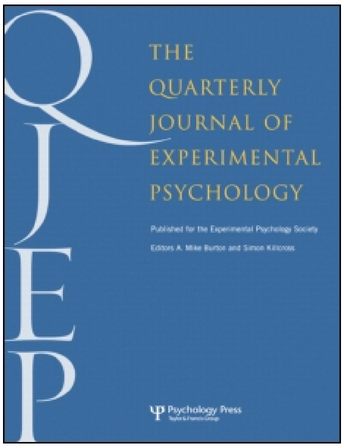

The Quarterly Journal of Experimental Psychology Section A
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713683590

Children's syllogistic reasoning

Philip N. Johnson-Laird[a]; Jane Oakhill[b]; Deborah Bull[b]
[a] MRC Applied Psychology Unit, Cambridge, U.K. [b] MRC Perceptual and Cognitive Performance Unit, Laboratory of Experimental Psychology, University of Sussex, Brighton, Sussex, U.K.

To cite this Article Johnson-Laird, Philip N. , Oakhill, Jane and Bull, Deborah(1986) 'Children's syllogistic reasoning', The Quarterly Journal of Experimental Psychology Section A, 38: 1, 35 — 58
To link to this Article: DOI: 10.1080/14640748608401584
URL: http://dx.doi.org/10.1080/14640748608401584

The Quarterly Journal of Experimental Psychology (1986) **38A**, 35–58

Children's Syllogistic Reasoning

Philip N. Johnson-Laird

MRC Applied Psychology Unit, Cambridge, U.K.

Jane Oakhill and Deborah Bull

MRC Perceptual and Cognitive Performance Unit, Laboratory of Experimental Psychology, University of Sussex, Brighton, Sussex, U.K.

Most theories of the development of deductive ability propose that children acquire formal rules of inference. An alternative theory assumes that reasoning consists of constructing a mental model of the situation described in the premises, scanning the model for an informative conclusion, and then searching for alternative models that refute this conclusion. Hence, performance should reflect two principal factors: the difficulty of constructing a model, which depends on the "figure" of the premises, and the number of models that have to be evaluated to respond correctly. In Experiment 1, two groups of children (9- to 10- and 11- to 12-year-olds) drew conclusions from 20 pairs of syllogistic premises. The results confirmed that children are affected both by figure and by number of models. Experiment 2 corroborated these findings for all 64 possible forms of syllogistic premises. The development of reasoning ability may therefore depend on the acquisition, not of formal rules of logic, but of procedures for manipulating models.

GENERAL INTRODUCTION

Many psychologists believe that children learn to reason by acquiring an internal system of logic. They gradually develop a set of formal rules of inference, which enables them to draw conclusions from any premises that meet the abstract specifications of the rules. For example, Braine and Rumain (1983) postulate the following mental rule of inference:

Requests for reprints should be sent to P. N. Johnson-Laird, MRC Applied Psychology Unit, 15 Chaucer Road, Cambridge, CB2 2EF.

The authors are grateful to Alison Black for carrying out some of the statistical analyses, and to Jonathan Evans and Don Mitchell for their criticisms of an earlier draft.

From the premises: All A are B, All B are C,
Conclude: All A are C.

The formal schema, or "logical form", of this rule is matched by the premises:

All artists are beekeepers.
All beekeepers are chemists.

and so the rule can be used to derive the conclusion:

All artists are chemists.

Reasoning is thus an essentially syntactic or pattern-matching process in which formal rules are applied to the premises regardless of their meaning. Indeed, the power of a formal calculus resides in the feasibility of this application.

Psychologists, linguists and computer scientists have adopted this approach so widely that it must rest on an assumption that is made by default—namely, deductive reasoning is impossible without access to a logic consisting of formal rules of inference (see, e.g., Inhelder and Piaget, 1964; Beth and Piaget, 1966; Osherson, 1975; Braine, 1978; Rips, 1983). However, there is an alternative theory, which makes sense of many phenomena: the reasoner imagines a state of affairs based on the meaning of the premises, formulates an informative conclusion that is true in this model of the state of affairs, and then searches for alternative mental models that may lead to refutation of the conclusion (see Johnson-Laird, 1983).

The present study has three aims. First, and most important, it is intended to throw light on how the ability to reason develops by examining the respective merits of theories based on mental logic and those based on mental models. Second, it provides some systematic data on children's syllogistic reasoning and, in particular, results based on all 64 possible forms of syllogistic premise. Although strong claims have been made about children's competence, for example, that they have an inadequate grasp of quantifiers until they attain the stage of formal operations (Beth and Piaget, 1966), perhaps surprisingly there are no data in the literature based on a full investigation of children's performance with syllogisms. Third, the study aims to rebut a criticism of the theory of mental models. Some authors have claimed that it lacks systematic results in its support (see, e.g., Rips, in press). Since the theory makes a number of predictions about syllogistic inference, and these predictions should also apply to children, the study was intended to be a step towards meeting that criticism.

Syllogisms are an excellent test case, because they lie precisely on the border of human competence: some are very easy, but others are so difficult that even adults hardly ever get them right. They are a special case of deductions that hinge on quantifiers, and quantifiers are crucial for the existence of mathematics and science as well as for much of everyday thinking, because they include such terms as "all", "some", "none", "few", "more than half", "less than five", "finitely many", and so on. If certain quantifiers are combined with the affirmation or denial of simple predicates, they yield the basic building blocks of syllogisms—namely, the four types of statement (known technically as "moods') from which the premises and conclusion of a syllogism are constructed: "All A are B", "Some A are B", "No A are B", and "Some A are not B". There are two such premises in a syllogism, and the terms in them can be arranged in only four possible ways (known technically as "figures"):

A — B	B — A	A — B	B — A
B — C	C — B	C — B	B — C

The syllogism in the first paragraph of this paper is in the first of these figures. Given that there are four figures, and four moods for each premise, there are 64 possible syntactic forms for the pairs of premises of a syllogism; 27 pairs yield valid conclusions interrelating the end terms, granted that the conclusion may relate A to C or C to A. When the conclusions are restricted to just one form, say C — A, as in Scholastic logic, then only 19 pairs of premises yield valid conclusions.

The details of the mental model theory as it applies to syllogisms, and a computer implementation of it, have been published elsewhere (Johnson-Laird and Bara, 1984), and so we will outline only its main features here. The theory assumes that deductive reasoning depends on three main stages:

Stage 1. Reasoners imagine a typical state of affairs in which the premises are true. The premises:

All the artists are beekeepers.
All the beekeepers are chemists.

could be interpreted in a form equivalent to Euler circles or Venn diagrams, but the theory of mental models postulates that logically naive individuals represent finite sets by finite sets of mental tokens and accordingly form the following sort of representations:

```
artist = beekeeper = chemist
artist = beekeeper = chemist
        obeekeeper = chemist
                     ochemist
```

where there are two artists who are beekeepers, three beekeepers who are chemists, and the prefix "o" indicates that the relevant individual may or may not exist. The actual numbers are chosen arbitrarily, because the truth conditions of the premises do not demand sets of a specific cardinality. In other words, instead of perceiving real scenes, people construct similar but internal models of the premises. Such models may be experienced as vivid images, or they may be entirely outside phenomenal experience. What is important to the theory is not the subjective character of models, but the hypothesis that reasoning depends on constructing and manipulating them, and that their underlying structure consists of sets of mental tokens corresponding to sets of individuals.

Stage 2. Reasoners scan the models they construct to determine whether they yield any conclusions. Any set of premises whatsoever always yields an infinite number of different valid conclusions, but the majority of them will be totally banal, such as a mere conjunction of the premises or a long series of disjunctions of a premise with itself. Adults with no formal training in logic tend to assume that a conclusion should be informative—that at the very least it should express a relation that is not explicitly stated in the premises—and that if there is no such conclusion that follows validly, then there is no valid conclusion at all. The model above obviously supports the following informative conclusion:

All the artists are chemists.

The fact that people draw informative valid conclusions, but not trivially valid ones, has an important implication for psychology: any theory that assumes merely that there is an internalized logic is not sufficient to account for human deductive competence. The theory needs additional principles to constrain its deductive apparatus so that it delivers only sensible, non-trivial conclusions. This point has often been overlooked (see, e.g., Inhelder and Piaget 1958, p. 305).

Stage 3. In order to guarantee the validity of a conclusion, reasoners have to consider whether there is any other model of the premises in which it fails to hold. The theory assumes that adults have some procedure, neither totally random—though in principle such a procedure would work—nor totally systematic, by which they seek variants of the initial model that refute the conclusions it supports. The theory makes no strong claims about the nature of these procedures. What matters are the numbers of different models and the consequences of failing to consider all of them. The essential characteristic of what is to count as an alternative model of the premises is that it should support a different set of conclusions. Thus, merely changing the overall numbers of tokens does not yield a genuinely different model of syllogistic premises, since it does not lead to a different set of conclusions.

The example above should be easy because there is no alternative model of the premises, and the conclusion is accordingly valid. Other premises, however, require more than one model to be constructed to arrive at a valid conclusion.

Here are some premises that require two models to be considered to make the right response for the right reason:

Some of the artists are beekeepers.
Some of the beekeepers are chemists.

They support the model:

artist = beekeeper = chemist
artist = beekeeper = chemist
oartist obeekeeper ochemist

which suggests the conclusion, "Some of the artists are chemists", or its converse. A second model refutes these conclusions:

artist = beekeeper
artist = beekeeper ochemist
oartist beekeeper = chemist
beekeeper = chemist

and so there is no valid conclusion interrelating the end terms.

Finally, as an example of a problem that requires three models, the premises:

Some of the artists are beekeepers.
None of the beekeepers is a chemist.

can be represented by the model:

artist = beekeeper
artist = beekeeper
oartist obeekeeper

---

chemist
chemist

which, given that the barrier represents negation—that is, that none of the beekeepers is a chemist—maximizes the number of identities (and negations of identities) on the smallest number of individuals. It accordingly supports the maximally negative conclusion:

None of the artists is a chemist.

or its converse. However, these invalid conclusions are refuted by a second model of the premises:

```
   artist = beekeeper
   artist = beekeeper
           obeekeeper
  ___________________________
  oartist              chemist
                       chemist
```

because the artist below the barrier could be identical with a chemist. The two models together support only the weaker conclusion:

Some of the artists are not chemists.

or its converse. A third model of the premises:

```
   artist = beekeeper
   artist = beekeeper
            beekeeper
  ___________________________
  oartist             ochemist
  oartist             ochemist
```

establishes that the converse conclusion, "Some of the chemists are not artists", is in fact invalid, because here all the chemists could be artists. The three models together support only the conclusion:

Some of the artists are not chemists.

which no further variants refute.

The 64 pairs of premises divide up into those that require one, two, or three models to be constructed in order to make the right response for the right reason.

The theory makes three main predictions, which have all been confirmed in adult performance (see Johnson-Laird and Bara, 1984, who present two versions of the theory with similar empirical consequences). First, the greater the number of models that are required to make the correct response, the harder the task will be. Second, a failure to construct the full set will lead to errors of a specific sort, namely, conclusions that are compatible with the subset of models that are constructed. Third, the form of the conclusion and the construction of the initial model depend on the figure of the premises. The latter effects are complicated, and we shall only outline them here.

The most noticeable effect of figure is on the form of the conclusions. Premises in the A — B, B — C figure show a striking bias towards conclusions of the form A — C. Conversely, premises in the B — A, C — B figure show a bias towards conclusions of the form C — A. This figural response bias was first reported by Wason and Johnson-Laird (1972). It appears to be a consequence of the "first in, first out" operation of working memory: people prefer to recall items in the same order in which they entered memory rather than in the opposite order (see Broadbent, 1958). One corollary is that if a pair of premises supports only a conclusion contrary to this preferred direction of working, the task should be harder. In fact, we have only ever tested one adult subject who performed reliably better than chance with three-model problems of this sort.

The other major effect of figure is on the difficulty of forming an initial model. Frase (1968) proposed that there was a figural effect analogous to the mediation patterns of verbal learning. His theory predicts that syllogisms in the B — A, C — B figure should be easiest and those in the A — B, B — C figure should be the hardest, and he reported some data that showed a weak effect of this sort. As Dickstein (1978) pointed out, however, the experiment used only conclusions of the form C — A, and a conclusion of this form goes against the response bias for premises in the A — B, B — C figure, while it is compatible with the bias for premises in the B — A, C — B figure. Where subjects are allowed to draw their own conclusions, the results are opposite to those predicted by Frase's theory. We therefore believe that Dickstein's explanation, which is also essentially based on the order of the occurrence of the terms, is correct. It is supported by the latencies of adults' correct responses to one model problems only; latencies are otherwise a poor measure of syllogistic performance, since they fail to correlate with accuracy (see Johnson-Laird and Bara, 1984). These authors spell out a theory that predicts that the difficulty of forming a model should increase over the four figures (in the order presented at the start of this paper).

Success in syllogistic reasoning must depend on the ability to understand the premises, to combine the information in them, and in some way to derive a valid conclusion from them. How, then, do children develop the ability to make such inferences? Paradoxically, the main findings of Piaget and his colleagues provide support for the mental model theory rather than for formal rules of inference. Inhelder and Piaget (1964) have shown that young children have difficulty in interpreting quantified assertions. If the formal theory were correct, however, it is not obvious why this semantic problem should affect an allegedly syntactic process of deduction—a process that in principle is carried out on uninterpreted symbols. Certainly, children at the upper end of 'pre-operational' thought, between the ages of 5 and 7 years, show a characteristic difficulty in understanding universal quantifiers such as "all". A typical demonstration uses the following procedure: The child is presented with a collection of blue circles, together with some red and blue squares. The experimenter then asks: "Are all the circles blue?" The typical error is to reply: "No", either because "there are some blue squares" or because "there are some red and blue squares". Inhelder and Piaget (1964) found that this sort of mistake is more frequent when the quantified term concerns colour rather than shape. It is still further enhanced when the term concerns weight (Lovell, Mitchell, and Everett, 1962). Hence, it seems that performance is better with more concrete, or more graphic, terms: similarly, Markman and Seibert (1976) have found enhanced set-theoretic judgements when the superordinate set is one that is characteristically relational, such as a family; and Donaldson (1978, p. 67) has shown that the saliency of some aspect of the situation may also mislead children.

We regard these difficulties as a result of problems that arise in forming an accurate interpretation of assertions: they reflect a tenuous grasp of the truth conditions of quantified assertions, which is easily overborne by salient aspects of the situation. If children have trouble in forming a correct mental model of quantified assertions, they are unlikely to be able to reason syllogistically. When we tested a group of 7-year-olds with a set of syllogisms (in an unpublished study), we found little evidence of their ability to select a correct conclusion from a set of only three alternatives. To obtain a significant number of correct responses, we therefore had to study older children, and we decided to investigate two age groups (9- to 10-year-olds and 11- to 12-year-olds). Children in the younger of these two groups are unlikely to have attained the stage of formal operations and might therefore be expected to have difficulty with the task, according to Piagetian theory.

Since there are adults who find most syllogisms very difficult (Johnson-Laird and Bara, 1984), we were particularly concerned that the task should not be impossible for our subjects, and we therefore decided to select the more intelligent children from our population. We assumed that they would be able to represent at least some quantified assertions correctly, and they would be less likely to produce bizarre responses. We wanted to know whether they would be able to combine the information from separate premises and to draw their own valid conclusions from them. We were interested in whether the theory of mental models could give an adequate account of performance. In particular, would the number of models predict the children's difficulty with the syllogisms? And would children be affected by the figure of the premises?

EXPERIMENT 1

Method

Subjects

Forty children were tested: twenty 9- to 10-year-olds, and twenty 11- to 12-year-olds. Given the poor performance of the pilot group of 7-year-olds, our major worry was that the task would be beyond our subjects and might even perhaps upset them. We therefore took three precautions to try to maximize the chances of their success. (1) Our selection of syllogisms contained many theoretically easy ones. (2) We allowed our subjects as much time as they wanted to draw a conclusion; as we have already noted, latency of response is not a good measure of syllogistic performance since it tends not to correlate with accuracy. (3) Our subjects were sampled from intelligent populations. The younger group were selected from two classes in an English 'middle' school, and we asked the teachers to select the five brightest girls and the five brightest boys in each class. The older group were selected on the basis of a verbal reasoning test (Nelson's Cognitive Abilities Test) in an entirely analogous way, yielding the five brightest girls and the five brightest boys from two separate classes. Despite these precautions, the data from two subjects had to be excluded from the analysis, one from each group. The younger child failed to understand the instructions properly, and the older child tried to use outside knowledge in order to perform the task.

Design and Materials

There were two groups of subjects differing in age, and both groups received the same set of 20 pairs of syllogistic premises: five problems in each of the four figures. We deliberately restricted the number of problems so as not to pose too challenging a task on the children. Likewise, since we were primarily interested in whether the children could make valid deductions, 16 of these problems supported valid conclusions interrelating the end terms, whereas the remaining four, one from each figure, did not. In the 64 possible pairs of premises, there are 11 one-model problems, 4 two-model problems, and 12 three-model problems, that yield valid conclusions. Because adults find three-model problems very difficult, we decided to give the children only six of them, and to present ten one-model problems. The four problems that do not yield valid conclusions each call for two models. This design is summarized in Table I.

The content of the problems was of a sensible everyday sort, selected so that neither one conclusion nor another was particularly plausible, as in:

Some of the footballers are musicians.
All of the footballers are runners.

We chose 20 triplets of lexical items that the children would understand, and that denoted occupations and interests, such as footballers, musicians, runners. These triplets were allocated to the 20 different forms of syllogism at random: half the subjects had one allocation, and the other half another allocation. Each problem was presented on a single file card, with one premise typed above the other. The presentation order was random for each subject, with the constraint that the four problems with no valid conclusion were roughly equally spaced out

Table I

A Summary of the Types of Problem Used in Experiment 1

	Figure of the premises				
	A–B B–C	B–A C–B	A–B C–B	B–A B–C	Totals
With valid conclusion:					
One model	3	3	2	2	10
Three models	1	1	2	2	6
With no valid conclusion:					
Two models	1	1	1	1	4
Totals	5	5	5	5	20

within the series. The first problem was always a one-model syllogism, since we felt that an easy problem should inspire some confidence in the children.

Procedure

The children were tested individually in a quiet room. Once the experimenter had established rapport with a child, she explained the general nature of the task: she was interested in how people make conclusions and wanted the children to help her. The children were told to imagine that the descriptions were about some people in a room, and they were to say what followed about these people from what they were told about them on the card. If there was nothing that followed, they could simply say "nothing". They could take as long as they liked to decide upon their answer. The experimenter then checked that the children understood the task by giving them some simple three-term series problems. When it appeared that the child had grasped the nature of the task, the experiment proper began.

Results and Discussion

The overall performance of the children was surprisingly good. We scored as a correct valid conclusion any form of words that was logically appropriate. If, for example, children included a reference to the middle term in their conclusion, such as "Some of the A are B and C", then, provided that "Some of the A are C" was valid, the response was treated as correct. The mean numbers of correct responses were 5.0 for the younger children and 6.5 for the older children, but the difference between the groups was not significant (Mann-Whitney $U = 134$, $p > 0.05$, one-tailed). The best group of adults tested by Johnson-Laird and Steedman (1978) had a mean score of 11.8 for these 20 syllogisms.

Table II

Percentages of Correct Valid Conclusions in Experiment 1

	Number of models of premises	
	1 model	3 models
9- to 10-year-olds	39	0
11- to 12-year-olds	51	0

The number of models required by a syllogism had a striking effect on performance, as is shown in Table II, which presents the percentages of valid conclusions to the one-model and the three-model problems. Obviously, the three-model problems were impossible for our subjects: not a single child in either group made any correct responses to them, whereas every subject except one of the 9-year-olds made at least one correct response to a one-model problem: 39 subjects out of 40 made more correct responses to one-model problems than to three-model problems (Sign test, $p<0.00001$). Table III presents the percentages of correct valid conclusions to the 10 one-model problems as a function of their figure. As the theory predicts, there is a reliable decline in performance across the four figures for both groups (Page's $L=523$, $z=3.81$, $p<0.001$, for the 9-year-olds; Page's $L=538$, $z=5.00$, $p<0.001$, for the 11-year-olds; see Page, 1963). The children showed some competence with the four problems that do not yield valid conclusions relating the end terms: 38% correct for the 9-year-olds, and 35% correct for the 11-year-olds.

The figure of the syllogisms biassed the order of the terms in the

Table III

Percentages of Valid Conclusions to One-Model Problems as a Function of Figure in Experiment 1

	Figure			
	A–B B–C	B–A C–B	A–B C–B	B–A B–C
9- to 10-year-olds	60	37	29	18
11- to 12-year-olds	74	56	47	16
Overall	67	46	38	17

children's responses. Table IV shows the overall effect on the conclusions drawn to all 20 problems: it presents the percentages of the responses in the form A — C and in the form C — A for all four figures, with the balance of the percentages consisting in "no valid conclusion" responses and errors of omission. Although there is a bias towards A — C conclusions in the first, third, and fourth figures (shown by every single subject), this bias disappears in the second figure. These data are slightly different from those typically obtained from adults tested with the full set of syllogisms, but they are similar to those that adults yield when drawing their own conclusions in three-term series problems (see Johnson-Laird and Bara, 1984, Experiment 2). In both cases, they seem to reflect a bias towards building the initial model on the first premise.

The results support the theory of mental models. The number of models seems to be an important factor determining the difficulty of a syllogism, and the effect of figure on constructing and interpreting models is also borne out by the data. Contrary to Piagetian lore, there appeared to be no reliable difference in the reasoning ability of the two groups of children: the 9- to 10-year-olds were not markedly less competent with one-model problems than the 11- to 12-year-olds. The next step was plainly to test children with all 64 possible pairs of premises. There was an additional reason for carrying out such an experiment. It is possible that the difficulty with the three-model problems stemmed from children's unfamiliarity with statements of the form: "Some of the X are not Y." All the valid conclusions to three-model problems are of this form, and yet there were no such statement in the premises of any of the 20 problems used in the experiment. Children may have eschewed the correct conclusions because they had not encountered statements in the appropriate form. A test of all the forms of syllogism, together with improved instructions, would be a more stringent examination of the children's difficulty with these problems.

Table IV

Percentages of Types of Conclusion in Experiment 1 as a Function of Figure (Balance Consists of NVC Responses)

	Figure			
	A–B B–C	B–A C–B	A–B C–B	B–A B–C
A–C conclusions	63	26	36	35
C–A conclusions	13	26	11	6

EXPERIMENT 2

The aim of this experiment was to test children with all 64 possible forms of syllogism and to ensure that they were aware of all the acceptable forms of conclusion. In addition, however, we were interested in determining estimates of two measures that might correlate with the ability to do syllogisms: accuracy in interpreting statements in the four syllogistic moods and the processing capacity of working memory. We accordingly obtained empirical measures of both of these variables after the experiment proper had been completed by the children. Since there were no striking age differences within the range in which some syllogistic competence is detectable, we tested only one group of children in a series of repeated sessions.

Method

Subjects

Sixteen 11-year-olds participated in the experiment (8 boys and 8 girls). They were selected from two classes in an English 'middle' school, and we asked the teachers to choose the four brightest girls and the four brightest boys in each class.

Design and Materials

All 64 possible pairs of premises were presented to the subjects in four experimental sessions separated by a week. The lists of 16 problems that were presented each week contained four problems from each figure and one syllogism in each of the 16 possible pairs of moods of premises. The proportion of valid and invalid syllogisms, and of syllogisms in each of the three model classes, remained approximately the same from one list to another.

The content of the syllogisms was similar to the materials of the first experiment, except that there was just one set of 16 lexical contents that was used in the four sessions. Each triple of words appeared once in each of the four lists, but the triples were rotated so that they occurred in a different figure and mood in each list. The order in which the four lists occurred was counterbalanced over subjects in a Williams square design; the order in which the materials were presented within a list was randomized for each subject. At the end of the final session, the children were given a diagram-checking test to measure the accuracy of their interpretation of the different mood of statement and then the test of working memory capacity.

Procedure

The children were tested individually in a quiet room following the same general procedure and instructions as before. Unlike Experiment 1, however, the experimenter spelt out explicitly an example of each of the possible forms of conclusion (including "no valid conclusion"), using a card that summarized the possibilities and that was laid out in front of the subjects. The aim was to make sure that the children were aware of the possibility of drawing a conclusion of the form "Some A are not C" or its converse. No instructions were given on how the children were to reason, but as before the experimenter checked that the children had understood the task by giving them a three-term series problem. There was one procedural modification: the problems were again presented on cards, but in addition the experimenter read aloud the premises. At the beginning of subsequent test sessions, the experimenter reminded the subjects of the sorts of conclusions that they could draw. The same set of examples was laid out for them to look at, and they were told again that some problems did not have a valid conclusion.

At the end of the experiment, the children were given a sheet of paper on which were drawn the five set-theoretic relations in the form of Euler diagrams. They were told to tick the diagram or diagrams that were "correct pictures of each statement" for each of the four moods of syllogistic premise presented in an abstract form, e.g. "All the A are B".

Finally, the children were given a test of the processing capacity of working memory. The procedure was based on the alphabet transformation task devised by Hamilton, Hockey, and Rejman (1977). The children had to transform sets of letters by counting through the alphabet for a designated number of places. Given a pair of letters, the task might be to state the corresponding pair two letters later in the alphabet (2 letters + 2), for example, for "A B" respond "C D". The subjects were supposed to remember the results of individual letter transformations until they could report the whole of the transformed sequence as a single response. We used three levels of memory load: 2 letters + 2, 2 letters + 3, and 3 letters + 3. The task was explained to the children, with examples. The lists of strings to be transformed were laid out on separate sheets of paper. The three versions of the task were presented in increasing order of difficulty, and the children were given one minute in which to transform as many of the strings as possible, working through the list in the order of its presentation. The test was administered to the children individually, and they recorded their own responses on the sheets.

Results and Discussion

Syllogistic Reasoning

Of the premises, 27 pairs yield valid conclusions, and Table V shows the percentages of them that were drawn as a function of number of models and figure. Both these variables had their predicted effects. The decline in the number of valid conclusions as the number of models increased was highly reliable (Page's $L = 217.5$, $z = 4.51$, $p < 0.0005$). The subjects were able to cope with some two-model problems (26% correct), but three-model problems defeated them (only 2% correct). The predicted decline in valid conclusions across the four figures was also reliable (Page's $L = 426.5$, $z = 2.29$, $p < 0.01$). This decline is matched by a concomitant increase in the number of "no valid conclusion" responses. Table VI presents the percentages of conclusions in the A — C and C — A forms for all 64 problems as a function of figure. The balance of the percentage, as before, consists of "no valid conclusion" responses and

Table V

Percentages of Valid Conclusions in Experiment 2 as a Function of Figure and Number of Models

	Figure				
	A–B B–C	B–A C–B	A–B C–B	B–A B–C	Overall
1-model problems	69	65	66	54	63
2-model problems			19	34	26
3-model problems	2	2	3	2	2
Overall	36	34	29	26	31

errors in which an end term was omitted. Every single subject showed a bias towards A — C conclusions in the first figure and a bias towards C — A conclusions in the second figure ($p = 0.5^{16}$); and the pattern was reliable in an analysis by materials (Wilcoxon $T = 21.0$, $N = 16$, $p < 0.01$).

Table VII presents the proportion of erroneous "no valid conclusion" (NVC) responses as a function of number of models and figure. Since these responses are not independent of the results in Table V, we have not analyzed them statistically. It is worth noting, however, that although subjects drew some erroneous conclusions from these premises, the proportion of NVCs shows an increasing trend with the number of models: as one would expect, the harder the task, the more likely that a subject responds that there is no valid conclusion.

The 37 problems that do not yield valid conclusions all call for the construction of two models in order to make the correct NVC response for the right reason. However, these problems can be divided into three categories: those with two particular premises (i.e. containing the quantifier "some"), those with two negative premises, and the remain-

Table VI

Percentages of Types of Conclusion in Experiment 2 as a Function of Figure (Balance Consists of NVC Responses)

	Figure			
	A–B B–C	B–A C–B	A–B C–B	B–A B–C
A–C conclusions	65	27	40	43
C–A conclusions	13	44	26	19

Table VII

Percentages of Erroneous "No Valid Conclusion" Responses in Experiment 2 as a Function of Figure and Number of Models

	Figure				
	A–B B–C	B–A C–B	A–B C–B	B–A B–C	Overall
1-model problems	13	17	19	17	16
2-model problems			16	21	17
3-model problems	25	25	34	31	29
Overall	19	21	23	24	22

der, which yield a valid conclusion in other figures. (Following Johnson-Laird and Bara, 1984, we have included in the first category those problems with both particular and negative premises). The percentages of correct responses (i.e. "no valid conclusion") to these three categories of problems as a function of the figure of the premises are shown in Table VIII. There was a reliable increase in correct responses across the four figures (Page's $L=447$, $z=4.07$, $p<0.0001$). Likewise, the majority of children (15 out of 16) made a greater proportion of errors on those problems that do yield a valid conclusion in other figures than on those that do not (Sign test, $p<0.0005$).

Post-Experimental Test Results

Although we asked the children how they had carried out the task, their responses were not very informative—few adults can say how they reason syllogistically, either. The children's ability to make valid deduc-

Table VIII

Percentages of Correct "No Valid Conclusion" Responses in Experiment 2 as a Function of Figure and Type of Problem

	Figure				
	A–B B–C	*B–A C–B*	*A–B C–B*	*B–A B–C*	*Overall*
Particular premises	34	39	45	48	41
Negative premises	27	38	54	50	42
Other premises	6	6	13		8
Overall	24	29	38	49	34

tions is, not surprisingly, related to their ability to make accurate interpretations of the premises. The number of diagrams that they correctly chose as consistent with each mood of premise correlated significantly with the number of valid deductions they made [Spearman's $\rho = 0.61$, $t(14) = 2.90$, $p < 0.01$].

The performance on the working memory test was measured by counting the number of complete letter strings correctly transformed. One subject was eliminated from this test because she was obviously performing a letter-by-letter translation rather than translating the sets in memory before responding. There was a small and non-significant correlation between the number of valid conclusions and the number of sets of letters correctly transformed in working memory [Spearman's $\rho = 0.36$, $t(13) = 1.37$]. The children found the transformation task extremely difficult, and some of them may not always have waited until the end of a string before recording their response.

Discussion

The results of the experiment corroborate and extend those of the first experiment. Children are adversely affected by the number of models, and they very rarely draw a correct conclusion from a three-model problem. Likewise, they show reliable effects of figure on performance in the three predicted ways: (1) it creates the expected response biases; (2) the task of drawing a valid conclusion grows progressively harder over the four figures; (3) there is a concomitant rise in the proportion of "no valid conclusion" responses, whether correct or incorrect.

A small proportion of errors are uninterpretable because they fail to include both end terms. Of the remaining erroneous responses, 68% can be directly explained as a result of failing to construct all possible models or failing to scan models in both directions. For example, the three-model problem that we illustrated in the Introduction yielded the following erroneous, though predictable, responses:

None of the A are C:	4 subjects
None of the C are A:	2 subjects
No valid conclusion:	4 subjects

One subject drew the correct conclusion:

Some of the A are not C

and one subject made an uninterpretable error by drawing a conclusion that contained the middle term, B. The four remaining subjects drew the conclusion:

Some of the A are C.

This class of error, which we have also observed in adults (Johnson-Laird and Bara, 1984), accounted for a further 22% of the interpretable errors. These errors may occur because the children take conclusions of the form "Some X are not Y" to imply that "Some X are Y", or because they make a similar inference from a premise of the form "Some X are not Y" and then reason appropriately. But, as with the example above, the refutation of an initial conclusion often depends on constructing a model that for the first time allows that some of the As are Cs. This relation may be sufficiently striking as to lead the reasoners to draw the corresponding conclusion. We believe this tendency to concentrate on one model at the expense of others is the major cause, since it is the only fact common to all the problems giving rise to the error.

GENERAL DISCUSSION

How does the ability to make syllogistic deductions develop? One possibility, which is based on the doctrine of mental logic, is that children gradually acquire, or construct, formal rules of inference that they can apply to particular patterns of premises (see e.g. Inhelder and Piaget, 1964; Braine and Rumain, 1983). We have proposed a different theory: children learn to form mental representations of the premises based on understanding their truth conditions, to draw informative conclusions from such representations, and to search for alternative representations that serve as counterexamples to putative conclusions. This theory makes detailed predictions about the relative difficulty of the 64 different sorts of syllogistic problem and about the sorts of erroneous responses that should occur in each case; existing theories of mental logic make no such predictions. The present results bear out the predictions of our theory; they also support our conception of mental models as opposed to some alternative hypotheses about the nature of representations. We will substantiate these claims by considering the major phenomena that the experiments reveal.

The more models that are required to make an inference (according to our theory), the harder the children found the task. The results of the first experiment were very striking in that not a single child ever made a correct response to a three-model problem. This phenomenon could have been an artefact arising from the lack of an explicit example of the required type of conclusion. We remedied this defect in the second experiment, and the children certainly drew conclusions of the required type to other problems, but they still made only 2% correct responses to the three-model problems. In both experiments, however, performance was best with one-model problems and better than chance with two-model problems, including those that support a valid conclusion and those that do not. Moreover, in the second experiment 90% of the errors that the children made could be explained in terms of a failure to construct sufficient alternative models, a failure to scan models contrary to the figural effect, or a failure to distinguish between what was true in one model and what was true in all. Hence, the theory of mental models certainly predicts both relative difficulty and the particular errors that children make, whereas no account based on a mental logic explains this pattern of results.

Several theorists have proposed that subjects use representations that differ in form from our conception of mental models. They argue that reasoners represent premises in a way that is equivalent to Euler circles, either directly (see, e.g., Erickson, 1974) or indirectly by using strings of symbols that correspond to them (Guyote and Sternberg, 1981). People who have been taught such techniques could undoubtedly have recourse to them to reason syllogistically, but the proposal that logically naive children use them spontaneously is more difficult to maintain. The major empirical difference between the theories based on Euler circles and our theory concerns the relative difficulty of different types of syllogism. As Johnson-Laird and Bara (1984) have shown, a simple one-model syllogism may call for as many as 16 Euler representations, whereas in contrast a difficult three-model problem may call for only 6 Euler representations. Since both Erikson (1974) and Guyote and Sternberg (1981) place limits on the numbers of representations that subjects are supposed to be able to entertain at any one time, it is not clear that this objection is decisive, but then these workers have not published their predictions about the relative difficulty of the 64 pairs of premises. However, since Euler circles are merely an alternative conception of mental models, it ought to be possible to use them in a way that would mimic the predictions of our theory—at least for syllogisms. There would remain a theoretical advantage for our conception of models: they can be readily extended to deal with multiple quantification, spatial inference, and other forms of reasoning that cannot be represented by Euler circles (see Johnson-Laird, 1983). In addition, the theory can account for the way in which ordinary individuals reason with conditionals and other connectives (see Wason, 1983; Johnson-Laird, in press), and for the way in which beliefs and prejudices influence the process of reasoning (see Oakhill and Johnson-Laird, 1985), since it allows that the construction of a model can take into account knowledge specific to a domain and thereby accommodate the striking effects of content on reasoning—a phenomenon that is entirely at odds with the formal approach.

There were marked effects of figure on children's performance, especially in Experiment 2. Thus, a problem in the figure A — B, B — C gave rise to a preponderance of conclusions of the form A — C, whereas a problem in the figure B — A, C — B gave rise to a preponderance of conclusions of the form C — A. We also observed that children made fewer valid conclusions across the four figures. Once again, such effects are not predicted by the theories based on Euler circles. The response bias, however, can be explained by postulating rules of inference that match the pattern of the first figure (Braine and Rumain, 1983), or it can be explained by assuming that responses tend to follow the order in which items enter working memory. The increasing difficulty of the four figures might also be explicable in terms of the form of the rules of inference—though no complete account of this sort is as yet available; the present theory explains both the response bias and the increasing difficulty in terms of an increase in the number of operations required to form a mental model (see Johnson-Laird and Bara, 1984).

Although our theory gives a better account of the results than do other theories, perhaps our subjects, or the task that we gave them, are not truly representative. By selecting bright children, we have obviously limited our sample to those who reason in ways that are closer to the methods used by adults. Yet the children did differ in ability, and they all demonstrated an effect of figure and all but one of them an effect of number of models. Of course, we cannot rule out the possibility that other children reason by quite different methods—even perhaps by following mental rules of inference or by using Euler circles—but there are no grounds for making this assumption, and it certainly makes the task of explaining how children develop adult competence still harder, since a wide variety of adult subjects—British, American and Italian—conform to the predictions of mental model theory. Critics of the theory (e.g. H. W. Reese, personal communication) sometimes argue that psychologists should study ecologically valid forms of reasoning, that syllogisms are not ecologically valid, and that psychologists should therefore not study them. There is a profound irony here, since the argument is itself syllogistic in form. And, of course, if it were correct (which we do not accept, see Johnson-Laird, 1983, p. 71) then the task of explaining how formal rules could have evolved for syllogistic inference is made still harder. Another tack is to argue that if we had not instructed the children to imagine that the descriptions were about some people in a room, they might not have reasoned in a manner predicted by our theory. We cannot reject the possibility, but once again we see no good reason to accept it.

Our results suggest that children have some understanding of the four types of quantified assertion, that they can spontaneously formulate informative conclusions (see Experiment 1, where they were given no examples of possible conclusion), and that they follow the principle of searching for counterexamples. Yet they are still unable to deal satisfactorily with problems that call for more than one model. Development to adult competence therefore calls for a better grasp of truth conditions, and for an improvement in the capacity to search for alternative models of the premises. Not surprisingly, the accuracy of the interpretation of quantified assertions correlated with syllogistic performance (see also Erickson, 1974). Of course, the interpretation task and the reasoning task are different and have sometimes been interpreted as giving rise to discrepant data about the tendency of subjects to convert premises, especially those of the form "All X are Y" (see Evans, 1982, Ch. 6). There are grounds for assuming that children do sometimes make errors of conversion in reasoning, especially in the more difficult figures. Thus, for example, in both experiments the most frequent response to premises of the form:

All the B are A

All the B are C

was:

All the A are C.

That such errors are more likely to occur in this figure is a direct consequence of the theory of the operations required to form a model, which postulate that here the order of the terms in at least one premise must be swapped round so as to bring the two occurrences of the middle term into contiguity. Our data fail to substantiate the thesis that reasoners as a matter of course convert all premises or are universally subject to "atmosphere" effects (see Revlis, 1975).

Our failure to correlate a measure of the processing capacity of working memory with syllogistic ability suggests that the measurement procedure may be inappropriate for children. Alternative methods need to be developed to determine whether an improvement in working memory is a major factor in the development of syllogistic reasoning. When we gave adults the working memory test (in an unpublished study), we found that their scores correlated highly significantly with their accuracy with a set of one-model syllogisms [Spearman's $\rho = 0.76$, $t(15) = 4.53$, $p < 0.001$]. In addition, when adults have only 10 sec in which to respond to syllogistic premises (see Johnson-Laird and Bara, 1984), their performance drops to a level that is remarkably similar to that of 11-year-old children. Hence, there is a prima facie case that the processing capacity of working memory constrains syllogistic performance.

Finally, there is a major theoretical difficulty in explaining the growth of reasoning ability in terms of the development of a mental logic. The problem is to explain how children could acquire formal rules of inference, given that before they possess them they are supposed to be unable to make valid deductions. It is sometimes suggested that children may abstract rules of inference from the particular deductions that they encounter in much the same way that they allegedly acquire grammatical rules (Falmagne, 1980). Since adults are hardly given to sustained public demonstrations of valid deduction, this hypothesis is viable only as a supplementary method of learning since it presupposes that children can distinguish valid from invalid arguments. Likewise, the Piagetian doctrine that rules are gradually constructed by the internalization of action has never been made sufficiently explicit to be modelled in a computer program: the theory works in a mysterious way. Faced with such problems, it is not surprising that some commentators have argued for an innate basis for mental logic (Fodor, 1980, pp. 148–149).

At no stage does the theory of mental models postulate formal rules of inference, and it is not confronted with the difficult problem of explaining their acquisition. The first stage of interpreting the premises calls for procedures that are required in any case in order to explain comprehension. Similarly, the search for counterexamples is entirely analogous to the search for counterexamples to a simple generalization, such as "All swans are white". Hence, even if one accepted mental logic, it would still be necessary to account for the two processes of comprehension and search for counterexamples. In fact, the theory postulates only one process that may not be needed to explain other abilities: the procedure for formulating informative conclusions. This process, however, has been ignored by mental logicians, who have failed to explain how people eschew trivially valid conclusions. Despite these difficulties, many psychologists hold fast to mental logic, as though to abandon it were heretical. Their conviction, we believe, rests neither on solid evidence nor on a firm theoretical foundation. It seems to be based on a lack, until recently, of any other possible explanation of deductive reasoning.

REFERENCES

Beth, E. W. and Piaget, J. (1966). *Mathematical epistemology and psychology.* Dordrecht: Reidel.

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. and Rumain, B. (1983). Logical reasoning. In J. Flavell and E. Markman (Eds.), *Carmichael's manual of child psychology: Cognitive development,* 4th edition. New York: Wiley.

Broadbent, D. E. (1958). *Perception and communication.* London: Pergamon.

Dickstein, L. S. (1978). The effect of figure on syllogistic reasoning. *Memory and Cognition,* **6,** 76–83.

Donaldson, M. (1978). *Children's minds.* London: Fontana.

Erickson, J. R. (1974). A set analysis theory of behavior in formal syllogistic reasoning tasks. In R. Solso (Ed.), *Loyola symposium on cognition,* Vol. 2. *Theories of cognitive psychology.* Hillsdale, N.J.: Lawrence Erlbaum Associates.

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

Falmagne, R. J. (1980). The development of logical competence: A psycholinguistic perspective. In R. H. Kluwe and H. Spada (Eds.), *Developmental models of thinking.* New York: Academic Press.

Fodor, J. A. (1980). Fixation of belief and concept acquisition. In M. Piatelli-Palmarini (Ed.), *Language and learning: The debate between Jean Piaget and Noam Chomsky.* Cambridge, Mass.: Harvard University Press.

Frase, L. T. (1968). Effects of semantic incompatibility upon deductive reasoning. *Psychonomic Science,* **12,** 64.

Guyote, M. J. and Sternberg, R. J. (1981). A transitive-chain theory of syllogistic reasoning. *Cognitive Psychology,* **13,** 461–525.

Hamilton, P., Hockey, B. and Rejman, M. (1977). The place of the concept of activation in human information-processing theory: An integrative approach. In S. Dornic (Ed.), *Attention and performance,* Vol. 7. Hillsdale, N.J.: Lawrence Erlbaum Associates, pp. 463–489.

Inhelder, B. and Piaget, J. (1958). *The growth of logical thinking from childhood to adolescence.* London: Routledge & Kegan Paul.

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

Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness.* Cambridge, Mass.: Harvard University Press. Cambridge: Cambridge University Press.

Johnson-Laird, P. N. (in press). Conditionals and mental models. In C. Ferguson, J. Reilly, A. ter Meulen, and E. C. Traugott (Eds.) *On conditionals.* Cambridge: Cambridge University Press.

Johnson-Laird, P. N. and Bara, B. (1984). Syllogistic inference. *Cognition,* **16,** 1–61.

Johnson-Laird, P. N. and Steedman, M. J. (1978). The psychology of syllogisms. *Cognitive Psychology,* **10,** 64–99.

Lovell, K., Mitchell, B. and Everett, I. R. (1962). An experimental study of the growth of some logical structures. *British Journal of Psychology,* **53,** 175–188.

Markman, E. M. and Seibert, J. (1976). Classes and collections: Internal organization and resulting holistic properties. *Cognitive Psychology,* **8,** 561–577.

Oakhill, J. V. and Johnson-Laird, P. N. (1985). The effects of belief on the spontaneous production of syllogistic conclusions. *Quarterly Journal of Experimental Psychology,* **37A,** 553–569.

Osherson, D. (1975). Logic and models of logical thinking. In R. Falmagne (Ed.), *Reasoning: Representation and process.* Hillsdale, N.J.: Lawrence Erlbaum Associates.

Page, E. B. (1963). Ordered hypotheses for multiple treatments: A significance test for linear ranks. *Journal of American Statistical Association*, **58**, 216–230.
Revlis, R. (1975). Two models of syllogistic reasoning: Feature selection and conversion. *Journal of Verbal Learning and Verbal Behavior*, **14**, 180–195.
Rips, L. J. (1983). Cognitive processes in propositional reasoning. *Psychological Review*, **90**, 38–71.
Rips, L. J. (in press). Mental muddles. In M. Brand and R. M. Harnish (Eds.), *Problems in the representation of knowledge and belief*. Tucson, Ariz.: University of Arizona Press.
Wason, P. C. (1983). Rationality and the selection task. In J. St. B. T. Evans (Ed.), *Thinking and reasoning*. London: Routledge & Kegan Paul.
Wason, P. C. and Johnson-Laird, P. N. (1972). *The psychology of reasoning: Structure and content*. London: Batsford. Cambridge, Mass.: Harvard University Press.

Revised manuscript received 23 June 1985