Proceedings of the Sixteenth Annual Conference of the Cognitive Science Society

Edited by Ashwin Ram and Kurt Eiselt

August 13 to 16, 1994
Georgia Institute of Technology

LEA LAWRENCE ERLBAUM ASSOCIATES, PUBLISHERS
1994 Hillsdale, New Jersey Hove, UK

Mental Models in Propositional Reasoning

Bruno G. Bara
Monica Bucciarelli
Centro di Scienza Cognitiva
Universita' di Torino
via Lagrange 3, 10123 Torino, Italy
bara@psych.unito.it
monica@psych.unito.it

Philip N. Johnson-Laird
Department of Psychology
Princeton University
Green Hall, Princeton, NJ 08544
phil@clarity.princeton.edu

Vincenzo Lombardo
Dipartimento di Informatica
c.so Svizzera 185, 10149 Torino, Italy
Centro di Scienza Cognitiva
via Lagrange 3, 10123 Torino, Italy
Universita' di Torino
vincenzo@di.unito.it

Abstract

A cognitive account of propositional reasoning must consider both the representation of the propositions (premises and states of affairs) and the context in which the propositions are used. This paper is concerned with reasoning processes involving three different connectives (conjunctive, conditional and disjunctive connectives) in three different tasks (accomplishing a request for action expressed by a premise, judging a state of affairs as true or false with respect to a premise, drawing an inference from two premises). Our claim is that the ability to reason with connectives is explained in terms of construction and manipulation of mental models. We present a computer model that takes as input the modelistic representations of the premises and the specific state of affairs, compares such models and gives rise to a series of model manipulations in order to produce a result, i.e. an action, a judgement or an inference. A computer program reproduces the performances of subjects of different age groups, predicting both correct and erroneous inferences.

Introduction

One of the challenges of a psychological theory of deduction is to explain propositional reasoning, i.e. making deductions involving connectives such as *and*, *if-then*, *only-if*, *but*, *not*, *or*. Mental models theory (MMT) claims that each connective is represented by a specific number of models. E.g., suppose that *p* and *q* are tokens which identify two different assertions; a possible representation for the conjunction *p and q* is the model

p	q

(1)

where the fact that p and q lie on the same line represents that they occur together. Conventionally, in MMT several items on a single line are meant to belong to the same model; different models lie on different lines. The exclusive disjunction *p or q, but not both* requires in fact two models that are represented as

p	
	q

(2)

where the notation indicates that when p occurs q does not occur, and vice versa. Finally, the representation of a connective can also include an implicit model (represented by dots), as the initial models for the conditional *if p then q*

p	q
...	

(3)

People are supposed to represent the possibility of further alternative models for the conditional, but they are made explicit only if necessary[1]. The construction of the models proceeds from a representation which is as implicit as possible, according to principles of parsimony, to a representation which is as explicit as necessary.[2]

The reasoning process starts by <u>integrating</u> the models of the assertions joined by the connective and the model of the actual situation into a further model. A first conclusion is generated by "reading" this model. E.g., given the model (2) and a state of affairs where p is negated, the integration procedure will generate a model of the situation which supports the conclusion "q".

The procedure of <u>falsification</u> may invalidate this result by providing a different integrated model, which supports a different conclusion. If the second model does not support a different conclusion, the first conclusion can be accepted as correct. For instance, if we have the model (3) and the state of affairs p, we can conclude q by integration, as in the further models for conditional (see note 1), the conclusion is still valid.

A traditional alternative approach to human propositional reasoning is advocated by rule-based theories (RBTs). Such theories argue for a mental logic consisting of a series of formal rules of inference, e.g., a *natural deduction system*. In this view, deductive reasoning is a syntactic process where syntactic rules of inference are applied to the logical form of the premises (Braine, 1978; Braine & Rumain, 1983). The reason why MMT is a plausible candidate to explain human performances in reasoning is that it accounts for three phenomena that RBTs do not explain, that is: (a) systematic errors in reasoning, (b) difference in difficulty among deductions, (c) effect of content on the reasoning process (see Johnson-Laird & Byrne, 1993).

---

[1] Note that the models represent the true occurrences of the connectives. In particular, the complete set of the explicit models for the conditional is:

p	q
not p	not q
not p	q

[2] A mental model represents a class of tarskian models. In order to validate a deduction the reasoners must explore just this class and, as a consequence, they will draw conclusions which are only possibly true rather than necessarily true.

Our aim is to produce further evidence in favor of MMT in propositional reasoning. We refined the factors that contribute to the difficulty of the connectives, by proposing an analytical view of the steps of integration and falsification of mental models. We defined a context of use of a connective, given by the type of the task to accomplish and the actual situation. We devised a computational model based on MMT that defines the complex operations of integration and falsification from a finite set of basic procedures. Finally, we compared the performances recorded in a psychological experiment with the result generated by the computer program in the same conditions. A cognitive model simulates the production of both correct responses and errors. Error prediction is essential to assess the validity of a cognitive model, since it provides some cues about the elementary steps of the reasoning process. A main point is to correlate different kinds of errors with different ages of subjects (Bara, 1994).

Ontology

In the tasks we considered, the elementary items can be states of affairs which are *described* or *perceived*, and actions on states of affairs. By "described items" we mean elements of representation built after a verbal description; by "perceived items" we mean elements built through the visual perception of a state of affairs. An action item refers to an action on a state of affairs which is yet to be executed by the subject herself. All such items are represented in mental models as tokens of several types, listed in Figure 1. In the following we assume that the letters p, q, ... are tokens referring to states-of-affairs, while the letters a, b, ... refer to actions on states of affairs. perc and desc are two monadic functors, that, applied to tokens identifying states-of-affairs, distinguish between states of affairs respectively perceived and described. Not-p represents explicitly the falsity of a state of affairs p, that either (1) has been mentioned in a proposition but is not perceived in the actual situation, or (2) has been mentioned in the first premise but not in the second, or (3) has been explicitly mentioned as negative. Action identifiers, like a, b, ..., are functors that apply to the items that result from the application of perc and desc. The action functors are a rough notation for complex objects of a theory of action, which is outside of the scope of the paper.

p, q	= are states of affairs
perc(p)	= a state of affair *p* perceived
desc(p)	= a state of affair *p* described
α(perc(p))	= an action α on a state of affair *p* perceived
α(desc(p)	= an action α on a state of affair *p* described
not-p	= a false state of affairs

Figure 1. Examples of elementary items.

We assume that the tokens of the form perc(x) require a lower cognitive load than the tokens desc(x) in order to remain active in the working memory. This assumption is consistent with the notion of different degrees of activation due to the strength and the continuity of the stimulation of a stable visual perception with respect to a verbal description (Baddeley, 1986).

In order to account for the processing mechanisms that underlie propositional reasoning, we postulate the elementary capabilities in Figure 2, which are naturally present in human reasoners even at the early stages of development. The first three are without a common standard of comparison. Thus, we do not make any prediction about their relative difficulty. For the next four we assume an increasing degree of difficulty from the fourth to the seventh. The rationale for this assumption has been given in (Bara, Bucciarelli, Johnson-Laird, 1994).

1. processing capacity of the working memory
2. generate tokens
3. establish relations among tokens
4. detect identity of tokens ($p = p$)
5. detect difference of tokens ($p \neq q$)
6. identify references
7. generate a negative token not-p

Figure 2. Seven basic abilities.

All these abilities operate on elementary tokens. In particular, the sixth one establishes the relationship "referent of a description" between a state of affairs perceived perc(p) and a described token introduced by a verbal description desc(p). The seventh ability, i.e. the generation of negative tokens, is essential to represent explicitly the falsity of a state of affairs, or the absence of some entity.

Integration and falsification, the two complex procedures at the core of the reasoning process, are based on the elementary capabilities given in Figure 2 (see the section on the computer model).

The experiment

The connectives taken into account are *and*, *or* and *if-then*. The tasks are of three types: *action*, *truth-value judgement* and *inference*. Each connective is considered in all of the three tasks, except conjunction, which is analyzed only in the case of *truth-value judgement*. In fact, conjunction in the tasks of *action* and *inference* would be interpreted as an *if-then* connective. In order to render comprehensible the computer model and the simulation, we present here the experiment.

Experimental subjects 20 subjects (10 females and 10 males) from each of the following age groups: 3-4, 5 -6, 8-9, 11-12 and over 21. They had no previous formal training in logic.

Design and Procedure Subjects are told they are attending a test regarding the way people reason. They are presented individually with the test in a single session. The trials are presented in the following order: first, subjects deal with action and inference trials (we adopted four randomizations balanced for number and sex), then with truth-value judgement trials (we adopted a different randomization for each subject). The reason is that in the pilot experiment we found that younger subjects got confused when dealing with truth-value judgement and inference together. They sometimes tried to judge the premises of an inference and to derive a conclusion from an assertion regarding a state of affairs in the case of truth-value judgement.

The material consists of one box and a series of toy animals, which we will identify with the symbols *p*, *q*, etc. The general argument is the presence/absence of the animals inside the box. The trials consist of instances of the three tasks: let us consider them in detail.

1. (*Action*). The entity p is in the box and the entity q is outside the box. Both are plainly visible to the reasoner. One of the following requests is uttered:
"If a p is in the box, then you put a q in the box" (*if-then*)
"Either a p is in the box or you put a q in the box" (*or*)

2. (*Truth-value judgements*). One of the following situations occurs: either both entities p and q are in the box (TT - it's true that p and q are in the box), or p is in the box and q is outside the box (TF), or q is in the box and p is outside the box (FT). The arrangement is plainly visible to the reasoner. Subjects have to consider each of the following judgements:
"There is a p in the box and there is a q in the box. True or false?" (*and*)
"If there is a p in the box then there is a q in the box. True or false?" (*if -then*)
"Either there is a p in the box or there is a q in the box. True or false?" (*or*)

3. (*Inference*). The reasoner is presented with a first premise of the following type:
"If p is in the box, than q is in the box" (*if*)
"Either p is in the box or q is in the box" (*or*)

Then a second premise follows of the form: a) p is in the box; b) p is not in the box; c) q is in the box. d) q is not in the box.

The reasoner is invited to draw a conclusion.

A computer model: procedures and global strategies

The computer model simulates the human propositional reasoning in MMT. We provide an analytical account of the strategies involved, by introducing the two fundamental procedures, *match* and *make-models-explicit*.

Match takes as input two mental models and, if it succeeds, returns as output an integrated model, from which the conclusion is drawn. If match fails, the initial models are left unchanged, and no integrated model is produced. The elementary steps of *match*, which looks like a simplified version of the unification algorithm, are described by the algorithm in Figure 3. Actually, the *match* algorithm is more complex, because it also includes false states of affairs explicitly mentioned in the premises. In Figure 3 we have limited the algorithm to the cases that appear in the trials of the experiment (see the previous section). The first argument M1 is the model of the actual situation (or, in the case of the inference task, the second premise); the second argument M2 is the model of the premise against which the specific situation is to be matched. Match operates in the resolution style, by "resolving" the two models into the integrated model.

```
match(M1, M2):
if for each x [perc(x) is in M1 and desc(x) is in M2] then
   /* identify referents */
   return an integrated model M (where all the desc(x) are
   turned into perc(x))
else if for each x, f [f(x) is in M1 and f(x) is in M2] then
   /* detect identities */
   return M1
else fail
fi
```

Figure 3. The *match* algorithm.

The operation *make-models-explicit* increases the accessible informative content of a mental model. The name is a general identifier for two methods (*token-explicitation* and *model-explicitation*) that are triggered by different conditions. By "explicitation" we mean the operations that flesh out a model which is not exhaustively represented.

The first method is called *token-explicitation*. The procedures of model construction introduce only the occurrences of tokens explicitly detected or mentioned. If we need to reason by taking into account a false state of affairs, we have to make it explicit via the generation of a negative token. A token not-x is introduced into a model M for each token x mentioned in the premise and not contained in M (this treatment closely resembles the approaches to defaults and non monotonic reasoning in AI). There exist a hierarchy of difficulty: a negative token is easier to generate if it refers to a state of affairs than to an action yet to occur; dealing with a (not) perceived state of affairs is easier than dealing with a (not) described state of affairs.

The second method, called *model-explicitation*, makes explicit the models which are implicit in the initial representation of the premise. Note that also this second method employs the generation of negative tokens, since the introduction of further models occurs basically via the transformation of the initial explicit model. For instance, the explicitation of the initial model for the conditional is accomplished by, first, generating a negative token for both the antecedent and the consequent and, second, generating a negative token for the antecedent and leaving the consequent unaltered (see note 1).

Match, which involves "detect identities" and "identify referents" (see Figure 2), is simpler than *token-explicitation*, which involves the generation of negative tokens not-x. *Token-explicitation* has operationally the same difficulty of *model-explicitation*, but the acquisition of *model-explicitation* requires a longer life experience, since it implies having already experienced further situations in which the connective is true beyond the trivial case. *Match* implements integration of mental models; *make-models-explicit* followed by a *match* implements falsification.

Propositional reasoning is guided by the global strategy described in Figure 4. The framed expressions represent mental models, the expressions in italics are operations and tests on models. All the tasks start by trying to match the model of a state of affairs against the model of a premise. If the match succeeds, it produces an integrated model and

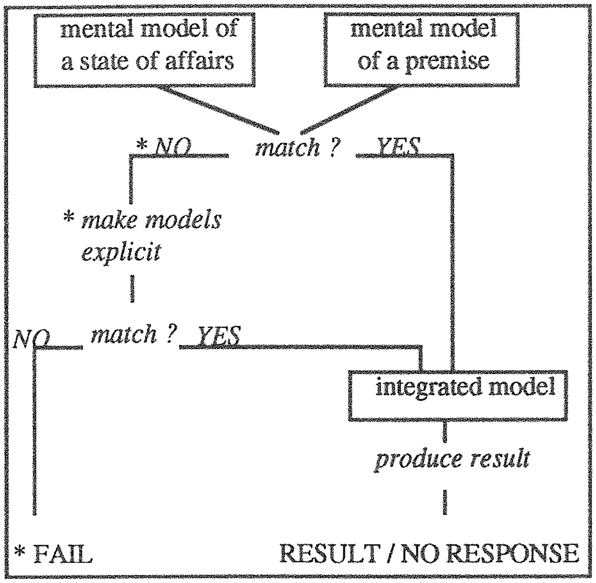

Figure 4. The global strategy

generates a result. This could be either a decision about performing an action, assigning a judgement, or drawing an inference, but if it happens that the integrated model supports both a result and its opposite, no response is possible[3]. On the contrary, if the match fails, the implicit information contained in the representation is made explicit and the match is newly attempted. If it finally succeeds, the result is extracted from the integrated model; otherwise the whole process fails.

The "mental model of a premise" represents the acontextual meaning of the connective, that is the meaning of the premise expressed by one of the representations (1), (2) and (3), introduced at the first page, that possibly include implicit information.[4] The "mental model of a state of affairs" involves either perceived entities (perc(x)) in action and judgement, or described entities (desc(x)) in inference. Together with the actual task to accomplish, the model of a state of affairs defines the context for use of a connective. The specific sequence of operations involved dynamically determines the difficulty of a "contextualized" connective.

The asterisks indicate (erroneous) possible intermediate exits of the reasoning process: if the capacity of the working memory is exhausted or the following operations are not successfully completed the response is given on the basis of the manipulation executed so far.

Beyond the general reasoning schema in Figure 4, the various tasks refine the global strategy in specific ways.

Comparison between human and artificial subjects

The computational model accounts for the analytical decomposition of integration and falsification in terms of a finite set of basic procedures. It is assumed that the construction of the initial models, that precedes these two operations, is correctly executed by all the subjects of any age group. The errors are explained in terms of failures that occur either on the execution of the basic procedures or in the sequence of operations, that may produce a large number of intermediate results, thus exceeding the capacity of the working memory.

As we cannot present all data, the comparison between the results produced by the human reasoners and the computer program will focus on a relevant case, of which we shall show the reasoning steps and the breakdown points. The example chosen involves the disjunction of the states of affairs *p* and *q* (i.e. "either there is a *p*, or you put a *q*" for action; "either there is a *p* , or there is a *q*" for judgement and inference); in all the three tasks the initial situation is always the state of affairs *p* (see the initial representations in Figure 5).

		Human subjects					Artificial subjects		
		3-4	5-6	8-9	11-12	>21	young	middle	adult
Inference	say 'there is *p*'	11	8				+++	+	+
	say 'there is *q*'	2	4	1	1		++	++	+
	SAY 'THERE IS NOT*Q*'	1	7	19	19	20	+	++	++
	say 'I don't know'	6	1				unpredicted	unpredicted	unpredicted
Action	put *q* in the box	15	15	10	5	1	++	+	+
	DO NOTHING	4	5	10	15	19	+	+	++
	'I don't know'	1					unpredicted	unpredicted	unpredicted
Judgement	SAY 'TRUE'	11	13	4	9	16	++	+	++
	say 'false'	9	7	16	11	4	+	+	+

Table 1. Type of responses with the frequences associated.

---

3 Consider, for instance, the result of *matching* the state of affairs "not-p" against the explicit models for the conditional "if p then q" in note 1. Both "q" and "not-q" are possible conclusions.

4 The general term "premise" is specialized into "request" in action, "assertion" in judgement, "first premise" in inference.

The predictions of the computer model are summarized in the right side of Table 1 under the denomination "artificial subjects". We have individuated three groups: young, which roughly corresponds to the first two groups (3-4 and 5-6) of human subjects (left side of Table 2), middle, which corresponds to the second two (8-9 and 11-12), and adult. The artificial subject entries express predictions on the competence but not on the performance of human reasoners: what is provided is a preferential order of the possible responses. The more "+" there are, the higher is the preference. The entry unpredicted concerns subjects' responses that are not accounted for by the computer model. Responses within each task are vertically arranged according to the temporal order of the possible exits in the reasoning process. The correct responses are in capital letters. Such arrangement also reflects the appearances of responses from a developmental point of view.

Let us turn now to the reasoning process described in Figure 5. All the three tasks start with a *match*, that generates an integrated model *p* (perc(p) in action and judgement, desc(p) in inference). After the integration, a *token-explicitation* is required, in order to make the integrated model explicit. Two possibilities are considered:

i) If a subject is not able to execute token-explicitation, MMT predicts two possibilities depending on the fact that subjects have a representation of the integrated model not fully explicit. The first incorrect output is based on the first disjunct desc(p). The corresponding responses are: "there is p" in inference, "true" in judgement. Such responses are expected from human subjects from 3 to 6 (young artificial subjects), whose working memory capacity may be exhausted by the integrated model, on which they are forced to base their answers. A correct answer may be given also on the basis of an integrated model not fully explicit. This explains the cross-over which is evident in Table 2, where children, in contrast with the other groups, perform judgements better then inferences. No response is supported in action, since the integrated model involves only a state of affairs. The second incorrect output is based on the model of the second disjunct desc(q). The corresponding responses are: "there is q" in inference, "put q in the box" in action, "false" in judgement. Such responses are expected from human subjects from 8 to 12 (middle artificial subjects), who take into account the second disjunct (q) on its own, but without thinking of it as an alternative model to the one that matched (p). The response "put q in the box" is also expected from younger subjects, since it is the unique possibility in action at this stage. This last MMT's prediction is strengthened by the pragmatic consideration that it is hard for a child to retain from acting, in our experimental setting.

Table 2 shows that the predicted trend is confirmed for subjects of 8-9, 11-12 and over 21 years (Page's L Test: L=439.5; p<.00001). Surprisingly, the highest percentage of correct responses of younger subjects is obtained exactly in judgement. Such results are accounted for by the analysis of the totality of possible responses in Table 1. Let us analyze whether human responses on the left follow the predictions on the right. The expectations are fully confirmed in all the tasks for the groups 3-4 and 5-6 with respect to young. Things are slightly less smooth for the groups 8-9 and 11-12 with respect to middle, since the tendency to answer

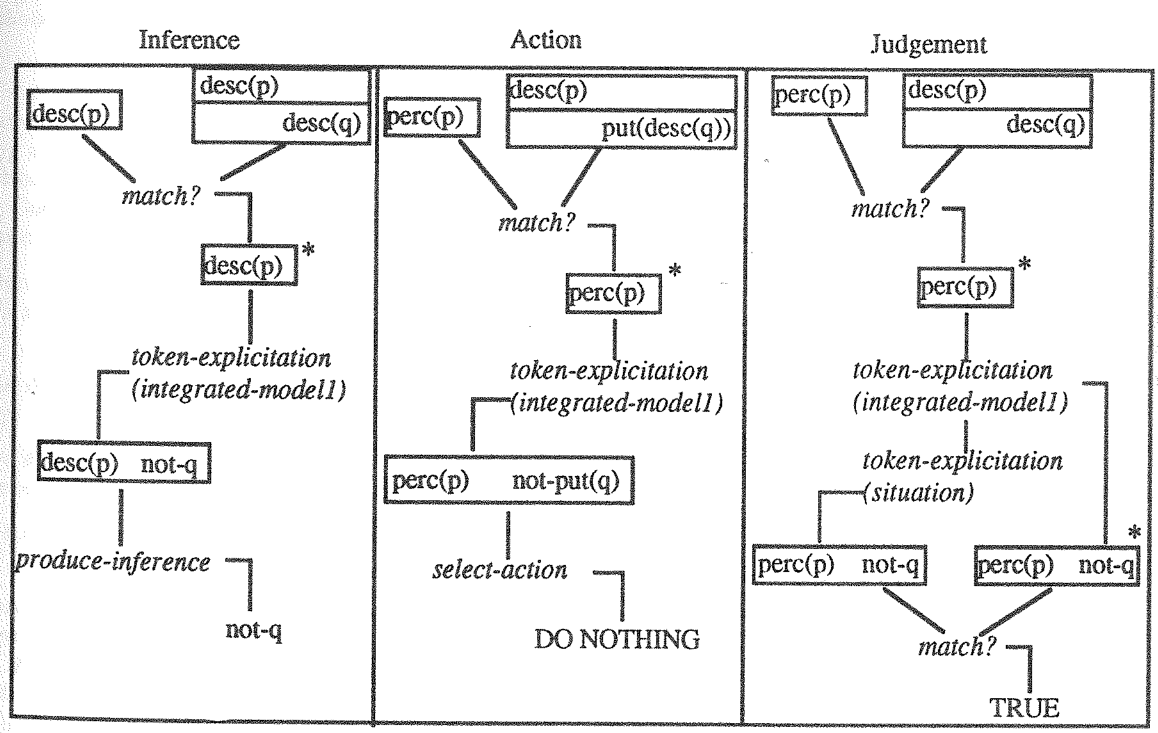

Figure 5. The three reasoning processes.

correctly runs faster than the model prediction, especially ininference (both groups) and action (11-12 group). Finally, the group >21 confirms the predictions given for adult, with only minor changes in inference, where no wrong conclusion occurs, even if some were predicted by the model.

age	Inference	Action	Judgement	global percentages
3-4	5	20	55	27
5-6	35	20	65	42
8-9	95	50	20	55
11-12	95	75	45	72
> 21	100	95	80	92
global percentages	66	53	53	57

Table 2. Percentages of correct responses.

ii) If subjects are able to execute *token-explicitation*, they meet different degrees of difficulty in dependence on the type of token. It is more difficult to generate a negative token that asserts the non-execution of an action, rather than to generate a token that falsifies a state of affairs: the former involves an action yet to occur, the latter involves a state of affairs immediately testable by perception or verbal description. The completion of *token-explicitation* should allow subjects to correctly answer in the tasks of inference and action, "there is not q" and "do nothing" respectively. But the difference of difficulty mentioned above predicts that a correct answer comes at a younger age in inference than in action. After the *token-explicitation*, judgement requires a further explicitation of the model of the situation and again a *match*. The correct response in judgement, that is "true", is expected from the adult group.

Thus, the computer model predicts for disjunction the following trend: judgment is much more difficult than both action and inference; action is more difficult than inference, but the difference between the latter two is less remarkable.

In conclusion, artificial subjects appear to realistically reproduce the actual behavior observed in experimental subjects.

Conclusion

Our interpretation of the presented results is:

1) MMT accounts for propositional reasoning in the three situations described: action, judgement and inference;

2) Development in propositional reasoning is explained by a refinement of MMT;

3) A computer model is presented, which simulates the development of competence in propositional reasoning.

Acknowledgements

Italian authors were supported in this research by the Italian National Research Council (C.N.R., bilateral project Italia-Usa, 1994).

The names of the authors are ordered alphabetically.

References

Baddeley A.D. (1986). *Working memory*. Oxford University Press, Oxford.

Bara B. G. (1994). Developing induction. *International Studies in the Philosophy of Science*, 8 (1), 31-34.

Bara B. G., Bucciarelli M., Johnson-Laird P. N. (1994). The development of syllogistic reasoning, The American Journal of Psychology, in press.

Braine M.D.S. (1978). On the relation between the natural logic of reasoning and standard logic, *Psychological Review*, 85, 1-21.

Braine M.D.S, Rumain, B. (1983). Logical reasoning. In J.H. Flavell and E.M. Markman (Eds.), *Carmichael's handbook of child psychology. Vol. III: Cognitive development* (4th Edn.), Wiley, New York.

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

Johnson-Laird P. N., Byrne R. M. J. (1991). *Deduction*, Erlbaum, U.K.

Johnson-Laird P. N., Byrne R. M. (1993). Precis of Deduction, *Behavioral and brain sciences*, 16, 323-380.