story. Moreover, the naming task showed that the inference was drawn before any consequent action could be described in the following sentence. In spite of that, the or-elimination inferences were reliable drawn.

In sum, the experiment showed that the prediction of the on-line application of the schemas of the mental predicate logic is correct for the or-elimination schema. This result contradicts some of the influential theories of text comprehension, such as the minimalist theory of McKoon and Ratcliff (1992), but fits within the constructionist model of Graesser et al. (1994) and text comprehension theory of Kintsch (1998). Up till today, three schemas of the mental propositional logic (Lea, 1995; Lea & Mulligan, 2001) and one schema of the mental predicate logic (present experiment) have been tested for their automatic application. The results showed that they are all reliably drawn online. It can be expected that most or all of the remaining schemas specified by Braine and O'Brien (1998c) for the mental propositional logic and Braine (1998) for the mental predicate logic would behave in a similar way. The mental logic theory provides solid predictions for drawing forward logical inferences during text comprehension.

CHAPTER 6: GENERAL DISCUSSION

The study aimed to test four of the prediction of the mental logic theory introduced by Braine and O'Brien (1998a). Experiment 1a and 1b tested the errorless and effortless use of eight of the propositional mental logic model when introduced separately in short texts. The Experiment 2 introduced several such schemas in one story and aimed to detect whether these schemas are used according to the proposed Direct Reasoning Routine, feeding the output of one schema as a premise for the following schema. Also in this experiment the objective was to see whether the participants come to the correct conclusion without errors and effortlessly. The Experiment 3 tested the prediction of automatic on-line use of such schemas in text processing.

Summary of the Results

All the above-mentioned predictions seem to be supported by the results. Experiment 1a showed that subjects have no difficulty in applying one of the core mental predicate logic schema in short texts, drawing the correct conclusion in up to 97% of the cases. When there were two to three schemas presented in the texts of Experiment 2, the participants successfully applied the Direct Reasoning Routine and came to the correct conclusion in 92% of the cases.

Comparing the logical inferences to paraphrases of the text tested the effortless use of the mental predicate logic schemas. Participants judged the model predicted inferences as being presented in the text in 58% to 63% of the times depending whether the texts contained only one schema (Experiment 1b) or several schemas (Experiment 2). These percentages are close to those of true paraphrases that were remembered as being presented in the text in 77% to 78% of the times. On the other hand, there is a significant step in difficulty between the schemas of the mental logic and other valid schemas of formal logic. Inferences of formal logic that do not take part of the mental logic model, were judged as being presented in the text only in 14% of the times for single schema texts and in 17% of the cases in multiple schema texts.

Experiment 3 provided evidence that one of the predicate mental logic schemas is drawn on-line during reading independently of any demand for coherence. The minimalist hypothesis of McKoon and Ratcliff (1992) cannot predict the logical inferences found by this experiment. The constructionist theory of Graesser et al. (1994) seems to be more adequate, although even this theory is rather skeptical in relation to forward logical inferences. A theory

that is capable to accommodate the logical inferences in the text is the construction-integration model of Kintsch (1998). Contradicting McKoon and Ratcliff (1986), Potts et al. (1988), Murray et al. (1993) and others, **t**he inferences applied in Experiment 3 were reliably drawn even though they could be characterized as purely elaborative.

Relative Difficulty of the Schemas

Besides these main findings, the experiments included in this study raised some additional issues that could enrich the theory of mental logic of Braine and O'Brien (1998). One of them is related to the relative difficulty of the core schemas of mental logic. Research of Yang et al. (1998) already showed that the predicate mental model core schemas are not equally easy. The authors relate the relative difficulty of the schema to the number and type of logical particles it uses. The present study suggests another possible explanation for the differences in relative difficulty of the schemas when used in narrative texts: The difficulty of the schema could be related to the number of premises the schema requires. When the schema requires a bigger number of premises, the arguments necessary for the conclusion are repetitively mentioned in the text. Readers draw more often inferences that are highly activated from multiple information sources (Graesser et al., 1994). When a logical schema requires only one premise, the probability of drawing an inference would be somewhat smaller than when the topic would be activated several times due to a higher number of premises. Other factors, such as the presence of negations, or type of logical particle, can also play a role in schema difficulty. This issue could be clarified in future research.

Organization of the Premises

Experiment 2 suggested that in texts containing premises for applying multiple schemas, the organization and order of presentation of the premises in the text could influence the probability of correct conclusion. This effect is related to the role of working memory in inference drawing and text processing. For an inference to be made, the premises have to be simultaneously present in the working memory. It has been suggested that when premises relevant to several mental logic schemas are not well organized in the text, the working memory has difficulties to keep track and maintain them active during long stretches of intermediate text. Such working memory overload could cause that some premises are dropped out, which would interrupt the correct functioning of the Direct Reasoning Routine. Nevertheless, questions like how does the reader organize the premises, or how does he/she decide which premises to apply in which schema, have to be investigated in future research.

Invited Inferences

The frequency of logical inferences was influenced by the trustworthiness of the premises, as has been pointed out by O'Brien (1993), Evans (2002) and others. What factors can influence the trustworthiness of the premises? It was showed that when premises cannot be doubted, as, for example, when logical problems are presented in abstract contexts, the mental logic schemas are reliably applied (e.g., Braine et al., 1984; Lea et al., 1995; Yang et al., 1998). The trustworthiness of premises becomes a factor when logical schemas are embedded in narratives. During text comprehension, readers create representations of the text, that contain not only information from the text, but also a number of inferences based on the reader's knowledge. These inferences are considered together with the premises of the logical schema, and the logical inference is drawn from the whole set of such premises. The exact number and type of these invited premises is impossible to determine, as they depend on individual knowledge and experience of the reader. Nevertheless, the present experiments detected some relevant types of invited premises that influence the probability that the mental logic schema will be applied.

Inferences related to pragmatic issues of the discourse and to the principles of conversational implicatures sometimes seem to add such invited premises and influence the probability of drawing the mental logic inference. The models of Cooren and Sanders (2002) and Grice (1975) could provide a useful framework for explaining such invited inferences.

When premises were presented as information shared by the characters of the narrative, the conclusion had to be presented as conclusions of the characters, too. Inferences about projected knowledge and projected co-presence, as suggested by Lea et al. (1998) and Gerrig et al. (2001) have to be taken in account while evaluating logical inferences in texts.

Logic and Language

We use natural language to describe our thoughts. Logic uses an abstract language to describe cognition. Therefore, "…to squeeze all human cognition into a logical formalism greatly compounds the distortion problem…:Compared with narrative language, logic…is a very inflexible system, not notably suited for the representation of natural language, and even less for lower levels of mental representation" (W. Kintsch, 1998, p.33). All three experiments of this study address the interplay between logic and text comprehension.

The experimental texts had to contain premises that would trigger a logical inference. For example, in the story *Experimental Drug* there is a premise saying "*there are no patients that have*

dengue and receive the drug". In the natural language, this sentence is almost equal to "*the dengue patients do not receive the drug"*, or even "*the drug is not given to the dengue patients"*. The exact surface structure as well as the propositional predicate-argument representation of these three paraphrases differs; however, the situation model constructed from these three sentences could be the same. Moreover, each of the above mentioned utterances could serve as a premise for another schema. The first premise would trigger an inference according to schema 3b: "*the patients that have dengue do not receive the drug"*. The second version of the sentence could be premise for schema 2a and together with the premise that "*Angela is a dengue patient*" lead to the conclusion that "*Angela does not receive the drug"*.

Kintsch (1998) proposes that text is represented on three levels: surface structure, propositional representation, and situational model. At which of these levels do the logical premises trigger the schema? Should we assume that several sentences can have a different surface structure yet they can lead to a common situational model, and, at the same time, they lead to different logical inferences, then it could be deduced that logical inferences do not occur on the situation model level, but only on the higher levels of text representation, such as surface structure or semantic/propositional representation.

The mental models theory of Johnson-Laird (1983) claims that people draw logical inferences from models that are iconic representations of the premises. Should we assume that the situational model of the text is of such a mental model sort, then the above presented argument shows that Johnson-Laird's proposition is not likely to be true. The mental logic inferences seem to be drawn from higher level representational models of text and not from the situational model.

Future research could address the issues of surface and deep structures of language and how different surface structure influences logical inferences. It would be interesting to analyze whether there would be, for example, a difference between premises expressed as "*there are no shirts that are blue and have short sleeves"*, or "*there are no blue shirts having short sleeves*" or even "*there are no blue shirts that have short sleeves"*.

Quantifiers

Some questions have been raised in relation to the representation and scope of quantifiers. Previous research tackled the difference between the universal quantifiers *all – each – every* (Brooks & Braine, 1996; Ioup, 1975). The present study pointed out some more topics to be investigated in the future. For example, in some of the experimental stories the universal

quantifier is not explicitly mentioned. This is because in natural language the use of quantifiers is not always obligatory. We could mention "*the green apples are expensive*" and mean "*all the green apples are expensive*" – the quantifier could be skipped. Why does mental logic apply the appropriate schema when a quantifier is missing? The explanation might be related to the fact that sometimes we consider *the green apples* as one single entity (as compared, for example, with *the red apples*) and other times we need to see it as a multiple entity (for example, a set of green apples where some are expensive and others might not be).

Another interesting effect is related to the existential quantifier *some*. In formal logic, *all* is a special case of *some*. For reasons related to Gricean implicatures, in natural language *some* would normally exclude *all*. According to Grice (1975), should one know that, for example, "*all the paintings were sold*", he/she would not say "*some of the paintings were sold*", because any utterance should be maximally informative and the speaker should not hide any relevant information. In the experimental stories of this study, the quantifier *some* was hardly used. The premises almost always referred to specific sub-sets. For example, *the yellow dress* was a subset of *all the dresses*, or *the speakers from Manaus* were a subset of *all the speakers*, and *tonight* was a subset of *always*. In the last example, the text-base of the narrative did not provide the information that, for example, *tonight* is a subset of *always*, the reader had to infer this from his own knowledge. Could this have increased the difficulty of the logical inference?

Some researchers have started to investigate quantifiers that are not expressible in standard predicate logic, and found out that people use, for example, the quantifier *most* in a similar way as *all*, *few* in a similar way as *some*, (Oaksford & Chater, 2001), or *at least* in a similar way as *some* (Geurts, 2003). Geurts (2003) complains that deductive reasoning theories are not dealing with quantified statements used in natural language, such as, *most A are B* or *three A are B*. These authors criticize logic-based approaches to deduction (mental logic theories) as unsatisfactory psychological models as they "are incapable of capturing even the simplest non-standard quantifiers" (Geurts, p.233). This accusation is unjust. The present study proved that mental logic theory is not this inflexible. The texts used in the experiments showed that, for example, readers had no problems interpreting quantifiers in the premises *all the Germans* and *my two German friend*s. The fact that the mental predicate logic schemas are defined in Braine's (1998) theory with the help of the standard four quantifiers (*all, some, some…not, none*) does not mean that the theory predicts that only those quantifiers would be used correctly in everyday reasoning. The mental logic proposal is open to comprise many of the non-standard quantifiers used in everyday communication. Braine (1998) does include in his theory quantifiers such as

many, *few*, or specifications such as *Sam and Harry* as sub-set of *the boys*. Nevertheless, future research could provide more insight in the process that allows to switch between different types of quantifiers, treating, for example the cardinal quantifier *two of them* as a specific case of the particular quantifier *some*.

Logical Particles

Not only quantifiers but also logical particles are sometimes used in natural language in a different way than what is prescribed by formal logic. For example, *or* can be used both inclusively and exclusively in natural language, with context providing the resolution. When one of the narratives of Experiment 2 mentioned "*the women on the lunch were from Casa Forte or Boa Viagem*" the readers might have understood the particle *or* inclusively or even in the same sense as the particle *and:* There were some women from Casa Forte and others from Boa Viagem. In this case, so we could also say "*the women on the lunch were from Casa Forte and Boa Viagem*". Another possible interpretation of the *or* in this sentence would be that the author wanted to express doubts over the information: He/she was not sure whether the women were from Casa Forte or from Boa Viagem.

In natural language we do not use the particle *or* when we are confident about the truth or falsity of one of the members of the disjunction. For example, should we be sure that the women were from Casa Forte and not from Boa Viagem, we would not say that they are from Casa Forte *or* Boa Viagem, even though this conclusion is valid in formal logic. For natural language there has to be a certain level of probability of both options for the speaker to use a disjunction. As Grice (1975) explains, the speaker would not mention Boa Viagem if this would not be at least probable.

Mental logic theory found a solution to such ambivalence in the use of logical particles. By defining exactly the logical schemas where a certain logical particle is used, the theory redefines the meaning of such particles. For example, the use of *if* or *or* in mental logic is not defined by the four lines of the truth tables of formal logic, but instead by the few specific schemas in which they are needed in everyday reasoning.

Mental Logic and Other Theories of Deductive Reasoning

Several researchers in the field of deductive reasoning claim that people do not posses domain-general logical competence. Instead, correct logical deductions are presumably tied

with a small set of content specific pragmatic schemas. In the present study, the predicate mental logic schemas were tested in numerous short narratives. Each narrative introduced a different content and context. Contexts related to pragmatic schemas of permission or obligation as defined by Cheng and Holyoak (1985), or cheater detection (Gigerenzer & Hug, 1992), were avoided. Therefore, it can be reliable confirmed that the mental predicate logic schemas are domain-general, not related to any content-specific schemas.

The present study gave support to the mental logic theory by confirming its predictions in the area of text comprehension. Mental logic theory continues to prove that it is a useful description of human reasoning. In spite of all the empirical evidence by which the theory is continuously supported, the present deduction reasoning research gives a lot of attention to another theory of deduction - the mental models theory (e.g., Johnson-Laird, 2001). This victory seems to be unjust, as many authors refute mental logic theory because of misinterpretation or lack of information on the basic claims of the theory. For example, Meiser, Klauer, and Naumer (2001) investigated the role of working memory in propositional reasoning. Meiser and his colleagues erroneously assumed that mental logic reasoning is based on controlled and analytic reasoning strategies. Moreover, they based their experiment on "training" the subjects in mental logic rules. Mental logic rules are not susceptible to training as they are automatic, implicit rules acquired early in childhood by the process of language acquisition. Also, the rules Meiser et al. used in the training and experimental sessions included inferences that do not make part of mental logic, like the distinction between exclusive and inclusive *or*, or equivalence (*if and only if*).

Another example of misinterpretation of mental logic theory can be found in the work of Rader and Sloutsky (2001), who believe that "…syntactic (= mental logic) theories do not predict, a priori, that inference schemas for some logical forms are more available in memory than those for others" (p.847). Mental logic theory defined a set of primary schemas that are readily available in reasoning processes. Other schemas of formal logic are substantially more difficult than the ones detected by mental logic theory, as confirmed once more by the present study. Moreover, within the schemas of mental logic, differences in relative difficulty have been found for both mental propositional logic (Braine et al., 1984), as well as mental predicate logic (Yang et al., 1998).

In some cases the authors start from the assumption of validity of the mental models theory and make an arduous effort to explain the results of their research within this theory. When the results do not fit within the mental models theory, the researcher would suggest

adjustments of the mental models theory, instead of looking to the competing mental logic theory where their evidence could be comfortably accommodated (e.g., Klauer et al., 2000).

Conclusion

In sum, it can be concluded that mental logic theory describes well the basis of our everyday reasoning. The theory provides a basic repertory of inferential skills that are being enriched by pragmatic and domain-specific processes. People are very accurate at making the predicate mental logic inferences, and they make them easily enough that they often do not realize that they are making any inferences. These inferences are made on-line even when they are not required for coherence. Mental propositional logic used during reading is entrenched in the complex text comprehension processes. The present study made an attempt to integrate the two areas of cognition, reasoning and text processing, and offered some suggestions on how they can be are interrelated. Future research should address these issues in more depth on order to fully implement the mental logic theory in a discourse processing framework.

REFERENCES

Baddeley, A. D. (1986). *Working memory*. Oxford Psychology Series, Vol.11. Oxford: Oxford University Press.

Baddeley, A. D., Hitch, G. (1974). Working memory. In G. H. Bower (Ed.) *The psychology of learning and motivation,* Vol.8. pp.47-90. New York: Academic Press.

Barnet, V, & Lewis, T. (1994). *Outliers in statistical data*. 3rd ed. Chichester: John Wiley.

Bell, V., & Johnson-Laird, P. N. (1998). A model theory of modal reasoning. *Cognitive science*, 22, pp.25-51.

Bower, G. H. (1979). Experiments on story understanding and recall. *Quarterly journal of experimental psychology*, 28, pp.511-134.

Braine, M. D. S. (1998). Steps toward a mental-predicate logic. In M. D. S. Braine & D. P. O'Brien (Eds.), *Mental logic*, pp.273-332. Mahwah, NJ: Lawrence Erlbaum Associates.

Braine, M. D. S., & O'Brien, D. P. (1991). A theory of if: A lexical entry, reasoning program, and pragmatic principles. *Psychological review*, 98, pp.182-203.

Braine, M. D. S., & O'Brien, D. P. (1998a). *Mental logic*. Mahwah, NJ: Lawrence Erlbaum Associates.

Braine, M. D. S., & O'Brien (1998b). How to investigate mental logic and the syntax of thought. In M. D. S. Braine & D. P. O'Brien (Eds.), *Mental logic*, pp.45-61. Mahwah, NJ: Lawrence Erlbaum Associates.

Braine, M. D. S., & O'Brien (1998c). The theory of mental propositional logic: Description and illustration. In M. D. S. Braine & D. P. O'Brien (Eds.), *Mental logic*, pp.79-90. Mahwah, NJ: Lawrence Erlbaum Associates.

Braine, M. D. S., & Rumain, B. (1983). Logical reasoning. In J. H. Flavel & E. M. Markman (Eds.) *Handbook of child psychology*: Vol.III, Cognitive development. New York: Willey.

Braine, M. D. S., O'Brien, D. P., Noveck, I. A., Samuels, M., Fisch, S. M., Lea, R. B. & Yang, Y. (1995). Predicting intermediate and multiple conclusions in propositional logic inference problems: Further evidence for a mental logic. *Journal of experimental psychology: General*, 124, pp.263-292.

Braine, M. D. S., O'Brien, D. P., Noveck, I.A., Samuels, M.C., Lea, R.B., Fisch, S.M., & Yang, Y. (1998). Further Evidence For The Theory: Predicting Intermediate And Multiple Conclusions In Propositional Logic Inference Problems. In M. D. S. Braine & D. P.

O'Brien (Eds.), *Mental logic*, pp.145-197. Mahwah, NJ: Lawrence Erlbaum Associates.

Braine, M. D. S., Reiser, B .J., & Rumain, B. (1984) Some empirical justification for a theory of natural propositional logic, in G.H. Bower (Eds.) *The psychology of learning and motivation: Advances in research and theory*. Vol. 18. New York: Academic Press

Brainerd, C. J. (1977). On the validity of propositional logic as a model for adolescent intelligence. *Interchange*, 7, pp.40-45.

Brooks, P. J., & Braine, M. D. S. (1996). What children know about the universal quantifiers *all* and *each*? *Cognition*, 60, pp.235-268.

Calvo, M. G., & Castillo, M. D. (1996). Predictive inferences occur on-line, but with delay: Convergence of naming and reading times. *Discourse Processes*, 22, pp.57-78.

Carraher, T., Carraher, D., & Schlieman, A. L. (1989). *Na vida dez na escola zero*. São Paulo: Cortez.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. Cambridge: Cambridge University press.

Chater, N., & Oaksford, M. (1999). The probability heuristics model of syllogistic reasoning. *Cognitive psychology*, 38, pp.191-258.

Cheng, P. W. & Holyoak, K. J. (1985). Pragmatic reasoning schemas. *Cognitive psychology*, 17, pp.391-416.

Clark, H. H., & Chase, W. G. (1972). The process of comparing sentences against pictures. *Cognitive psychology*, 3. pp.472-517.

Cooren, F., Sanders, R. E. (2002). Implicatures: A schematic approach. *Journal of pragmatics*, 34, pp.1045–1067.

Cordeiro, A. A. A. (2003). *Resolução de problemas da lógica predicativa por surdos usuários da língua de sinais brasileira: um enfoque da teoria da lógica mental*. Unpublished doctoral thesis, UFPE, Recife, Brazil.

Cosmides, L. (1989). The logic of social exchange: Has natural selection shaped how humans reason? Studies with the Wason selection task. *Cognition*, 31, pp.187-275.

Cosmides, L., & Tooby, J. (2000). Consider the source: The evolution of adaptations for decoupling and metarepresentation. In D. Sperber (Ed.), *Metarepresentations* (pp.53-115). Oxford, England: Oxford University Press.

Ennis, R. H. (1975). Children's ability to handle Piaget's propositional logic. *Review of educational research*, 45, pp.1-41.

Ennis, R. H. (1976). An alternative to Piaget's conceptualization of logical competence. *Child development*, 47, pp.913-919.

Evans, J. St. B. T. (1972). Interpretation and "matching bias" in a reasoning task. *Quarterly journal of experimental psychology*, 24, pp.193-199.

Evans, J. St. B. T. (1982). *The psychology of deductive reasoning*. London: Routledge.

Evans, J. St. B. T. (1991). Theories of human reasoning: The fragmented state of the art. *Theory & Psychology*, Vol.1, pp.83-105.

Evans, J. St. B. T. (1993). Bias and rationality. In K. I. Mankeltow & D. E. Over (Eds.), *Rationality: Psychological and philosophical perspectives* (pp.6-30). London: Routledge.

Evans, J. St. B. T. (2002). Logic and human reasoning: An assessment of the deduction paradigm. *Psychological bulletin*, 128(6), pp.978–996.

Evans, J. St. B. T., Barston, J. L., & Pollard P. (1983). On the conflict between logic and belief in syllogistic reasoning. *Memory and cognition*, 11, pp.295-306.

Evans, J. St. B. T., Clibbens, J., & Rood, B. (1995). Bias in conditional inference: Implications for mental logic and mental models. *Quarterly journal of experimental psychology*, 48A, pp.644-670.

Evans, J. St. B. T., Handley, S. J., Harper, C., & Johnson-Laird, P. N. (1999). Reasoning about necessity and possibility: A test of the mental model theory of deduction. *Journal of experimental psychology: Learning, memory and cognition*, 25, pp.1495-1513

Falamagne, R. J., Gonsalves, J. (1995). Deductive inference. *Annual review of psychology*, 46, pp.525–559.

Fincher-Keifer, R. (1996). Encoding differences between bridging and predictive inferences. *Discourse Processes*, 22, 225-246.

Fisch, S. (1991). *Mental logic in children's reasoning and text comprehension*. Unpublished doctoral dissertation, New York University.

Fodor, J. (1975). *Representations*. Cambridge, MA: Harvard University Press.

Frensch, P. A., & Funke, J. (1995). *Complex problem solving: The European perspective*. Hilsdale, NL: Lawrence Erlbaum Associates.

Garnham, A. & Oakhil, J. (1991). *Thinking and reasoning*. Oxford: Blackwell.

Geis, M., & Zwicky, A. M. (1971). On invited inferences. *Linguistic inquiry*, 2, pp.561-566.

Gentzen, G. (1935). Untersuchugen uber das logische Schliessen. Matematische zeitschrift, 39,

pp.176-221. (II. Investigations into logical deduction. (1964), *American philosophical Quarterly*, 1, pp.288-306.

Gernsbacher, M. A., & Jescheniak, J. D. (1989). Cataphoric devices in spoken discourse. *Cognitive psychology*, 29, pp.24-58.

Gerrig, R. J., Brennan, S. E., Ohaeri, J. O. (2001). What characters know: Projected knowledge and projected co-presence. *Journal of memory and language*, 44, pp.81– 95.

Geurts, B. (2003). Reasoning with quantifiers. *Cognition*, 86, pp.223-251.

Gigerenzer , G., & Hug, K. (1992). Domain-specific reasoning: Social contracts, cheating and perspective change. *Cognition*, 43, pp.127-171.

Goel, V., Gold, B., Kapur, S., Houle, S. (1997). The seats of reason? An imaging study of deductive and inductive reasoning. *Neuroreport*, 8(5),pp.1305-10.

Graesser, A. C., Singer, M., Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological review, 101(3),* pp.371-393.

Grice, H. P. (1979). *Pragmatics: implicature, presupposition and logical form*. Academic Press.

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics III: Speech acts* (pp.41-58. New York: Academic Press.

Griggs, R. A., & Cox (1982). The elusive thematic-materials effect in Wason's selection task. *British journal of psychology*, 73, pp.407-420.

Henle, M. (1962). On the relation between logic and thinking. *Psychological review*, 69, pp.366-378.

Inhelder, B., & Piaget, J. (1958). *The growth of logical thinking.* New York: Basic Books.

Inhelder, B., & Piaget, J. (1964). *The early growth of logic in the child.* London: Routledge & Kegan Paul.

Ioup, G. (1975). Some universals for quantifier scope. In J. Kimball (Ed.), *Syntax and semantics (Vol.4)*. New York: Academic Press.

Johnson-Laird, P. N. (1980). Mental models in cognitive science. *Cognitive science*, 4, pp.71-115.

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

Johnson-Laird, P. N. (1999). Deductive reasoning. *Annual review of psychology*, 50, pp.109-135.

Johnson-Laird, P. N. (2001). Mental models and deduction. *Trends in cognitive sciences*, 5(10), pp.434–

442.

Johnson-Laird, P. N., & Bara, B. G. (1984). Syllogistic inference. *Cognition*, 16, pp.1-61.

Johnson-Laird, P. N., & Byrne (1991). *Deduction*. Hove, England: Erlbaum.

Johnson-Laird, P. N., Byrne, R. M. J. & Schaeken, W. (1992). Propositional reasoning by model. *Psychological review*, 99, pp.418-439.

Just, M. A., Carpenter, P. A. (1971). Comprehension of negation with quantification. *Journal of verbal memory and verbal behavior*, 10, pp.244-253.

Just, M. A., Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, 99, pp.122-149.

Kahneman, D., & Tversky, A.. (1972). Subjective probability: A judgment of representative-ness. *Cognitive psychology*, 3, pp.430-454.

Kahneman, D., & Tversky, A.. (1982). On the study of statistical intuition. *Cognition*, 12, pp.325-336.

Kearns, J. T. (1988). *The principles of deductive logic*. Albany, N.Y.: State university of New York Press.

Keefe, D. E., & McDaniel, M. A. (1993). The time course and durability of predictive inferences. *Journal of Memory and Language*, 32, pp.446-463.

Keenan, J. N., Golding, J. M., Potts, G. R., Jennings, T. M., Aman, Ch. J. (1990). Methodological issues in evaluating the occurrence of inferences. *The psychology of learning and motivation*, Vol.25., pp.295-312. New York: Academic Press.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction – integration model. *Psychological review*, 95(2), pp.163–182.

Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge: Cambridge University Press.

Kintsch, W. & Keenan, J. M. (1973). Reading rate and retention as a function of the number of propositions in the base structure of sentences. *Cognitive psychology*, 5, pp.257-279.

Kintsch, W., van Dijk, T.A. (1978). Towards a model of text comprehension and production. *Psychological review*, 85, pp.363-394.

Kintsch, W., Welsch, D., Schmalhofer, F., & Zimny, S. (1990). Sentence memory: A theoretical analysis. *Journal of memory and language*, 29, pp.133-159.

Klauer, K. C., Musch, J., Naumer, B. (2000). On belief bias in syllogistic reasoning. *Psychological*

review, 107(4), pp.852– 884.

Klin, C. M. (1995). Causal inferences in reading: From immediate activation to long-term memory. *Journal of experimental psychology: Learning, memory and cognition*, 32, pp.1483-1494.

Klin, C. M., Guzman, A. E., & Levine, W. H. (1999). Prevalence and persistence of predictive inferences. *Journal of Memory and Language*, 40, 593-604.

Klin, C. M., Murray, J. D., Levine, W. H., & Guzman, A. E. (1999). Forward inferences: From activation to long-term memory. *Discourse processes*, 27(3), pp.241-260.

Lea R. B., (1995). On-line evidence for elaborative logical inferences in text. *Journal of experimental psychology*: LMC, 6, pp.1469-1482.

Lea, R. B. (1998). Logical inferences in comprehension: How mental logic and text processing need each other. In M. D. S. Braine & D. P. O'Brien (Eds.), *Mental logic*, pp.233-255. Mahwah, NJ: Lawrence Erlbaum Associates.

Lea, R. B., Mason, R. A., Albrecht, J. E., Birch, S. L., Myers, J. L. (1998). Who knows what about whom: What role does common ground play in accessing distant information? *Journal of memory and language*, 39, pp.70–84.

Lea, R. B., Mulligan, E. J. (2002). The effect of negation on deductive inferences. Journal of experimental psychology: *Learning, memory, and cognition*, 28(2), pp.303-317

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

Markman, A. B., Gentner, D. (2001). Thinking. *Annual review of psychology*, 52, pp.223–247.

McDonald, M. C., & Just, M. A. (1989). Changes in activation levels with negation. *Journal of experimental psychology: learning, memory and cognition*, 15, pp.633-642.

McKoon, G., Ratcliff, R. (1992). Inference during reading. *Psychological review*, 99(3), pp.440-466.

Meiser,T., Klauer, K. Ch., Naumer, B. (2001). Propositional reasoning and working memory: The role of prior training and pragmatic content. *Acta Psychologica*, 106(3), pp. 303-327.

Millis K. K., & Just, M. A. (1994). The influence of connectives on sentence comprehension. *Journal of Memory and Language*, 33, pp.128 – 147.

Murray, J. D. (1997). Connectives and narrative text: The role of continuity. *Memory and Cognition*, 25(2), pp.227–236.

Murray, J. D., Klin, C. M., Myers, J. L. (1993). Forward inferences in narrative text. *Journal of memory*

and language, 32, pp.464–473.

Noveck, I. A., Lea, R. B., Davidson, G. M., & O´Brien, D. P. (1991). Human reasoning is both logical and pragmatic. *Intellectica*, 11, pp.81-109.

Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice-Hall.

Newstead, S. E., Handley, S. H., & Buck, E. (1999). Falsifying mental models: Testing the predictions of syllogistic reasoning. *Journal of memory and language*, 27, pp.344-354.

Oaksford, M. & Chater, N. (2001). The probabilistic approach to human reasoning. *Trends in cognitive sciences*, 5, pp.349-357.

O'Brien, D. P. (1987). The development of conditional reasoning: An iffy proposition. *Advances in child development and behavior*, 20, pp.61–90.

O'Brien, D. P. (1993). Mental logic and irrationality: We can put a man on the moon, so why can't we solve those logical reasoning problems? In K. I. Manktelow & D. E. Over (Eds.). *Rationality: Psychological and philosophical perspectives*, pp. 110-135.

O'Brien, D. P., Braine, M.D.S., & Yang, Y. (1994). Propositional reasoning by mental models? Easy to refute in principle and in practice. *Psychological Review*, *101*, 711-724.

O'Brien, D. P., & Lee, H-W. (1992). A cross-linguistic investigation of a model of mental logic: The same inferences are made in English and Chinese. Unpublished manuscript.

Osherson, D. N. (1975). Models of logical thinking. In R. Falamagne (Ed.) *Reasoning: Representation process in children and adults*. Hilsdale, NL: Erlbaum.

Osherson, D., Perani, D., Cappa, S., Schnur, T., Grassi, F., Fazio, F. (1998). Distinct brain loci in deductive versus probabilistic reasoning. *Neuropsychologia*, 36(4), pp.369-76.

Politzer, G. & Braine, M. D. S. (1991). Responses to inconsistent premises cannot count as suppression of valid inferences. *Cognition*, 38, pp.103-108.

Potts, G. R., Keenan, J. M., & Golding, J. M. (1988). Assessing the occurrence of elaborative inference: Lexical decisions versus naming. *Journal of memory and language*, 27, pp.399-415.

Rader, A., Sloutsky, V. (2001). Conjunctive bias in memory representations of logical connectives. *Memory and Cognition*, 29(6), pp.838–849.

Reder, L. M. (1982). Plausibility judgments versus fact retrieval: Alternative strategies for sentence verification. *Psychological review*, 89, pp.250-280.

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

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

Rumain, B., Connell, J. W., & Braine, M. D. S. (1983). Conversational comprehension processes are responsible for reasoning fallacies in children as well as adults: *If* is not the biconditional. *Developmental psychology*, 19, pp.471-481.

Schaeken, W. S. (1996). Mental models and temporal reasoning. *Cognition*, 60, pp.205-234.

Schaeken, W., & Schroyens, W. (2000). The effect of explicit negatives and different contrast classes on conditional reasoning. *British journal of psychology*, 91, pp.533-550.

Schank, R. C. (1979). Interestingness: Controlling Inferences. *Artificial intelligence*, 3, pp. 273-297.

Siegel, S. (1956). *Nonparametric statistics for the behavioral sciences*. NY: McGraw-Hill.

Singer, M. (1990). Inference processes. In M. Singer: *Psychology of language: An introduction to sentence and discourse processes*, pp. 167-189. Hillsdale, NJ: Erlbaum.

Singer, M, & Ferreira, F. (1983). Inferring consequences in story comprehension. *Journal of Verbal Learning and Verbal Behavior*, 22, pp.437-448.

Singer, M., Kintsch, W. (2001). Text retrieval: A theoretical exploration. *Discourse processes*, 31(1), pp.27–59.

Sperber, D., Wilson, D. (1995). *Relevance: communication and cognition*. Blackwell Publishers, Cambridge, MA.

Stanovich, K. E. (1999). *Who is rational? Studies of individual differences in reasoning*. Mahwah, NJ: Erlbaum.

Sternberg, R. J. (1994). *Thinking and problem solving*. London: Academic press.

Suchmann, L. A (1987). *Plans and situated actions: The problem of human-machine communication.* Cambridge, UK: Cambridge University Press.

Swinney, D. A. (1979). Lexical access during sentence comprehension: (re)consideration of context effects. *Journal of verbal learning and verbal behavior*, 18, pp.645-659.

Tabossi, P. (1999). Mental models in deductive, modal, and probabilistic reasoning. In C. Habel and G. Rickheit (Eds.), *Mental models in discourse processing and reasoning*, pp.229-331, Elsevier.

Thorndike, E. L. (1977). Reading as reasoning: study of mistakes in paragraph reading. *Journal of education psychology*, 8, pp.323-332.

Tooby, J. & Cosmides, L. (1989). Evolutionary psychology and the generation of culture: I. Theoretical considerations. *Ethological sociobiology*, 10(1-3), pp.29-49.

Trabasso, T., & Sperry, L. L. (1985). Causal relatedness and importance of story events. *Journal of Memory and Language*, 24, pp.595-611.

Tversky, A., & Kahneman, D. (1973). Availability: A heuristics for judging frequency and probability. *Cognitive psychology*, 5, pp.207-232.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185, pp.1124-1131.

Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy improbability judgment. *Psychological review*, 90, pp.293-315. probability. *Cognitive psychology*, 5, pp.207-232.

Uleman, J. S., Hon, A., Roman, R. J., & Moskowitz, G. B. (1996). On-line evidence for spontaneous trait inferences at encoding. *Personality and social psychology bulletin*, 22(4), pp.377-394.

Walker, C. H., & Meyer, B. J. (1980). Integrating different types of information in text. *Journal of verbal learning and verbal behavior*, 19(3), pp.263-275

Walker, C. H., & Yekovich, F., R. (1984). Script based inferences: Effects of text and knowledge variables o recognition memory. *Journal of Verbal Learning and Verbal Behavior*, 2, pp.357 – 370.

Wason, P. C. (1983). Realism and rationality in the selection task. In St. B. T. Evans (Ed.), *Thinking and reasoning: Psychological approaches* (pp.44 –75). London: Routledge & Kegan Paul.

Wason, P. C. (1960). On the failure to eliminate hypothesis in a conceptual task. *Quarterly journal of experimental psychology*, 12, pp.129-140.

Wason, P. C. (1968).Reasoning about a rule. *Quarterly journal of experimental psychology*, 60, pp.273-281.

Wason, P. C. (1977).The theory of formal operations: a critique. In B. Geber (Ed.), *Piaget and knowing* (pp.119-135). London: Routledge & Kegan Paul.

Wason, P. C., & Shapiro, D. (1971). Natural and contrived experience in a reasoning problem. *Quarterly journal of experimental psychology*, 23, pp.63-71.

Whitney, P., Ritchie, B. G., & Crane, R. S. (1992). The effect of foregrounding in readers' use of predictive inferences. *Memory and Cognition*, 20, pp.424-432.

Yang, Y., Braine, M. D. S., O'Brien, D. P. (1998). Some empirical justification of the mental-predicate-logic model. In M.D.S.Braine & D.P.O'Brien (Eds.), *Mental logic*, pp.333-365. 35)

Mahwah, NJ: Lawrence Erlbaum Associates.

Zimny, S. T. (1987). *Recognition memory for sentences from a discourse*. Unpublished doctoral dissertation, University of Colorado, Boulder, CO.

APPENDIX A – MATERIALS USED IN EXPERIMENT 1A AND 1B AND EXAMPLE OF THE EXPERIMENTAL PROTOCOLS

Stories requiring application of Schema 1a

Notation of the schema:
S1[All X] OR S2[PRO-All X]; NEG S2[α]; α ⊇ [X] /∴ S1[α]
Example: The boys either played with girls or fought with girls; Tom and Dick did not play with girls/ Tom and Dick fought with girls.

1. Lunch

The boss invited some of the people who work for him to a barbeque. All the people who came to the barbeque were from the same department.

The boss was famous for providing always delicious food. This time, all the guests could choose either steak or grilled fish.

Márcia, the boss's secretary, also came to the barbeque. When the boss saw that she had finished eating, he was curious about what she had chosen.

"I eat steak yesterday for dinner," said Márcia, "so I didn't choose it again today."

Validity task items:

Valid: "Ah," the boss replied to Márcia, "so you had grilled fish, then."
Invalid: "Ah," the boss replied to Márcia, "so you had lamb chops, then."

Recognition items:

Model Predicted: Márcia chose grilled fish on the barbeque of the boss.
Paraphrase: Márcia told the boss that she didn't choose steak.
Foil: Anyone from another department had not come to the barbeque.

2. Legs

Roberta was all excited when she phoned her girlfriend, Veronica.

"You're not going to believe it, but Paul asked me to go to dance forró with him! Now I only have to decide what to wear," she said.

"Well", her friend, Veronica, agreed, "I know that you do not like to show your legs, so don't wear a mini skirt.

"Yes, I hate to show my chubby legs," Roberta confirmed and continued:

"I always wear either a long skirt or pants and this time I will not wear a long skirt, because it would annoy me during dancing."

Validity task items:

Valid: Roberta decided to wear pants to the disco with Paul.

Invalid: Roberta decided not to wear pants to the disco with Paul.

Recognition items:

Model Predicted: Roberta decided to wear pants to the disco with Paul.

Paraphrase: Roberta said she does not like to show her chubby legs.

Foil: Roberta always wears mini-skirt, long skirt or pants.

Stories requiring application of Schema 1b

Notation of the schema:

S1[All X] OR S2[PRO-All X]; /∴ S2[All X: NEG S1[PRO]]

Example: The boys either played with girls or fought with girls/ The boys who did not play with girls fought with girls.

3. Concert

The children who study in the Musical Club can choose to play either piano, or violin, or percussions.

Renata was helping to organize some music concerts to present the small musicians.

When she was writing the program she asked the music teacher:

"What instruments will the children be playing at the concert in October?"

The teacher answered, "At the October concert all the children will be playing either piano or violin."

Validity task items:

Valid: Renata replied, "So, at the October concert, the children who don't play violin will play piano.

Invalid: Renata replied, "So, at the October concert, the children who don't play violin also will not play piano."

Recognition items:

Model Predicted: At the October concert, the children who don't play violin will play piano.

Paraphrase: The children studying at the Musical Club can choose between several instruments.

Foil: Any children who play the percussions will not play at the October concert.

4. Conference

Jaime was organizing a conference at the university.

The first day everything went all right an all the participants were a great success.

The second day he was becoming worried that the participants from Manaus who were supposed to speak tomorrow morning had not yet arrived.

In the afternoon his secretary found out that all of them were scheduled to arrive either on a 10 p.m. flight or on an 11 p.m. flight.

As soon as Jaime got this information he calmed down.

Validity task items:

Valid: The participants from Manaus who were not scheduled to arrive on the 10pm flight were scheduled to arrive on the 11 p.m. flight.

Invalid: The participants from Manaus who were not scheduled to arrive at 10 p.m. were not scheduled to arrive at 11 p.m.

Recognition items:

Model Predicted: The participants who were not scheduled to arrive on the 10pm flight were scheduled to arrive on the 11 p.m. flight.

Paraphrase: Jaime was not nervous anymore after he got the information about the flights.

Foil: A participant who was not successful did not speak on the first day.

Stories requiring application of Schema 2a

Notation of the schema:
S[All X]; $\alpha \supseteq$ [X]/$\therefore$ S[α]
Example: The girls all wore red jeans / The girls in sneakers wore red jeans.

5. Friends

Mario started a course at a college in Olinda.

A week later his cousin came to pick him up after class, and asked:

"Have you already made some friends?"

"In the beginning I didn't know anybody, but now at least I've met all the five women in tonight's class. All are very nice and every one of them lives in Casa Forte, like me," answered Mario.

As they were talking, an attractive woman approached them.

Mario smiled at her and said to his cousin: "Well, I am just going to introduce you to one of them."

Validity task items:

Valid: "You live in Casa Forte, then," said José.

Invalid: "You live in Olinda then," said José.

Recognition items:

Model Predicted: The colleague Mario was about to present to his cousin lives in Casa Forte.

Paraphrase: Mario did not know anybody in his class in the beginning.

Foil: All the women from the course are either from Casa Forte or from Olinda.

6. Sandwich

Roberto and Cátia had been on a holiday trip and were at the airport waiting for their flight, almost without any more money to spend.

As they entered the terminal, Roberto noticed only one place that sold snacks and soft drinks.

Cátia mentioned that she was very hungry and that she wanted a sandwich.

A large sign there stated that all of the sandwiches cost5 Reais.

Cátia asked Roberto to buy a cheese sandwich.

Validity task items:

Invalid: "Well," said Roberto, "a cheese sandwich from the food stand does not cost 5 Reais."

Valid: "Well, "Roberto complained, "a cheese sandwich from the food stand costs 5 Reais."

Recognition items:

Model Predicted: A cheese sandwich from the food stand on the airport costs 5 Reais.

Paraphrase: Cátia wanted to have a ham-and-cheese sandwich because she was hungry.

Foil: Items at the food stand that cost less than 5 Reais did not include any sandwiches.

Stories requiring application of Schema 2b

Notation of the schema:

NEG S[~Some X~]; $\alpha \supseteq$ [X]/$\therefore$ NEG S[α]

Example: None of the boys wore striped shirts / Sam and Henry did not wear striped shirts.

7. Fruit

Manuel was a Spanish tourist who was visiting Recife. He bought a book about plants and animals in Brazil and was reading it in his hotel room.

"The trees and bushes that are natural to Mata Atlantica have only small fruits, like cajú, goiaba, and pitanga...

…Nevertheless, during colonization, lots of plants, like coconuts and mangoes, were imported from Africa and Asia and planted here.

None of the large fruits are originally from Mata Atlantica."

Manuel stopped and started to think about the information he just read.

"For example, jaca, my favorite fruit, is one of these large fruits", he remembered.

Validity task items:

Valid: "So jaca can not be originally from the Mata Atlantica."

Invalid: "So jaca is originally from the Mata Atlantica."

Recognition items:

Model Predicted: Manuel's favorite fruit, the jaca, is not originally from the Mata Atlantica.

Paraphrase: Manuel stopped to think over the information about the plants in the Mata Atlantica.

Foil: The large fruits in Brazil are imported or they came from a region outside Mata Atlantica.

8. Restaurant

Silvano and his girlfriend heard that all their friends have already eaten in a recently opened restaurant and decided to try this place, too.

The girlfriend had stopped eating red meat one year ago, so they immediately started to discuss which dishes would include red meat and which fish.

They called the waiter to clear this up and he answered, "This restaurant serves only sea food, so none of the dishes on the menu contain meat."

Sarah asked Silvano whether he was interested in trying the dish called 'Specialty of the Chef'.

Validity task items:

Valid: "Well, the Specialty of the Chef does not contain red meat," said Silvano.

Invalid: "Well the Specialty of the Chef does contain red meat," said Silvano.

Recognition items:

Model predicted: The Specialty of the Chef does not contain red meat.

Paraphrase: The waiter explained that the restaurant is specialized in seafood.

Foil: All their friends have already eaten in that restaurant or they have tried 'Specialty of the Chef .

Stories requiring application of Schema 3a

Notation of the schema:

NEG E[~Some X: S1[PRO-All X] & S2[Pro]~]; S2[α]; $\alpha \supseteq$ [X]/$\therefore$ NEG S1[α]

Example: There were no boys who wore sandals and blue jeans; The boys that played with Mary wore blue jeans / The boys that played with Mary did not wear sandals.

9.Guide

A hotel owner received an e-mail message telling her that a group of Japanese tourists was about to arrive at the same time as a group from Korea.

She realized that all the guests coming next week would want Asian food

But a bigger problem was that she urgently needed a guide for them.

Nobody on her staff could speak both Japanese and Korean.

She thought: "I think we will have to hire an interpreter."

It’s a pity because, for example, Isabel speaks Japanese very well...”

Validity task items:

Valid: “But Isabel doesn’t speak Korean,” she commented. .

Invalid: “And Isabel speaks Korean,” she commented.

Recognition items:

Model Predicted: The hotel owner remembered Isabel who speaks Japanese but doesn’t speak Korean.

Paraphrase: The owner discovered that two groups of tourists from Japan and Korea would arrive together.

Foil: Guests who don’t want Asian food would not be among those coming next week.

10. School

Some women from the Casa Forte neighborhood were speaking about their children when they were having lunch.

Angela mentioned that she wanted to her oldest son go to a school that would give the curriculum both in English and in Portuguese so that he would learn to speak English properly, and complained:

“But there is no school in Recife that is cheap enough for us to afford and gives the curriculum in English.

“The American School of Recife in Boa Viajem is one of the schools I am speaking about,” Angela continued, “because it, in fact, does give the curriculum in English.”

Validity task items:

Valid: “and the American School of Recife isn’t cheap enough for us to afford.”

Invalid: “and the American School of Recife is cheap enough for us to afford.”

Recognition items:

Model predicted: The American School of Recife isn’t cheap enough for Angela to afford.

Paraphrase: Angela wanted her son to learn English properly.

Foil: The women were from Casa Forte ou Boa Viagem.

Stories requiring application of Schema 3b

Notation of the schema:

NEG (S1[All X] & S2 [PRO-All X])/∴ NEG S2[All X: S1[PRO]]

Example: The boys did not wear sandals with blue jeans/ The boys that wore blue jeans did not wear sandals.

11. Stealing

The manager of a company entered his personnel director's office and announced, "We've had some problems with theft in the warehouse, and the police think it's an inside job," he told the personnel director.

He continued, "I just want to check that what I told the police was right. I said that our employees do not work in the warehouse having a criminal record."

"Sir, you can stay calm, that is true. I know about it and I checked today the documentation of each of the workers in the warehouse", the personnel director answered.

Validity task items:

Valid: The workers who are in the warehouse do not have any criminal record.

Invalid: The workers who are in the warehouse have criminal records.

Recognition items:

Model Predicted: In that company, the workers who are in the warehouse do not have any criminal record.

Paraphrase: The manager said that there had been some cases of stealing in the warehouse.

Foil: Any worker not checked by the director today works in another part of the company.

10. Mini-skirts

André and his mom were walking past an evangelic church on their block.

People coming out from a service were mixing with the others on the sidewalk.

"Look, Mum, do you think those two girls are also evangelic?" asked André.

He was speaking about two girls in miniskirts on the sidewalk he never had seen there before.

"André, you know that evangelic girls use long skirts. There are no girls who are evangelic who wear a mini-skirt," Helena answered.

Validity task items:

Valid: André understood that girls who wear mini-skirts are not evangelical.

Invalid: André understood that some girls who wear mini-skirts are evangelical.

Recognition items:

Model Predicted: Mum explained that girls who wear mini-skirts are not evangelical.

Paraphrase: André and his mum saw some people in front of an evangelic church.

Foil. Non-evangelic girls wear mini-skirts or long skirts.

Stories requiring application of Schema 4

Notation of the schema:
S1[All X] OR S2[PRO-All X]; S3[All X: S1[PRO]]; S3[All X: S2[PRO]] /∴ S3[All X]
Example: All the cars in the lot have stickers or the guards tow them away; The cars that have stickers are Toyotas; The cars that the guards toy away are Toyotas/ All the cars in the lot are Toyotas.

13. Chairs

Luciano was buying furniture for his new house because all the furniture from his old apartment was very worn out.

He needed some nice chairs, but his financial condition was not good.

He knew he couldn't spend too much on furniture.

Luciano looked several times through the catalogue for the Super Furniture Store, and complained:

"This is crazy! All the chairs from the Super Furniture Store are either ugly or they are extremely expensive."

"Well, I will certainly not buy ones that are ugly and I will certainly not buy ones that are extremely expensive," concluded Luciano.

Validity task items:

Valid: "So, I won't buy chairs at the Super Furniture Store."

Invalid: "So, I will buy chairs at the Super Furniture Store."

Recognition items:

Model Predicted: After looking through the catalog, Luciano concluded that he will not buy chairs at the Super Furniture Store.

Paraphrase: Luciano thought that he couldn't spend much money on chairs for his new apartment.

Foil: If a piece of furniture was in good condition it was not from Luciano's old apartment.

14. Exam

Stefano, a student of psychology, had just finished an important exam.
Completely tired, he arrived at the café to meet with his classmates.
"So, Stefano, was it difficult?" asked the friends.
"Well, all the questions were about memory or about perception," answered Stefano.
"Did you know all the answers?" inquired the friends.
"I answered all the questions about perception," Stefano explained.
"And what about the questions about memory?" the classmates asked.
"I answered all the questions about memory, too", Stefano replied with relief.

Validity task items:

Valid: so I'm sure that I answered all the questions correctly."

Invalid: so I might not have answered all the questions correctly."

Recognition items:

Paraphrase: In the café Stefano's classmates asked him if he knew all the answers.

Model predicted: Stefano told his friends that he answered all the questions of the exam.

Foil: If there was a question about perception Stefano didn't answer, then it was on another exam.

Stories requiring application of Schema 5

Notation of the schema:
S1[All X] OR S2[PRO-All X]; S3[All X: S1[PRO]]; S4[All X: S2[PRO]]
/∴ S3[All X] OR S4[PRO-All X]
Example: All the cars in the lot have stickers or the guards tow them away; The cars that have stickers are Datsuns; The cars that the guards toy away are Toyotas/ The cars in the lot are all Datsuns or Toyotas.

15. Dance

My neighbors are a young couple who like to dance very much.

On one evening during the San João festival week they decided to go to their favorite bar.

On the way to the bar, the wife, who usually likes to dance to samba, asked her husband what type of music the bands would be playing that night.

"All of the bands that will be playing at the bar during Saõ João are either from Recife or from Caruaru," he replied.

"I spoke with the owner of the bar and he explained to me that the bands from Recife all play forró, and the bands from Caruaru all play xote," explained the husband.

Validity task items:

Valid: "The bands will be playing either forró or xote," said the wife.

Invalid: "Some bands will be samba tonight," side the wife.

Recognition items:

Model predicted: During São João the bands at the bar will be playing either forró or xote.

Paraphrase: During São João the couple went to dance at their favorite bar.

Foil: The wife likes to dance samba or forró.

16. Exhibition

A student of journalism had to write an article about an upcoming exhibition at the university Art Club.

She was wondering whether there would be any nudes on the exhibition.

She discovered that she could check all the paintings going to the exhibition during

her visit.

She also noticed that all of the painters were working either outdoors or in the studio and asked the art teacher why the painters were not all working in one place.

The teacher answered that, “Those working in the studio are painting portraits. We asked some models to pose for us.

And those working outdoors are painting landscapes.

Validity task items:

Valid: The painters preparing pictures for the exhibition are painting portraits or landscapes.

Invalid: Some painters were working on nudes.

Recognition items:

Model Predicted: The painters preparing pictures for the exhibition are painting portraits or landscapes.

Paraphrase: The journalism student was curious why the painters were not all in one place.

Foil: If the student could not check a painting during her visit, it will not be in the exhibition.

EXAMPLE OF THE EXPERIMENTAL PROTOCOL - EXPERIMENT 1A

Idade_____

Sexo_____

Leia as estórias e marque o final mais apropriado.

Restaurante

Sílvio e sua namorada souberam que todos os amigos deles já comeram num restaurante recentemente aberto e decidiram provar este lugar também.

O cardápio tinha um prato especial chamado 'Surpresa do Chefe'.

A namorada parou de comer carne vermelha há um ano, então, eles logo começaram a discutir se este prato seria de carne ou de peixe.

Chamaram o garçom e ele respondeu:

"Olhe, este restaurante serve somente frutos do mar e, portanto, nenhum dos pratos do cardápio tem carne."

Final 1: "Bem, a 'Surpresa do Chefe' não é um prato de carne", avisou Sílvio.

Final 2: "Bem, a 'Surpresa do Chefe' é um prato de carne", avisou Sílvio.

EXAMPLE OF THE EXPERIMENTAL PROTOCOL - EXPERIMENT 1B

Leia a estória a seguir.

Restaurante

Sílvio e sua namorada souberam que todos os amigos deles já comeram num restaurante recentemente aberto e decidiram provar este lugar também.

A namorada parou de comer carne vermelha há um ano, então, eles logo começaram a discutir quais pratos seriam de carne e quais de peixe.

Chamaram o garçom para esclarecer isso e ele respondeu:

"Olhe, este restaurante serve somente frutos do mar e, portanto, nenhum dos pratos do cardápio tem carne."

A namorada perguntou a Sílvio se ele estaria interessado em pedir o prato chamado 'Surpresa do Chefe'.

Após ter terminado de ler, por favor, vire a página para responder as três perguntas sobre esta estória.

Por gentileza, não voltar para traz após ter ido para a próxima página.

<u>NÃO</u> VOLTAR PARA A PÁGINA ANTERIOR!

Cada uma das três frases a seguir contém informações que foram apresentadas na estória que você acabou de ler (embora não com palavras idênticas) ou, então, não foram apresentadas na estória, mas podiam ser inferidas (deduzidas) da estória.

Para cada uma destas três frases, por favor, indicar se as informações foram apresentadas na estória, ou se foram inferidas, ou se você está em dúvida entre estas duas opções. Indique a sua resposta numa escala, fazendo um círculo em cima de um dos três pontos.

<u>NÃO DÊ A MESMA RESPOSTA A TODOS OS TRÊS ITENS (por exemplo, não indique que todos os três itens foram apresentados na estória)</u>

1. O prato 'Surpresa do Chefe' não é um prato de carne.

❑ Foi apresentado na estória (embora não com palavras idênticas) ❑ Estou em dúvida ❑ Não foi apresentado na estória (mas poderia ser inferido)

2. O garçom explicou que o restaurante é especializado em frutos de mar.

❑ Foi apresentado na estória (embora não com palavras idênticas) ❑ Estou em dúvida ❑ Não foi apresentado na estória (mas poderia ser inferido)

3. Os amigos do casal já provaram a 'Surpresa do Chefe' ou comeram naquele restaurante.

❑ Foi apresentado na estória (embora não com palavras idênticas) ❑ Estou em dúvida ❑ Não foi apresentado na estória (mas poderia ser inferido)

APPENDIX B – MATERIALS USED IN EXPERIMENT 2 AND EXAMPLE OF THE EXPERIMENTAL PROTOCOL

1. A Fancy Present

Maria forgot about her friend's birthday. She entered a shop with accessories that seemed to have reasonable prices to look for a present.

Maria thought about what kind of a present her friend would like and picked up some nice shoes and some hand bags.

"So all the presents I would buy in this shop are either shoes or bags," she thought.

"Hey, but all these shoes are leather imitation," she discovered.

She asked the about the handbags and the salesman confirmed that they were all from leather imitation, too.

Maria thought for a while.

She realized that there was no present from leather imitation that would satisfy her friend.

Validity task items:

Valid: Maria concluded: "My friend would like to get a present from this shop."

Invalid: Maria thought: "A leather imitation present from this shop would not satisfy my friend. I have to look somewhere else…"

Recognition items:

Paraphrase: Maria was looking in a shop to find a present for her friend.

Model predicted: All the presents Maria thought to buy in this shop were from leather imitation.

Foil: If Maria bought a leather handbag then it was from another shop.

2. Peter's Friends

Peter was worried because two of his friends went with a group of tourists on a trip to the Amazons, and he heard that two tourists from that group got lost.

He phoned to the lodge and spoke with the receptionist. But the receptionist did not know the names of the two lost tourists.

"Can you check what was the nationality of the lost tourists? My two friends are German," asked Peter.

"On the day the tourists got lost we only had American and German guests in the hotel. And all the German tourists took a canoe trip", she told him.

"So?" Peter asked impatiently.

"Well, sir, it is confirmed that there were no tourists that took the canoe trip and got lost", the receptionist remembered.

Validity task items:

Valid: "That means that my friends did not get lost", Peter thought.

Invalid: "So that means that my friends could be lost", Peter worried.

Recognition items:

Paraphrase: The receptionist of the hotel did not know the names of the lost tourists.

Model predicted: Peter concluded that his German friends took a canoe trip.

Foil: A tourist who took the canoe trip was not Germans or was not Americans.

3. When Will Marina Date?

Marina got a bit overweight and she decided to follow a strict diet. She was explaining her plan to her boyfriend.

"I will do physical exercise every day, either in the swimming pool in the mornings or jogging in the afternoons.

"That's cool. But –you will continue going out with me in the evenings, won't you?" asked her boyfriend a bit worried.

"Well, I won't go out with you when I am tired from jogging. There won't be any days when I jog and then go out in the evening."

"So what exercise will you do on the days we go out?"

Validity task items:

Valid: "I'll do exercise in the swimming pool in the morning."

Invalid: "I'll do jogging."

Recognition items:

Paraphrase: Marina explained to her boyfriend her plan for a strict diet.

Model predicted: The days Marina will not exercise in the swimming pool she will jog.

Foil: The days Marina goes out with her boyfriend she had either jogged or exercised in the pool.

4. Who will go to the game?

Caroline works at a small hotel which receives guests for various professional meetings. Last weekend all the groups were from São Paulo, Rio and Curitiba. The guests from São Paulo all came for a trial-lawyers' conference. The guests from Rio came for a dentists meeting, and the guests from Curitiba came for the same meeting for dentists.

Caroline arranged a bus to take some of the guests to a football game. She checked her notes as she waited for the bus to arrive, and noticed that none of the guests who were going to the game were from São Paulo.

Validity task items:

Invalid: "Gee," she thought, "some of the guests who are going to the game aren't here for a dentists meeting."

Valid: "Gee," she thought, "every guest who is going to the game is here for a dentists' meeting."

Recognition items:

Paraphrase: Caroline arranged for a bus to take some of the guests to a football game.

Model predicted: The guests going to the game were either from Rio or Curitiba.

Foil: The guests who didn't come for a dentists meeting aren't going to the game.

5. The Lunch Specials

Juliana met her sister for lunch. They each wanted to order one of the lunch specials.

"I see that all the lunch specials come with either Coke or with beer," said her sister, when she first looked at the menu.

"Yeah," said Juliana, "and that's not good for our waist lines. I also see that the lunch specials that come with Coke come with french fries!"

"And he lunch specials that come with beer come with potatoes au graton!" said Kate.

Juliana sighed and said, "I guess we better not worry about our weight today."

Validity task items:

Valid: "The lunch specials that don't come with potatoes au graton come with french fries."

Invalid: "The lunch specials that don't come with potatoes au graton don't come with French fries."

Recognition items:

Paraphrase: Juliana and her sister wanted to order lunch specials when they had lunch together.

Model predicted: The lunch specials that didn't come with beer came with Coke.

Foil: The lunch specials came with beer, Coke, or potatoes au graton.

6. Experimental Drug

Luciano got worried when he heard that his colleague, Angela, was in the hospital. He knew that some patients at that hospital were getting an experimental drug, which he didn't trust. He talked to the head nurse on Angela's ward, who told him that all the patients on that ward had either pneumonia or dengue.

Luciano checked and discovered that none of the patients who had dengue were getting an experimental drug. When he saw Angela he asked her which disease she had, and she told him that she did not have pneumonia.

Validity task items:

Valid: "So," said Luciano to Angela, "you're not getting an experimental drug."

Invalid: "So," said Luciano to Angela, "you're getting an experimental drug."

Recognition items:

Paraphrase: All the patients in Angela's ward had pneumonia or dengue.

Model predicted: Luciano concluded that Angela had dengue.

Foil: Some patients with pneumonia were receiving the experimental drug.

7. Languages

The International Youth Club soccer championship was held in Scotland. The players stayed in hotels or in the dorms at the athletes' village.

The place had few hotels so all the foreign players stayed in the dorms at the athletes' village.

All the players who scored goals in the final game were from Brazil or Argentina. The Brazilian players all spoke Portuguese. The Argentine players all spoke Spanish.

During the championship many players were interviewed by TV stations. Every player was interviewed by either an English-language TV station or by a South American TV station. None of the final game goal scorers were interviewed on English-language television.

Validity task items:

Valid: As the players who scored in the final game spoke Portuguese or Spanish, they were not interviewed by an English-language TV station.

Invalid: As the players who scored in the final game spoke English, they were interviewed by an English-language TV station.

Recognition items:

Paraphrase: The foreign players at the Youth Club championship stayed in the dorms in the athletes' village because there were not enough hotels.

Model predicted: The players who were not interviewed by an English-lanugage TV were interviewed by a South American TV station.

Foil: The players who stayed in hotels were all either foreigners or Scots.

8. The House that Mom Wants

Emilia's mother was looking for a house to buy in Florida. She found some information about a new condo and saw that there were houses she thought she might buy. She asked Emilia to find out what her husband, Jack, thought about it. Emilia showed Jack the brochures, and said, "Look, every house comes with a swimming pool or with a Jacuzzi."

"Yes, it says here in the brochure," Jack replied, "that all the houses with a swimming pool have a built-in barbeque. I think it's important for mom to have a good barbeque."

Emilia checked the brochure and said, "It also says that all the houses with a Jacuzzi have a built-in barbeque."

"Yes, but there's still a choice to be made," said Jack, "Look at the size of the gardens: None of the houses have a swimming-pool and a big garden. Mom would certainly prefer a house with a swimming pool.

Validity task items:

Valid: "So all the houses that Mom would buy do not have a big garden but have a barbeque," concluded Emilia.

Invalid: "So some of the houses that Mom would buy do not have a barbeque but all have big gardens," concluded Emilia.

Recognition items:

Paraphrase: Emilia showed Jack the brochures with information about the houses.

Model predicted: All the houses that Emilia's Mom was interested in come with a built-in barbeque.

Foil: If there was a house with a Jacuzzi that Emilia's Mom liked, it was in another complex.

9. Where is the Musical Mouse?

Leandro and Betty were reading their newspapers during breakfast. Leandro began, as always, by reading the financial pages. Betty was commenting an article about an art exhibit they had gone to last week.

"Wasn't that strange painting of the musical mouse by Roberto Britto at that exhibit?" asked Leandro.

"That's right," replied Betty, "it was."

"It says that every painting either got sold, or was donated to the Children's Museum, and none of the paintings given to the Children's Museum are still at the exhibit." said Betty, as she continued reading.

"Is the musical mouse painting still at the exhibit?" asked Leandro as he stirred his coffee.

"I see that none of the paintings by Roberto Britto got sold," said Betty.

Validity task items:

Invalid: "So the musical mouse painting is still at the art exhibit, " concluded Leandro.

Valid: "So the musical mouse painting is not at the art exhibit anymore," concluded Leandro.

Recognition items:

Paraphrase: None of the paintings of Roberto Britto were sold at that exhibit.

Model predicted: The Musical Mouse was not among the paintings that were sold.

Foil: If Leandro is reading the arts section, he has finished with the financial pages.

EXAMPLE OF THE EXPERIMENTAL PROTOCOL - EXPERIMENT 2

---

*Idade*_____

Sexo_____

Leia a estória e indique o final mais apropriado.

Droga Experimental

Luciano ficou preocupado quando soube que a sua colega, Ângela, estava no hospital. Ele sabia que alguns pacientes no bloco onde Ângela estava estavam sendo medicados com uma droga experimental na qual ele não confiava. Ele falou com a enfermeira, e ela lhe disse que todos os pacientes naquele bloco tinham pneumonia ou dengue.

Luciano verificou e descobriu que nenhum dos pacientes com dengue estavam recebendo a droga experimental. Quando viu a Ângela logo perguntou qual doença ela tinha e ela lhe disse que não tinha pneumonia.

Final 1: "Então," disse Luciano a Ângela, "você não está recebendo a droga experimental."

Final 2: "Então," disse Luciano a Ângela, "você está recebendo a droga experimental."

Após ter terminado de ler, por favor, vire a página para responder as três perguntas sobre esta estória.

Por gentileza, não voltar para traz após ter ido para a próxima página.

<u>NÃO</u> VOLTAR PARA A PÁGINA ANTERIOR!

Cada uma das três frases a seguir contém informações que foram apresentadas na estória que você acabou de ler (embora não com palavras idênticas) ou, então, não foram apresentadas na estória, mas podiam ser inferidas (deduzidas) da estória.

Para cada uma destas três frases, por favor, indicar se as informações foram apresentadas na estória, ou se foram inferidas, ou se você está em dúvida entre estas duas opções. Indique a sua resposta com um X.

<u>Não dê a mesma resposta a todos os três itens.</u>

1. No bloco da Ângela os pacientes que não tinham dengue tinham pneumonia

❑ Foi apresentado na estória (embora não com palavras idênticas) ❑ Estou em dúvida ❑ Não foi apresentado na estória (mas poderia ser deduzido)

2. Luciano ficou preocupado porque não confiava na droga experimental.

❑ Foi apresentado na estória (embora não com palavras idênticas) ❑ Estou em dúvida ❑ Não foi apresentado na estória (mas poderia ser deduzido)

3. Alguns pacientes com pneumonia estavam recebendo a droga experimental.

❑ Foi apresentado na estória (embora não com palavras idênticas) ❑ Estou em dúvida ❑ Não foi apresentado na estória (mas poderia ser deduzido)

APPENDIX C – MATERIALS USED IN EXPERIMENT 3