UNIVERSITY OF ILLINOIS PRESS

Development of Syllogistic Reasoning
Author(s): Bruno G. Bara, Monica Bucciarelli, Philip N. Johnson-Laird
Source: *The American Journal of Psychology*, Vol. 108, No. 2 (Summer, 1995), pp. 157-193
Published by: University of Illinois Press
Stable URL: http://www.jstor.org/stable/1423127
Accessed: 04/06/2009 23:57

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at http://www.jstor.org/page/info/about/policies/terms.jsp. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at http://www.jstor.org/action/showPublisher?publisherCode=illinois.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is a not-for-profit organization founded in 1995 to build trusted digital archives for scholarship. We work with the scholarly community to preserve their work and the materials they rely upon, and to build a common research platform that promotes the discovery and use of these resources. For more information about JSTOR, please contact support@jstor.org.

University of Illinois Press is collaborating with JSTOR to digitize, preserve and extend access to *The American Journal of Psychology.*

STOR®

http://www.jstor.org

Development of syllogistic reasoning

BRUNO G. BARA and MONICA BUCCIARELLI
Centro di Scienza Cognitiva, Università di Torino, Italia

PHILIP N. JOHNSON-LAIRD
Princeton University

We investigated the syllogistic reasoning of children 9–10 years of age, adolescents, and adults. Their performance on five tasks that theoretically might measure components of such reasoning was examined: the interpretation of quantifiers such as *some* and *all;* the referential integration of assertions; the search for counterexamples to generalizations; the perception of identical shapes within figures; and the processing capacity of working memory. Syllogistic ability improved reliably with age, though even the youngest subjects were able to draw valid conclusions well above chance to one-model syllogisms. Performance on two of the component tasks also improved reliably with age: the detection of identities, and the capacity of working memory. Multiple regressions showed that performance on these two tasks also accounted for some of the variance in syllogistic reasoning. Performance on the other three tasks was at about adult level by the age of 9. We accordingly examined performance with a group of 7-year-old children and discovered that they also performed at better than chance with one-model syllogisms. Our results support three main conclusions: young children are capable of syllogistic reasoning (contrary to the claims of Inhelder & Piaget, 1964); there is a significant development of ability from childhood to adulthood; and it is possible to identify some of the major components of this improvement.

Development of the ability to reason

How could children who do not know how to reason validly acquire the ability to do so? This question is deeply puzzling—so puzzling, in fact, that none of the three sorts of answers to it made by psychologists is viable (see Johnson-Laird, 1990). The first sort of answer is that children acquire logical ability through the normal operations of conventional learning. Given some existing ability, children might learn other procedures by the operations of generalization and specialization (Falmagne, 1980), but it is probably impossible for such mechanisms to yield the concept of validity or procedures to establish it. The second sort of answer, Piaget's, is that intellectual development is governed by an automatic tendency to self-regulation that he called *equilibration.* Each new equilibrium is the outcome of compensatory reversible operations, but each is also an occasion for further cor-

AMERICAN JOURNAL OF PSYCHOLOGY
Summer 1995, Vol. 108, No. 2, pp. 157–193

rection (see Inhelder & Piaget, 1964, pp. 292–293). Hence, development passes through a series of stages that culminates in the ability to carry out operations, not just on external objects, but on the objects of thought. These procedures, which inaugurate the final stage of "formal operations," are made possible by the development of a complete mental logic, which is supposed to correspond to the propositional calculus (see Inhelder & Piaget, 1958, p. 305). Notwithstanding the greatness of Piaget as a psychologist, this account now appears to have little substance. The experimental evidence fails to substantiate it, the theory has little explanatory power, and Piaget's logic is so idiosyncratic that in the words of some other notable proponents of formal rules of inference in the mind: "It is too problematic to stand as a psychological model of anything" (Braine & Rumain, 1983). The third sort of answer is that the learning problem for formal rules is so intractable that rules are innate (Fodor, 1980). The problem with this argument is that it also appears to rule out the possibility that formal rules could have evolved (see Johnson-Laird, 1983, pp. 142–144).

The theory of mental models suggests a radical alternative. Deductive ability does not depend on the acquisition of formal rules of inference, because there are none in the minds of logically untrained individuals. *Reasoning depends instead on a mastery of referential language and on the ability to search for counterexamples.*

That reasoning depends on an understanding of the meaning of premises may seem obvious, but no such dependence is postulated by theories based on formal rules. Our claim is consistent, however, with Carey's (1985) view that intellectual development depends, not on structural changes in mental architecture, but on the acquisition of knowledge. The logical incompetence of very young children is probably a consequence of their not yet having mastered the semantics of quantifiers. *No one has ever defended as psychologically plausible any system of formal rules for syllogistic reasoning.*

The ability to search for counterexamples depends on a knowledge of the relevance of the procedure and on a competence to carry it out. It also depends on a working memory with a sufficient capacity for such a search. The mechanism underlying the development of working memory has not yet been identified with any certainty. It may depend on maturational changes or on knowledge that enables information to be chunked more economically.

Deductive competence

For many centuries after Aristotle (330 B.C./1968), who was the first to analyze them, syllogisms[1] were the heart of logic, and perhaps for this reason they have been much studied by psychologists (for reviews, see Evans, 1982; Wason & Johnson-Laird, 1972). Psychologists have discovered at least two robust phenomena of syllogistic reasoning. The first phenomenon is that syllogisms vary greatly in their difficulty. Some are so easy that even 9-year-old children are able to draw valid conclusions from their premises (Johnson-Laird, Oakhill, & Bull, 1986); others are so difficult that hardly anyone is able to make a correct response to them.

The second phenomenon is that logically untutored individuals vary greatly in their ability to make syllogistic deductions. One way to proceed is therefore to establish the nature of the development in syllogistic ability. We know that 9-year-olds can make some syllogistic inferences, but there has been no study that shows that adolescents or adults perform any better than such children. Accordingly, the chief aims of the present study are to examine children's syllogistic reasoning, to check whether ability improves with age, and to throw light on the component processes that might underlie this development. Our approach is based on the mental model theory (Johnson-Laird, 1983) and on our earlier studies of adult syllogistic reasoning (e.g., Johnson-Laird & Bara, 1984). Our first task, however, is to address the long-standing and vexed issue of deductive competence. Next, we present a synopsis of the latest version of the model theory of syllogistic reasoning. After these preliminaries, we turn to our study of the development of syllogistic reasoning.

To what extent are logically untutored individuals able in principle to reason syllogistically? Theorists are strongly divided about this question. Some (e.g., Hamill, 1990; Henle, 1978) assume innate logical competence; hence, they claim that no one ever makes a mistake in logic. In strong contrast to this view, other psychologists have argued that untrained subjects have little or no competence to make syllogistic inferences (see Wetherick & Gilhooly, 1990). This second position goes back to the atmosphere effect proposed by Sells (1936) and Woodworth and Sells (1935): Subjects draw a conclusion that matches the mood of one or another of the premises. According to Johnson-Laird and Bara (1984), matching strategies are largely chimerical—they have been devised post hoc by theorists who have happened to notice the match between many conclusions and the mood of one or another of the premises.

In our view, most individuals are logically competent in principle. They intuit that an argument is good if the truth of its premises necessitates the truth of its conclusion. They also transcend logic in drawing their own conclusions: They grasp that a conclusion should be parsimonious, and, where possible, establish a relation that is not explicitly asserted in the premises (see Johnson-Laird & Byrne, 1991, for a defense of this claim). But, reasoners err in practice, especially if the interpretation of the premises places a heavy load on working memory. In summary, individuals who have not been trained in logic grasp the fundamental principle of validity, but they do not invariably succeed in making valid deductions. They are neither logically impeccable nor intrinsically irrational.

The model theory of syllogistic reasoning

No theory of syllogistic reasoning based on formal rules of inference has so far been proposed by psychologists, probably because the length of formal derivations fails to account for differences in difficulty among valid syllogisms (see Johnson-Laird & Byrne, 1991, pp. 116 ff.). Most theories have accordingly been based on models. Some theories have been based on Euler circles or equivalent strings of symbols (e.g., Guyote & Sternberg, 1981). The number of Euler representations does not correlate with difficulty: There are easy problems that call for many representations, and difficult problems that call for few representations (see Johnson-Laird & Bara, 1984; however, Stenning & Oberlander, 1991, have devised a new method based on Euler circles that gets round this problem in a similar way to the model theory). The theory of models postulates a different sort of representation in which assertions about finite sets are represented by a small finite number of mental tokens (Johnson-Laird, 1983). The normal process of comprehension, according to this theory, depends on procedures that construct models of the situations that are described by discourse. An assertion that describes a particular situation is represented by a single model, even if the description is incomplete or indeterminate, and this model may embody plausible assumptions based on general knowledge and even arbitrary assumptions if relevant information is lacking. If such assumptions turn out to be wrong in the light of subsequent discourse, the procedures can, if possible, revise the model so as to make it consistent with the discourse as a whole. The content captured in a model is therefore a function of both the model and the processes that revise and evaluate it. There are limits on the revision of a model: People forget the original description, the process of revision may place too great a cognitive load on the system, and so on. Nevertheless, it is possible to advance a psychological theory of inference based on the idea of manipulating models.

The general semantic principle that governs all valid deduction is that an inference is valid if its conclusion is true in every possible interpretation of its premises. In their study of syllogistic inference, Johnson-Laird and Bara (1984) argued that the difficulty of a syllogism depends on two main factors:

1. The number of models to be constructed in trying to establish a valid conclusion, given that the processing capacity of working memory is limited;
2. The particular figure of the premises, given that the order in which information is produced from working memory is optimally the order in which it entered working memory.

In this section, we will describe the latest version of the model theory (which is a slight modification of the theory in Johnson-Laird & Byrne, 1991) and the figural hypothesis.

Syllogistic reasoning depends on three principal stages: the interpretation of the premises, the formulation of a putative conclusion, and the testing of its validity. According to the model theory, reasoners proceed as follows: (a) they begin by constructing a model of the premises that makes explicit the minimum amount of information; (b) they then use this model to try to formulate a parsimonious conclusion that expresses a relation not asserted by any of the premises; and (c) to test the validity of a conclusion, they search for an alternative model of the premises that falsifies the conclusion, perhaps fleshing out the initial models more explicitly in order to do so. The interpretation of a premise of the form

All the athletes are bakers

yields the following sort of model:

[a]	b
[a]	b
. . .	

Each line represents a separate individual. There is an arbitrary number of athletes (represented by lowercase *a*'s), and each is also represented as a baker (lowercase *b*'s). There are implicit individuals represented by the three dots: They may, or may not, exist, and their properties are not explicitly represented in this initial model. The representation of the athletes is ***exhaustive*** in relation to the bakers, as is indicated by the square brackets. The term ***exhaustion*** means that any further athlete that might be added to the model must also be represented as a baker. In particular, athletes who are not bakers cannot occur in fleshing out explicitly the content of the implicit individuals represented by the three dots. Because the bakers are not exhausted in relation to the athletes, this constraint does not apply to them. Hence, a wholly explicit model of the same premises can be of the following sort:

[a]	[b]
[a]	[b]
[¬a]	[b]
[¬a]	[¬b]

where the symbol "¬" represents negation (for a defense of such "propositional annotations," see Johnson-Laird & Byrne, 1991; Newell, 1990; Polk & Newell, 1988). The form of the model is akin to a tableau in which various actors are instructed to play various parts: Two actors are playing the role of athletes (and thus bakers), another actor is not an athlete but is a baker, and another actor is neither an athlete nor a baker. A model is an internal version of such a tableau, which may be experienced as a vivid image or as something that is beyond the scope of introspection. What is crucial about a model is not its phenomenology, but its structure.

The interpretation of a premise of the form

Some of the A are B

yields the following sort of model:

a	b
a	
	b

. . .

in which neither set of individuals is exhausted. Connoisseurs of scholastic logic will recognize that the notion of exhaustion is closely related to the traditional concept of a distributed term.

The interpretation of a premise of the form

None of the A is a B

yields the following sort of model:

[a]	
[a]	
	[b]
	[b]

. . .

Because the *b*'s are exhausted, it follows at once that the *a*'s can be explicitly represented as not-*b*'s:

[a]	¬b
[a]	¬b
	[b]
	[b]

. . .

Hence, no A can be added to a B in the model, and vice versa.

Finally, the interpretation of a premise of the form

Some of the A are not B

yields the following sort of model:

a	¬b
a	¬b
	b
	b

. . .

This initial model supports the converse assertion:

Some of the B are not A

which is a fallacious inference that is often made (see e.g., Wilkins, 1928). But, this conclusion is falsified by an alternative model of the premise:

a ¬b
a ¬b
a b
a b
. . .

The number of tokens representing a set is assumed to be arbitrary, though always small and plural.

The principles for forming an integrated model that combines the interpretations of the two premises are straightforward. First, following Johnson-Laird and Bara (1984), two operations may be used to bring the tokens corresponding to the middle terms into contiguity: (a) reversing the order of the two models before they are combined, and (b) inverting the order of the items in a model. These operations account for various effects of figure, and they apply according to the figure of the premises in the following way:

Figure A-B B-C:	No operations needed.
Figure B-A C-B:	Reverse the order of the two models.
Figure A-B C-B:	Invert the order of elements in the second model (or reverse the order of the two models and then invert the order of elements in the A-B model).
Figure B-A B-C:	Invert the order of elements in the first model (or reverse the order of the two models and then invert the order of elements in the B-C model).

The operation for combining models is simple: Ensure that the models of the two premises contain the same number of tokens corresponding to the middle term, and then join these individuals one-to-one. For example, given premises of the form

Some of the A are B
All the B are C

the first premise yields the model

a b
a
  b
. . .

and the second premise yields the model

[b] c
[b] c
. . .

and so the result is

a	[b]	c
a		
	[b]	c
. . .		

Likewise, for premises of the form

None of the A is a B
All of the B are C

the initial combined model is

[a]	¬b	
[a]	¬b	
	[b]	c
	[b]	c
	. . .	

Conclusions are formulated by describing the relation between the tokens representing the two end terms (i.e., the two terms that occur in separate premises). If there are no negative tokens in a model, then the description is affirmative: If each end token, x, occurs in an individual that also contains the other end token, y, then the conclusion is

All the X are Y.

Otherwise, if at least one end token, x, occurs in an individual that also contains the other end token, y, then the conclusion is

Some of the X are Y.

And if this condition fails, then no valid conclusion can be drawn. If there is a negative token in a model, then the conclusion is negative: If the end tokens, x and y, do not occur in the same individual, and at least one of these two sets is exhaustively represented, then the conclusion is

None of the X is a Y.

Otherwise, if at least one individual is an x but not a y, then the conclusion is

Some of the X are not Y.

And if this condition fails, then no valid conclusion can be drawn.

The theory postulates that human reasoners attempt to search for alternative models that will falsify putative conclusions. The evidence, as we shall see, implies that they have no simple or certain algorithm for making such searches. Indeed, little is known about how they attempt to make searches beyond the fact that the task is difficult and likely to defeat them. The computer program implementing the theory accordingly draws conclusions from the initial model of the premises, scanning the model first in one direction and then the other; it then uses three sorts of operation to search for alternative models—breaking individuals into two, adding new individuals to the model, or joining two separate individuals into one. The operations may not be psychologically realistic; they are designed merely to ensure that the program's search is ultimately complete. For an affirmative conclusion, if there is an individual containing both end tokens, a and c, and the middle term, b, which is not exhausted, the first operation *breaks* this individual into two separate individuals:

```
a    b    c
```

becomes

```
a    b
     b    c
```

Otherwise, if an end term is not exhausted, the second operation *adds* further such end terms to the model, one at a time unless more are needed to falsify a conclusion. For example, given the initial model

```
[a]   [b]   c
[a]   [b]   c
   . . .
```

and the conclusion

All the C are A

the operation of adding *c*'s yields a new model that refutes the conclusion

```
[a]   [b]   c
[a]   [b]   c
            c
   . . .
```

For a negative conclusion, if an individual corresponds to one exhausted end term but not the other, then the third operation moves the individual and *joins* it to one corresponding to the other end term. Such moves are made one at a time unless more are needed to falsify a conclusion. For example, the initial model

```
[a]   ¬b
[a]   ¬b
      [b]   ¬c
      [b]   ¬c
            [c]
            [c]
```

yields the conclusion

None of the A is a C.

This conclusion is refuted by two joins:

[a]	¬b	[c]
[a]	¬b	[c]
	[b]	¬c
	[b]	¬c

Otherwise, if there is a nonexhausted end term, *add* further instances of it to the model one at a time unless more tokens are needed to falsify a conclusion. For example, the initial model

[a]	¬b	
[a]	¬b	
	[b]	c
	[b]	
		c

supports the conclusion

None of the A is a C.

This conclusion is refuted by adding a token:

[a]	¬b	c
[a]	¬b	
	[b]	c
	[b]	
		c

Each model is a possible situation described by the premises, and a valid conclusion must hold in all of them—unless it merely states a possibility. In some cases where the premises yield three models, a slight change in the algorithm will yield only two models. At present, we have no way of discriminating between these alternatives (cf. Johnson-Laird & Bara, 1984, who also described two distinct computer programs, one of which never constructed more than two models of the premises, and the other of which, where possible, constructed three models of the premises). The theory accordingly draws the critical distinction between one-model problems and multiple-model problems. The Appendix presents all 64 possible pairs of premises and indicates the 10 one-model problems and the 17 multiple-model problems with valid conclusions (Tables A1–A4). The remaining 37 pairs of premises are multiple-model problems with no conclusion that holds for all the models, and so they have no valid conclusion interrelating the end terms.

Componential study of syllogistic inference in children, adolescents, and adults

If the model theory is correct, the development of syllogistic reasoning occurs under the influence, not of a specific mechanism for formal rules, but of a few basic underlying modules that concern the construction of models, their linguistic descriptions, and the search for falsifying models. This claim is akin to Sternberg's (1985) analysis of intellectual components, where a component is an elementary information process that operates on internal representations. The goal of the present study is therefore to determine the basic component skills of syllogistic reasoning according to the model theory; to devise experimental tests to measure each of these components; and to assess the extent to which competence at each of these components accounts for syllogistic ability at all stages of development from childhood through adolescence to adulthood.

The theory of models implicates five major components in syllogistic inference:

1. The interpretation of the premises depends on the ability to understand the meaning of quantified expressions. Likewise, the formulation of a conclusion that holds in one or more models depends on a comprehension of the meaning of quantifiers.

2. The integration of the information from the second premise depends on establishing that it refers to entities introduced into the model according to the first premise. Unless reasoners grasp this relation, which is based on the two occurrences of the middle term, they will be unable to integrate the two premises. This process is evidently sensitive to the figure of the syllogism. Conversely, individuals who have difficulty in establishing referential relations will have problems in integrating information from premises in which the two occurrences of the middle term are not contiguous.

3. After the formulation of a conclusion, reasoners must search for counterexamples if they have to reach a valid conclusion for multiple-model problems or to declare correctly that there is no valid conclusion (holding between the end terms). We need to ascertain whether subjects of different ages realize that it is necessary to test a putative conclusion by searching for counterexamples to it, and whether they are able to do so. This ability is clearly primarily relevant to coping with multiple-model problems: The correct response to a one-model problem does not call for any ability to search for counterexamples.

4. The construction and falsification of models depends on the ability to notice identities between them. This ability is obviously necessary if subjects are to discern what conclusion, if any, holds over a set of models. It is also likely to play a part in the integration of information from the two premises. Hence, it should predict inferential ability with all sorts of syllogisms, valid and invalid.

5. The formulation of conclusions and the search for counterexamples depend on a working memory that has a sufficient processing capacity. This capacity develops with age (see Baddeley, 1986). Hence, we can predict that syllogistic ability—granted that it is an inferential process, rather than a superficial matter of matching premises—should also develop with age. Although it is commonly asserted that a development in syllogistic ability does occur, it does not seem to have been studied in any comprehensive cross-sectional study (though see Johnson-Laird et al., 1986, for a comparison of children 9–10 and 11–12 years of age).

There is at least one other ability necessary for successful syllogistic reasoning. An important component in understanding is the process of parsing the premises and combining the meanings of their constituents according to the grammatical relations between them. We did not attempt to assess this ability, but rather assumed that it was part of some of the components that we did measure, (e.g., the ability to understand quantified assertions). There is no simple way in which to factor out this component and to examine it in isolation from other mechanisms.

EXPERIMENT

METHOD

Experimental tasks to test the five components

Interpretation of quantifiers. We assessed the subjects' interpretations of quantified assertions by asking them to classify a set of drawings in terms of those that were truthfully described and those that were falsely described by assertions in the four moods. The drawings corresponded to the main set-theoretic relations that can hold between two sets: A and B have the same members; A is a proper subset of B; B is a proper subset of A; A and B overlap in their members; A and B are disjoint. For this test to make sense to the children, the drawings were made so that A was a set of women and B was a set of dancers. The drawings were put on a table in front of a subject in a random order, and, for each sentence, the subject was invited to choose the drawings that were consistent with the description.

Referential integration of sentences. As in Ehrlich and Johnson-Laird (1982), the subjects listened to spatial descriptions of objects and then attempted to place actual objects into the appropriate positions on a table. To avoid confusion, the table was covered by a large sheet of paper labeled with the four cardinal points:

BEHIND
LEFT RIGHT
FRONT

We presented three series of problems in a predetermined order of increasing difficulty. Each series consisted of four trials in a different randomized order for every subject. The first series were problems with two coreferential sentences, such as

> The pear is behind the book.
> The book is to the right of the cup.

The second series were problems with three coreferential sentences, such as

> The ring is in front of the watch.
> The watch is to the left of the necklace.
> The necklace is in front of the pendants.

The third series were problems in which either the last two sentences or the first two sentences did not contain coreferential items, such as

> The album is behind the brush.
> The pen is in front of the bottle.
> The pen is to the left of the brush.

In this case, the subjects can integrate the information only after they interpret the third assertion.

Search for counterexamples. The subjects were given three trials with a task that tested their grasp of the "logic" of falsification. The experimenter put seven small toys ("Smurfs") on the table in a random order together with seven drawings representing each of them. The experimenter then removed all the toys and put some of the them in a box labeled with a description. The subjects' task was to find out whether or not the label was an accurate description of the contents of the box. They did so by picking out a drawing and asking the experimenter whether the corresponding toy was in the box. The essential insight is that those toys that do *not* fit the description are crucial, because if one of them is in the box the label is false. According to the model theory, the task bears on syllogisms because a test of the validity of a conclusion depends on searching for models that do *not* fit the conclusion. The three tasks progressively increased the complexity of the description, which has been shown to affect insight into falsification in this task (Oakhill & Johnson-Laird, 1985). The first trial used a simple description ("motorist"), the second trial used a conjunctive description ("musician with wind instrument"), and the third trial used a disjunctive description ("chef with spoon or food").

Perception of identities. To test subjects' ability to find identical elements in models, we presented them with pairs of drawings in which they were asked to mark in pen on each drawing that part that was identical in the pair. The seven pairs of drawings were presented in a predetermined order of increasing difficulty. The first pair of drawings consisted of (a) a triangle and (b) a triangle on the top of a square, whereas the seventh pair consisted of (a) a girl carrying a decorated box and (b) a country scene of a farmhouse with mountains in the distance—the identical element was the lid of the box, which matched the roof of one of the farm buildings.

Processing capacity of working memory. To assess the subjects' working memory capacity, we tested their ability to recall series of spoken digits. In the first part of the test, they had to repeat the digits in the same sequence as they had heard them. The initial trial was a series of two digits, and for each of the subsequent trials the number of digits to be recalled was increased by one. If subjects made an error, they were presented with a different series of the same length. If they erred again, the test was terminated; but if they were correct, the normal trials resumed with a longer series. In the second part of the test, the procedure was identical except that the subjects' task was to repeat the digits backward (i.e., in the opposite order to which they had heard them). The number of digits the subjects were able to repeat correctly was recorded.

Subjects

Sixty subjects belonging to three different age groups participated. We selected 20 children 9–10 years of age from a local school in Florence, Italy, on the basis of their teachers' assessment that they were progressing well at school and had the sort of home background that suggested that they were likely to go to university. We used this procedure to maximize our chances of selecting children who would be able to cope with the task (see Johnson-Laird et al., 1986). The group was made up of 7 females and 13 males. We tested 20 adolescents at high school (10 females and 10 males) who volunteered to take part in the experiment. The high school was one commonly considered to be an intermediate step toward university. We tested 20 adult volunteers (10 females and 10 males) who were university students over the age of 21 years and attending courses in the Department of Psychology at the University of Florence. None of the subjects had attended any courses in logic. The children received small gifts for participating in the experiment, and the other subjects were paid 8,000 lira (approximately $6.50).

Design and materials

Three groups of subjects participated: children 9–10 years of age, adolescents 14–15 years of age, and adults over the age of 21 years. The subjects took part in two sessions: In the first session, we examined their ability to draw their own conclusions from syllogistic premises. The children were tested with 28 logically distinct syllogisms: all 10 one-model syllogisms, 9 multiple-model syllogisms with valid conclusions (interrelating the end terms), and 9 multiple-model problems with no valid conclusions (interrelating the end terms). This subset of syllogisms contained the easiest syllogisms, selected according to the results of our previous experiments. The adolescent and adult groups were tested with all 64 possible pairs of syllogistic premises. We tested the children with a smaller number of syllogisms because they tired after about 30 syllogisms. The syllogisms were presented in two different random orders, randomly assigned to the subjects, balancing according to the gender of the subjects.[2] In the second session, a few days after the first, we gave the subjects the five tests designed to assess the basic components predicted to underlie syllogistic performance. They were presented in the following order: the perception of identities, the grasp of coreference, the capacity of working memory, the comprehension of quantified assertions, and the ability to search for counterexamples. Although the use of a fixed order made it impossible to control or to assess either effects of practice or residual effects from one test to another, in our view the procedural advantage of a fixed order outweighed the danger of gross effects of these types.

Procedure

The subjects were tested individually in a quiet room. They were told that they were going to take part in an experiment on how people reason. Their task was to state in their own words what conclusion, if any, followed from the statement of each problem. If they considered that nothing followed, then they were to say so. To familiarize the subjects with the reasoning task, the experimenter verbally presented six 3-term problems, two of them without valid conclusions interrelating the end terms. If the subjects made a wrong response to any of these practice problems, they were encouraged to try again. Once the subjects understood the nature of the reasoning task, the real test followed. Because the children were slower to read the premises than the other two groups, we minimized this difference using the following procedure. The experimenter read the first premise of a syllogism. When the subject nodded, the experimenter read the second premise, and at the same time put the typewritten premises on the table. The subjects were told that the maximum time they could take to respond to a syllogism was 2 min. The experimenter recorded the subjects' verbal response to the premises.

In the second session, the experimenter carried out the five tests of the component abilities. Each test was explained in sufficient detail for the subjects to grasp what was demanded of them.

RESULTS

Overall results for syllogisms

Table 1 presents the percentages of correct answers made by the three different age groups to the main sorts of syllogisms (in the subset of 28 presented to the children). The Appendix shows these percentages for each of the individual syllogisms (28 syllogisms for the children, and 64 for the adolescents and the adults), and also identifies whether the syllogism is one model or multiple model. The results support the predictions of the model theory. One-model syllogisms were easier than multiple-model ones with valid conclusions: This difference held for 59 of the 60 subjects (Wilcoxon Test, $z = 6.72$, $p < .00001$). Although the sorts of responses are entirely different, the one-model problems were also reliably easier than the multiple-

Table 1. Percentages of correct answers made by the three groups of subjects to the three sorts of syllogisms contained in the subset of 28 syllogisms presented to the children

Group	One model	Multiple model (valid)	Multiple model (invalid)	Overall percentage
Children	66	10	7	29
Adolescents	76	14	16	39
Adults	83	19	33	50
Overall percentage	75	14	19	39

model problems with no valid conclusions (this difference held for all 60 subjects, $p = .5^{60}$). As expected, an increase in age improved performance. The overall improvement was significant (Jonckheere Trend Test, $z = 5.54, p < .00001$); the trend with one-model problems was significant (Jonckheere Trend Test, $z = 2.56$, $p < .005$); and the trend with multiple-model problems was even more significant (Jonckheere Trend Test, $z = 3.61$, $p < .0001$).

The results also corroborate the figural effects. There are many possible ways in which to present the data, but they yield comparable effects. Table 2 shows the percentages of A-C conclusions for those problems where an A-C or a C-A conclusion is valid; the balance of the percentages are for C-A conclusions. In figure A-B B-C, as we predicted, there was a bias in favor of A-C conclusions: All 60 subjects conformed to it ($p = .5^{60}$). In figure B-A C-B, the predicted bias C-A was not reliable for the subjects overall (Wilcoxon Test, $z = 0.66$, $p > .25$), but 16 of the 20 adults conformed to the prediction (Wilcoxon Test, $z = 2.19$, $p < .01$). For the symmetric figures A-B C-B and B-A B-C, 54 of the 60 subjects conformed to the expected bias of an A-C conclusion (Wilcoxon Test, $z = 5.92$, $p < .00001$).

The model theory predicts a decline of correct conclusions over the four figures of premises, and an increase of "no valid conclusion" responses. Table 3 shows the percentages of correct valid conclusions to all syllogisms as a function of figure. The expected decline as the difficulty of figure increases is highly reliable (Page's L Test, $z = 6.46$, $p < .000001$). To retain the greatest possible amount of information, we have always considered (with the exception of Table 1) data coming from the entire set of 64 syllogisms for adolescents and adults. Because the subset of 28 represents the easiest syllogisms, the overall percentages shown in Table 3 (40% for children, 37% for adolescents,

Table 2. Percentages of A-C conclusions to all problems except those that have a valid conclusion of either an A-C or a C-A form; the balance of the percentages are C-A conclusions

	Figure of the premises				
Group	A-B B-C	B-A C-B	A-B C-B	B-A B-C	Overall percentage
Children	92	52	79	79	76
Adolescents	91	67	77	87	80
Adults	93	32	63	70	65
Overall percentage	92	50	73	79	74

and 43% for adults) may seem to lose the developmental trend shown in Table 1. But, if we consider the subset of 28 syllogisms for the three age groups, the overall percentages of correct valid conclusions do reflect the expected trend: 40%, 50%, 57%, respectively. Table 4 shows the percentages of erroneous "no valid conclusion" responses to all syllogisms. The effect of figure is again highly reliable (Page's L Test, $z = 6.25, p < .00001$). In both cases, however, the actual data suggest that the principal difference is between the asymmetric and symmetric figures; moreover, the trend suggests that the A-B C-B figure is easier than the B-A B-C figure. The percentage of "no valid conclusion" responses given by younger subjects is very low, which may be because these subjects are biased against such responses or, as the model theory predicts, because these subjects are often unable to construct more than one model of the premises.

Because model theory assumes that working memory capacity underlies both the effect of the number of models and the effect of figure, we carried out an analysis of variance (SPSS-PC) on the frequencies of correct valid conclusions as a function of these two variables (see Table 5), with age as covariate. The effect of models was reliable, $F = 2414.7, p < .0001$, as was the effect of figure, $F = 16.04$, $p < .01$, and they do interact according to our expectation, $F = 6.352$, $p < .0001$. In general, models and figure appear to exert about the same relative effect at each age.

Results for individual syllogisms

Certain aspects of the development of syllogistic performance become evident only if one examines performance on individual syllogisms (see Appendix). Many problems show a general tendency toward an improvement with age. These improvements can usually be ex-

Table 3. Percentages of correct valid conclusions as a function of the figure of the premises

	Figure of the premises				
Group	A-B B-C	B-A C-B	A-B C-B	B-A B-C	Overall percentage
Children	58	44	41	23	40
Adolescents	43	43	44	25	37
Adults	53	41	38	41	43
Overall percentage	51	43	41	30	41

plained in terms of an increasing ability to construct alternative models of the premises. For example, with the following problem in the A-B B-C figure

Some of the A are B
Some of the B are C

none of the children correctly responded that there was no valid conclusion, whereas a small number of adolescents and adults did respond correctly. The younger children accordingly appear to construct the following sort of model:

```
a   b   c
a
    b
        c
. . .
```

whereas the older subjects who responded correctly were able to construct the alternative model:

```
a   b
    b   c
a
    b
        c
. . .
```

This ability to construct alternatives, particularly where an alternative breaks a link between tokens of the end terms, gave rise to some qualitative differences in the performance of the three groups. We lack the space to examine all of the cases in detail, but in the typical problem

Table 4. Percentages of erroneous "no valid conclusion" responses as a function of the figure of the premises

	Figure of the premises				
Group	A-B B-C	B-A C-B	A-B C-B	B-A B-C	Overall percentage
Children	0	2	14	7	6
Adolescents	15	19	6	3	11
Adults	18	23	54	40	34
Overall percentage	11	15	25	17	17

All the A are B
Some of the B are C

12 of the children drew the conclusion:

All the A are C

but only 3 subjects in the other two groups did so; their typical conclusion was

Some of the A are C.

In fact, the correct response is "no valid conclusion." We infer that the children formed the following sort of model:

[a]	b	c
[a]	b	c
	b	¬c
	b	¬c

. . .

which supports their conclusion, and they go no further. The adolescents and adults find an alternative model that severs one of the links from a to c, but few subjects succeed in building a model of the premises in which all the links from *a*'s to *c*'s are severed:

[a]	b	¬c
[a]	b	¬c
		c
		c

. . .

The increasing ability to construct alternative models accounts for the qualitative differences in the conclusions of the three groups for the following problems: EO in figure A-B B-C; IA in figure B-A

Table 5. Percentages of correct valid conclusions as a function of the number of models and the figure of the premises

Group	Number of models by figures									
	Figure 1		Figure 2		Figure 3		Figure 4		Overall	
	1m	mm	1m	mm	1m	mm	1m	mm	1m	mm
Children	72	15	58	0	75	18	60	5	66	10
Adolescents	75	12	68	17	88	23	78	10	76	14
Adults	95	10	73	8	70	21	90	26	83	19
Overall	81	12	67	8	78	21	76	14	75	14

Note. One model, 1m; multiple model valid, mm.

C-B; AI in figure A-B C-B; AI, II in figure B-A B-C. It would probably account for still other problems, but as the reader will recall, the children were tested with only 28 of the 64 possible syllogisms.

In certain problems, an improvement in performance with age cannot be explained in terms of the construction of alternative models of the premises. These include problems where the children appear to have difficulty in dealing with negation. With premises of the form

All the A are B
None of the B is a C

9 of the children concluded

All the A are C.

Subjects in the older groups never made this error, and tended to conclude correctly

None of the A is a C.

One possibility is that children do not build the same sort of model for negative premises that we postulated earlier. Thus, for the premise

None of the B is a C

young children may construct a model corresponding to "All the B are not C":

$$\begin{array}{ll} [b] & \neg c \\ [b] & \neg c \\ \ldots & \end{array}$$

Such a model is entirely compatible with the earlier sort of model that we proposed:

[b]	¬c
[b]	¬c
	[c]
	[c]

. . .

However, when the children integrate the two premises, they may forget the negative element, and so construct the following model:

[a]	b	c
[a]	b	c

. . .

which supports the conclusion: All the A are C. A similar explanation applies to the EA problem in the B-A C-B figure.

Certain problems may reflect another aspect of negation. For example, premises of the form

Some of the B are not A
All the B are C

elicited a large number of "Some of the A are C" and "Some of the C are A" conclusions from both the children and adolescents, whereas the adults more often drew the correct conclusion: "Some of the C are not A." The younger subjects appear to construct the following model:

a	[b]	c
¬a	[b]	c
a		

. . .

and to describe it affirmatively. Indeed, they may draw the implication that "Some of the B *are* A" from the first premise. The adults, however, appear to be able to form the following models of the premises:

¬a	[b]	c
¬a	[b]	c
a		
a		

. . .

and

¬a	[b]	c
¬a	[b]	c
a		c
a		c

. . .

They are accordingly able to draw the correct negative conclusion:

Some of the C are not A.

Results of the five tasks

Table 6 summarizes the performance of the three age groups on the five component tasks.

Interpretation of quantifiers. None of the groups did well in this task, and there was no improvement with age (Jonckheere Trend Test, $z = 1.21, p > .1$). All three groups tended to interpret "All the A are B" as meaning "All and only the A are B": Each of the children made this interpretation, and so did 14 of the adolescents and 14 of the adults. The adults tended to treat "Some of the A are B" as inconsistent with the case where the A's are a proper subset of the B's, but the other two groups performed more in accordance with the logical interpretation of "some," treating it as meaning "at least some, but possibly all." Insofar as adults converge on an interpretation of quantifiers, the bulk of their learning appears to have occurred prior to the age of 9 years.

Referential integration. The subjects, as we predicted, performed better with the referentially continuous sentences than with semi-continuous and discontinuous sentences (Wilcoxon Test, $z = 4.87$, $p < .00001$). There was no reliable improvement in performance with age: Neither the trend over the three groups (Jonckheere Trend Test, $z = 0.52$, $p > .3$) nor the difference between the two extreme age groups was significant (Mann-Whitney U Test, $z = 1.39$, $p > .08$). Hence, the ability to integrate referentially continuous assertions, such as those that occur in syllogisms, has clearly been mastered before 9 years of age.

Search for counterexamples. We scored the subjects' performance by subtracting the number of positive examples they selected from the number of counterexamples. Because each of the three trials used three positive instances and four counterexamples, a subject could obtain a minimum score of –9, and so we added 9 to the scores to ensure that all were positive. The majority of subjects (of all ages) selected both positive and negative instances; there was no significant improvement in performance as a function of age, and even the difference between the adults and children was not significant (Mann-Whitney U Test, $z = 1.27, p > .1$). These results suggest that intelligent 9-year-olds grasp the need in principle to search for counterexamples, and this finding is borne out by the fact that they do spontaneously respond that there is no valid conclusion to certain syllogisms.

Perception of identities. Five of the seven pairs of drawings ex-

Table 6. Performance of the three age groups on the five component tasks

Group	Interpretation of quantifiers	Referential integration	Counter-examples	Perception of identities	Working memory capacity
	(20)	(12)	(21)	(7)	(36)
Children	15.9	6.1	8.5	6.0	20.7
Adolescents	16.5	6.0	8.4	6.1	22.7
Adults	16.2	7.4	10.3	6.7	27.9

Note. Maximum scores for each task are shown in parentheses.

hibited a ceiling effect, but the remaining two pairs of drawings showed a trend for performance to improve with age, and the adults were significantly more accurate than the children (Mann-Whitney U Test, $z = 1.79$, $p < .03$). Hence, performance should correlate with syllogistic ability.

Working memory capacity. Table 6 shows a clear trend in performance with age, which is highly reliable (Jonckheere Trend Test, $z = 5.39$, $p < .00001$). This improvement corroborates other findings in the literature (e.g., Baddeley, 1986; Case, 1985; Hitch & Halliday, 1983). It implies that performance should correlate with syllogistic ability.

Results relating syllogistic performance and the five tasks

We carried out a set of stepwise multiple regressions (SPSS-PC) on the syllogistic data and the results from the five tasks. The dependent variable was performance with the 28 syllogisms for the children and all 64 syllogisms for the adolescents and adults; the independent variables were the results from the five tasks. We carried out multiple regressions for both overall performance on the syllogisms and, for heuristic reasons, performance on the three main sorts of syllogisms: one model; multiple-model with valid conclusions; and multiple-model with no valid conclusions. Table 7 summarizes the results of these multiple regressions.

Overall, the ability to detect identities accounts for about 31% of the variance, and working memory capacity accounts for a further 8% of the variance. A comparable pattern occurs for the multiple-model problems with no valid conclusions, but the loading of the two tasks switches for one-model problems, where memory capacity accounts for about 16% of the variance and the detection of identities accounts for a further 6% of the variance. The detection of identities accounts for only 7% of the variance in performance with multiple-

Table 7. Results of the multiple regressions, with performance on the different sorts of syllogisms (and overall performance) as dependent variables and performance on the five cognitive tasks as independent variables

Type of syllogism	Task	R^2	Significance F
One model	Memory	.16	.001
	Memory + identities	.22	.0008
Multiple model, valid conclusions	Identities	.07	.04
Multiple model, no valid conclusions	Identities	.31	.00001
	Identities + memory	.40	.00001
Overall	Identities	.31	.00001
	Identities + memory	.39	.00001

Note. The 28 syllogisms were used for children, and all 64 syllogisms for adolescents and adults.

model problems with valid conclusions, and memory capacity is not significantly correlated with performance on these problems.

We also carried out a separate set of multiple regressions in which, in addition, the age of the subjects was treated as an independent variable. Age was always the best predictor of performance, accounting for 35% of the overall performance on syllogisms, with the detection of identities accounting for a further 6% of the variance. With age included as a factor, working memory capacity no longer produced any reliable correlations. For one-model syllogisms, age does not change the situation depicted in Table 7. Working memory maintains a highly reliable correlation with performance (Pearson's rho = .4, $p < .001$), and the detection of identities confirms its prediction value (Pearson's rho = .39, $p < .01$). Once again, the multiple-model problems with valid conclusions produced results for which there were few significant correlations. In fact, the only variable to emerge from the multiple regression was age, which accounted for 30% of the variance.

Syllogistic reasoning in 7-year-old children

In our study of the component tasks, the 9-year-old children performed at almost the same level as adults in interpreting quantifiers, searching for counterexamples, and making referential integrations. It therefore seemed plausible that still younger children might be able to reason syllogistically, notwithstanding the claims by Piaget and his colleagues that syllogistic inference is impossible until children acquire formal operations at about 12 years of age (e.g., Inhelder & Piaget, 1964). To test whether young children could, in fact, make syllogistic inferences at a level better than chance, we gave twenty 7-year-old children five pairs of syllogistic premises. The subjects came from the same general population as those in our main study. The syllogisms were presented in the guise of a fairy story: The prince had to answer five questions to rescue the princess. Three of the problems were simple one-model problems with valid conclusions in the A-B, B-C figure; their premises were in the following moods: AA, IA, and AE, which required conclusions in the A, I, and E moods, respectively. Two further problems had no valid conclusion interrelating the end terms: The premises were in the II and EE moods in figure A-B C-B.

Six of the 20 subjects were unable to carry out the task: In fact, 4 of them responded "no valid conclusion" to all the problems. The remaining 14 subjects demonstrated the ability to draw syllogistic conclusions. Overall, there were 43% correct conclusions to the one-model problems, and 45% correct responses to the problems with no valid conclusion. We can make a conservative estimate that the probability of guessing the correct response is .2 (there are four possible moods, and the "no valid conclusion" response). The estimate is conservative because it makes no allowance for guessing the appropriate terms in their correct places in the conclusion. Hence, the subjects were clearly performing reliably better than chance. Their errors with the invalid problems were almost always conclusions based on one model of the premises—a further 43% of their responses consisted of such errors. We conclude that some 7-year-old children are capable of simple syllogistic reasoning.

DISCUSSION

The first goal of our research was to show that model theory of syllogistic reasoning extends from adults to both adolescents and children. Our study has indeed shown that the same pattern of results occurs in all three age groups. The adults' performance corroborated the results of previous studies. They made 44% correct responses overall, which is well within the range reported in previous studies (e.g., Johnson-Laird & Bara, 1984; Johnson-Laird & Steadman, 1978). The adolescents did not perform quite so well, and children performed less well than adolescents. Nevertheless, the performance of the children, who ranged from 9 to 10 years, was well above chance. They spontaneously drew valid conclusions to 66% of the one-model problems, where the chance of guessing the correct response is conservatively .2. Hence, children are able to make certain syllogistic inferences long before the proposed Piagetian stage of "formal operations."

The improvement in syllogistic ability from children to adults shows a trend toward a greater improvement for multiple-model problems with valid conclusions than for one-model problems. In fact, children are remarkably competent with one-model problems.

If the model theory is on the right lines, then the ability to make syllogistic inferences depends, as we argued earlier, on the following component skills:

1. Parsing sentences and using lexical and compositional principles of semantics to assemble a propositional representation of the premises. Syllogisms, in particular, call for a knowledge of the meanings of quantifiers.
2. The use of propositional representations to construct models that integrate the information provided by the premises.
3. The formulation of a putative conclusion, based on the content of the models, which makes explicit relations that are not expressed by any of the premises.
4. The search for alternative models of the premises that might refute a putative conclusion.
5. The identification, where relevant, of what is common to a set of models.

Our results imply that 9-year-old children, and perhaps some 7-year-olds, already have a grasp of the meaning of quantifiers comparable to that of adolescents and adults, and so the first of these components had no detectable effect on their inferential ability. Syllogisms depend on only three terms arranged in two premises, and so the referential problems that arise with them are minimal. Reasoners have to establish the two occurrences of the middle term and to use this knowledge to form an integrated model. This step is comparable to the interpretation of simple coreference in everyday discourse, whereas our test of the second of the skills above called for much more difficult problems to be solved, such as the interpretation of two sentences with no referent in common followed by a third sentence that integrated the two previous sentences. This skill is undoubtedly an interesting psycholinguistic measure, but it seems that the children we tested perform comparably to adults, and performance does not correlate with syllogistic ability. The third of the skills above—the ability to describe the contents of models in an informative way—is again one that the children appeared to have mastered, because they spontaneously drew such conclusions (especially to one-model problems). If we had tried to measure this ability in an experimental task, we suspect that the results would again have shown neither an improvement with age nor any relation to syllogistic ability.

The most disappointing feature of the results was our failure to devise a task that reflected an improvement in the search for counterexamples. As children grow older, they undoubtedly get better with multiple-model problems—both those with valid conclusions interrelating the end terms, and those with no such valid conclusions. The capacity to respond correctly to a multiple-model problem depends on finding an alternative model of the premises that refutes a putative conclusion, and on determining what conclusion, if any, is supported by the set of models. Successful reasoners must accordingly search for alternative models. We have argued elsewhere that the origins of this ability lie in a grasp of the truth or falsity of general assertions; for example, an assertion of the form "All A are B" is falsified by a single counterexample of an A that is not a B (see Johnson-Laird, 1990). Hence, the mechanism for searching for counterexamples may derive from the system for evaluating truth values. We suspect that this mechanism may not be in place, or may not function entirely adequately, in certain individuals. Unfortunately, our task for measuring this ability failed to reveal any developmental trend. The problem, we believe, was twofold:

1. Our analysis of performance with individual syllogisms has shown that the main trend in development is an improved ability to find alternative models. However, the crux of this capacity is the ability to break existing links between end terms, and to forge such links where none exist. Young children are apt to be unable to do either of these tasks, and so to given premises, such as

Some of the A are B
Some of the B are C

they conclude:

Some of the A are C.

Unfortunately, we did not design a component task that directly assessed the ability to break existing links or to forge new links. Indeed, it is difficult to know how one could investigate this ability without testing reasoning. The essence of reasoning according to the model theory is the ability to ensure that a conclusion has no counterexamples, and in this case of syllogisms it depends on making and breaking links.

2. Our counterexample task may have been too explicitly "metalogical." In syllogisms, reasoners appear to grasp the need to search for alternatives in a tacit way; however, their introspections yield very little about the details of the process by which they have reasoned. The counterexample task, however, required subjects to search ex-

plicitly for evidence relevant to the truth or falsity of a generalization. It measured their insight into the principle that given a description, such as

musicians with wind instruments

that purports to describe individuals in a certain location, then it is necessary to know the location of individuals that do not fit the description. Their presence in the location falsifies the description.

If our diagnosis is correct, we have run into the central problems of designing tasks to measure component processes. On the one hand, if one seeks to isolate and to measure a component ability, then, inevitably, by isolating it from its normal place in a sequence of operations, one changes the ability that the subjects must display. They are now required to think in a conscious and explicit way about matters that normally lie outside awareness. On the other hand, the isolated ability may never occur by itself in any task other than a reasoning one. Hence, the component task turns out to be indistinguishable from an inferential one. Despite these difficulties, however, we cannot rule out the existence of an effective method to assess the ability to search for counterexamples by breaking and making alternative links in models. Such a method, we believe, would show a significant developmental trend, and a reliable correlation with syllogistic performance.

The task for measuring the fifth component—the detection of identities in different models—revealed a definite improvement in ability with age; it accounted for a significant proportion of the variance in syllogistic ability, and it did so even when age was taken into account in the multiple regression. Despite this success, we are not entirely convinced that the task measured the ability of subjects to establish the common features of models of syllogistic premises. Once again, the experiment calls for the explicit performance of a task that is rather different from the one that is needed in syllogisms. The perceptual task requires subjects to find identical shapes in complex figures, and the main improvement in age occurred with those figures in which the shapes are "hidden" by other features (e.g., one figure contains a pair of spectacles, and the other contains a lorry that has two wheels with the same geometrical shape). In syllogisms, however, reasoners have to appreciate that alternative models of the same premises support a common relation. Another relevant factor is the reliable correlation between the perception of identities and the capacity of working memory (Pearson's rho = .39, $p < .01$). Hence, although the detection of identities may measure an underlying component of syllogistic inference, it could be that both skills depend on another more fundamental process, namely, the processing capacity of working memory. A problem for the future is to explore the theoretical components of the perceptual task to determine whether there is any other elementary process likely to be common to syllogistic reasoning.

All five of the component skills require a working memory. If the information in such a memory fades rapidly, or if the capacity of such a memory is so small that, say, it is impossible to hold more than one syllogistic model in it at any time, then syllogistic reasoning will be very difficult. Our measure of working memory was simple and traditional, but it revealed a reliable improvement with age, and it accounted for a significant proportion of the variance in syllogistic performance over and above what was accounted for by the ability to detect perceptual identities. In general, we are confident that the pattern of results in Table 7 bears out the principle that as working memory improves—for whatever reason—it enables deductive reasoning to improve too. We have only one qualm about this claim, namely, the results for multiple-model problems with valid conclusions. None of the variables that we have examined, including working memory and age, account for much of the improvement in performance with this sort of syllogism. Older subjects do better with them, though their performance is never outstanding, but we have little idea about the cause of the improvement. These syllogisms are truly difficult and some of them are beyond the competence of nearly all logically untrained individuals. Yet, our adult subjects drew correct conclusions to them about one third of the time. The a priori probability of guessing a correct conclusion for these problems is approximately one in nine (only one of the eight sorts of conclusions All the A are C, All the C are A, etc., is correct, and the response "no valid conclusion" might also be guessed), and so this level of performance is clearly above chance on a binomial distribution. Beyond the detection of identities, which accounted for 7% of the variance, we do not know what accounts for our results. It is surprising that working memory capacity has no predictive power in this case.

Our findings corroborate the model theory of syllogistic reasoning. Every single subject performed better with one-model syllogisms than with multiple-model syllogisms. Indeed, we have never tested any subject in any experiment who has violated this law. Our findings also showed the usual effects of figure. Neither of these effects has been previously demonstrated with young subjects. There was a clear improvement in syllogistic reasoning as a function of age, and we have had some success in determining the likely causes of this improvement. Two of the component tasks that the subjects carried out provided measures that are relevant to this improvement: the detection of perceptual identities, which may reflect an ability to compare models of syllogistic premises; and working memory capacity, which seems central to most of the component processes of reasoning.

Three of our component tasks showed neither an improvement with age nor a relation to syllogistic reasoning. They failed because even the youngest of our subjects performed on these tasks at about the same level as adults. This phenomenon suggested to us that still younger children might be capable of syllogistic reasoning. They might have some knowledge of the meaning of quantifiers, the competence to use coreference to integrate premises, and the ability to describe informative relations in such models. We suspected that the weight of Piagetian tradition had prevented scholars from investigating this possibility. Indeed, we were able to show that some 7-year-old children *are* able to make one-model syllogistic inferences. Some of the component skills that we have explored in the present study may predict syllogistic ability among such younger children.

Appendix. Responses to syllogisms

The 64 pairs of syllogistic premises are shown, where each cell in Tables A1–A4 corresponds to a particular pair of premises and presents the frequencies of each sort of response that was made by at least 2 subjects. The correct responses are printed in capital letters. The data in the left-hand columns are the numbers of children (of 20) giving each sort of response, and these data are only for the 28 syllogisms that the children received; the data in the middle columns are for the adolescents (of 20); and the data in the right-hand columns are for the adults (of 20). For those 36 syllogisms presented only to adolescents and adults, the left-hand columns are the data for adolescents and the right-hand columns are the data for adults. Each cell also shows whether a problem is a one-model syllogism (as shown by a single square in the cell), a multiple-model syllogism with a valid conclusion interrelating the end terms (two squares in the cell), or a multiple-model syllogism with no valid conclusion interrelating the end terms (two squares shaded with black and white). NVC indicates a correct response; Nvc indicates an incorrect response.

Table A1. Responses given by children, adolescents, and adults to syllogisms of Figure AB-BC

Second premise	First premise				
A	A	☐			
		ALL A AREC	19	10	19
		SOME A ARE C		5	
		No A are C		2	
	I	☐			
		SOME A AREC	15	11	19
		SOME C ARE A	1		1
		All A are C	2	9	
	E	☐☐			
		No A are C		16	16
		No C are A		2	2
		SOME C NOT A		-	-
		Nvc			2
	O	◩◩			
		Some A not C		7	7
		NVC			3
		Some A are C		11	7
I	A	◩◩			
		All A are C	12	2	1
		Some A are C	4	14	16
		Some C are A	2	4	1
		NVC			2
	I	◩◩			
		Some A are C	19	17	17
		NVC		2	3
	E	☐☐			
		No A are C		14	15
		No C are A			3
		SOME C NOT A		1	1
		Some A are C		3	
	O	◩◩			
		Some A not C		7	5
		NVC			6
		Some A are C		11	8
E	A	☐			
		NO C ARE A	7	3	1
		NO A ARE C		16	17
		ALL A ARE NOT C	1		
		All A are C	9		
	I	☐☐			
		No A are C	8	8	6
		No C are A	1	2	2
		SOME A NOT C	3	6	5
		Some A are C	6	3	5
		Nvc			2
	E	◩◩			
		No A are C		19	12
		NVC		1	8
	O	◩◩			
		No A are C		3	5
		Some A not C		3	1
		NVC			8
		Some A are C		12	4
O	A	◩◩			
		No A are C		2	
		Some A not C		9	14
		Some C not A		2	
		NVC			1
		Some A are C		4	4
		Some C are A		2	1
	I	◩◩			
		Some A not C		9	15
		NVC		5	3
		Some A are C		5	2
	E	◩◩			
		No A are C	10	15	9
		Some A not C	5	2	
		NVC			10
		Some A are C	2	1	
		No A are (not) C	2	1	
	O	◩◩			
		Some A not C		14	13
		NVC			6
		Some A are C		4	1

Table A2. Responses given by children, adolescents, and adults to syllogisms of Figures BA-CB

	First premise															
Second premise	A				I				E				O			
				☐				◩◩				☐				◩◩
									NO C ARE A	3	2	10				
					All C are A	6	2		ALL C NOT A	2	1		No C are A		1	2
A	ALL C ARE A	12	6	9	Some C are A	3	4	12	NO A ARE C	5	14	7	Some C not A		3	10
	SOME A ARE C		1		Some A are C	8	12	7	All C are A	5	1	1	Some A not C		6	2
	Some C are A			2	NVC			1	All A are C	2	1		NVC			3
	All A are C	6	12	8	All A are C	2	1		Nvc		1	2	Some A are C		9	2
				☐				◩◩				☐☐				◩◩
I													No C are A			2
									No C are A	2		1	No A are C		2	
									No A are C	11	7	7	Some C not A		4	8
	SOME C ARE A	7	8	13					SOME C NOT A		4	3	Some A not C		3	
	SOME A ARE C	6	9	5	Some C are A	7	4	12	Some C are A	5	5	4	NVC		3	5
	All A are C	5	2		Some A are C	12	10	7	Some A are C		3		Some C are A		1	4
	Nvc			2	NVC		4	1	Nvc		1	5	Some A are C		7	1
				☐☐				☐☐				◩◩				◩◩
													No C are A	10	3	8
					No C are A		4	8					No A are C	3	8	2
	No C are A		8	13	No A are C		7	2					Some A not C	1	2	
E	No A are C		11	4	SOME A NOT C		5	2	No C are A		5	3	NVC			7
	SOME A NOT C		1		Some A are C		4	1	No A are C		13	9	Some C are A	2	1	2
	Nvc			2	Nvc			7	NVC		1	8	Some A are C	3	5	
				◩◩				◩◩				◩◩				◩◩
	No A are C		3	1												
	Some C not A		5	7	Some C not A		2	5					Some C not A		4	7
O	Some A not C		2	2	Some A not C		5	2	No A are C		8	1	Some A not C		6	4
	NVC			4	NVC		4	10	Some C not A		1	2	NVC		2	6
	Some C are A		4	5	Some C are A		2	3	NVC		2	12	Some C are A		1	3
	Some A are C		4	1	Some A are C		5		Some C are A		7	4	Some A are C		6	

Table A3. Responses given by children, adolescents, and adults to syllogisms of Figure AB-CB

Second premise	First premise				
A	A	◩◩			
		All A are C		16	8
		NVC		2	10
	I	◩◩			
		Some A are C		13	7
		Some C are A		1	2
		NVC		1	11
	E	□			
		NO A ARE C	13	17	12
		NO C ARE A	1		1
		Nvc	2		7
	O	□□			
		All C not A	2		
		SOME A NOT C	6	9	5
		Nvc	4	1	9
		Some A are C	4	7	5
I	A	◩◩			
		All A are C	8	1	
		Some A are C	3	6	2
		Some C are A	4	10	6
		NVC	3	3	12
	I	◩◩			
		Some A are C	9	8	
		NVC	7	11	13
	E	□□			
		No A are C		11	6
		No C are A		1	3
		SOME C NOT A		2	2
		Some A are C		3	
		Nvc		2	8
	O	◩◩			
		Some A not C		6	3
		Some C not A		1	2
		NVC		5	13
		Some A are C		5	2
		Some C are A		2	
E	A	□			
		NO A ARE C	7	7	5
		ALL A NOT C	6		
		NO C ARE A	3	11	10
		Nvc	2		5
	I	□□			
		No A are C	3	5	2
		No C are A	2	7	5
		SOME A NOT C	3	3	3
		Some A are C	6	1	3
		Nvc	3	4	7
	E	◩◩			
		No A are C		8	2
		NVC		10	18
	O	◩◩			
		No A are C		5	1
		No C are A		2	6
		Some A not C		7	1
		NVC		1	12
		Some A are C		3	
O	A	□□			
		No A are C		2	
		All A not C	8		
		Some A not C		7	2
		SOME C NOT A	2	4	7
		Nvc	2		4
		Some A are C	2	4	1
		Some C are A	3	2	3
		All A are C	2	1	
	I	◩◩			
		Some A not C		2	2
		Some C not A		3	
		NVC		10	16
		Some A are C		4	1
	E	◩◩			
		No A are C		8	5
		Some A not C		3	
		Some C not A		5	
		NVC		2	13
		Some C are A			2
	O	◩◩			
		Some A not C		7	4
		NVC		8	15
		Some A are C		3	1

Table A4. Responses given by children, adolescents, and adults to syllogisms of Figure BA-BC

Second Premise	First premise															
	A				I				E				O			
A	☐☐				☐				☐☐				☐☐			
													All C not A	3		1
									No A are C		15	10	Some A not C	4	3	1
	SOME A ARE C		1	3					No C are A		3	4	SOME C NOT A	1	4	11
	Nvc	4	5	8					SOME C NOT A			1	Some A are C	7	9	1
	All A are C	12	14	9	SOME A ARE C	14	15	8	Some A are C		2		Some C are A	2	2	1
	All C are A	3			SOME C ARE A	1	1	9	Nvc			3	Nvc			4
I	☐				◪◪				☐☐				◪◪			
									No A are C	12	13	6				
									No C are A	3		1				
									SOME A NOT C		1	2	Some A not C		4	1
	SOME A ARE C	6	13	14	Some A are C	12	11	6	Some A are C	1	2		Some C not A		2	3
	SOME C ARE A	3	2	5	Some C are A	3		3	Some C are A	3	1	1	NVC			13
	All A are C	7	1		NVC	3	8	11	Nvc	1	1	10	Some A are C		13	2
E	☐☐				☐☐				◪◪				◪◪			
													No A are C		6	1
	No A are C		12	13	No A are C		5	1					No C are A		1	2
	No C are A		2	2	No C are A		4	4					Some A not C		1	2
	SOME A NOT C			1	SOME A NOT C		1	4					Some C not A			2
	All A are C		3	1	Some A are C		7	2	No A are C		10	4	NVC		2	12
	Nvc		1	3	Nvc		1	9	NVC		9	14	Some A are C		4	2
O	☐☐				◪◪				◪◪				◪◪			
	SOME A NOT C	3	7	14												
	Some C not A	2	2						No A are C		7	2				
	Some A are C	4	4	2					Some A not C		2	1				
	Some C are A	2	1		Some A not C		8	9	Some C not A		2		Some A not C		9	5
	Nvc			3	NVC		6	9	NVC		5	16	NVC		8	13
	All A not C	7	2		Some A are C		6	1	Some A are C		2		Some A are C		2	

Notes

Bruno G. Bara and Monica Bucciarelli were supported in this research by the Italian National Research Council (Consiglio Nazionale delle Ricerche, bilateral project Italia-USA, 91.03815.CT08, 92.04523.CT08, 93.03663.CT08, and 94.03760.CT08). Philip N. Johnson-Laird was supported in part by the John S. McDonnell Foundation. We thank Renato Miceli and Luca Ricolfi for their statistical advice, and Martin Braine and Donelson Dulany for their critical review of earlier versions of this article. We are grateful to Davide Cipollone for editorial assistance.

Correspondence concerning this article should be addressed to Bruno G. Bara, Centro di Scienza Cognitiva, Università di Torino, via Lagrange, 3, 10123 Torino, Italia. E-mail: bara@psych.unito.it. Received for publication November 23, 1992; revision received January 26, 1994.

1. A syllogism consists of two premises and a conclusion, which each occur in one of four "moods" shown here with their customary mnemonics:

All A are B	(A: a universal affirmative premise)
Some A are B	(I: a particular affirmative premise)
No A are B	(E: a universal negative premise)
Some A are not B	(O: a particular negative premise)

To support a valid conclusion, the two premises must share a common term (the so-called "middle" term), and hence the premises can have four different arrangements (or "figures") of their terms:

1. A-B B-C 2. B-A C-B 3. A-B C-B 4. B-A B-C

2. The experiment was conducted with native speakers of Italian. The contents of the syllogisms were devised so as to minimize semantic relations between the terms within each premise pair while retaining plausibility for any possible conclusion, valid or invalid. This end was achieved by choosing occupations for the two end terms in each problem and an interest or preoccupation for the middle term. For example:

Ogni soldato e' ciclista (All the soldiers are cyclists)
Qualche ciclista e' facchino (Some of the cyclists are porters)
and
Qualche sportivo non e' insegnante (Some of the joggers are not teachers)
Nessun veterinario e' sportivo (None of the veterinaries is a jogger).

Care was taken to ensure that all the words in the premises were familiar to the children.

References

Aristotle. (1968). *Analytica priora et posteriora* [Prior and posterior analytics] (W. D. Ross, Ed.). Oxford: Clarendon Press. (Original work written about 330 B.C.)

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

Braine, M. D. S., & Rumain, B. (1983). Logical reasoning. In J. H. Flavell, & E. M. Markman (Eds.), *Carmichael's handbook of child psychology: Cognitive development* (4th ed., Vol. 3). New York: Wiley.

Carey, S. (1985). Are children fundamentally different kinds of thinkers and learners than adults? In S. F. Chipman, J. W. Segal, & R. Glaser (Eds.), *Thinking and learning skills: Research and open questions* (Vol. 2, pp. 485–517). Hillsdale, NJ: Erlbaum.

Case, R. (1985). A developmentally based approach to the problem of instructional design. In S. F. Chipman, J. W. Segal, & R. Glaser (Eds.), *Thinking and learning skills: Research and open questions* (Vol. 2). Hillsdale, NJ: Erlbaum.

Ehrlich, K., & Johnson-Laird, P. N. (1982). Spatial descriptions and referential continuity. *Journal of Verbal Learning and Verbal Behavior, 21,* 296–306.

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

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

Fodor, J. A. (1980). Fixation of belief and concept acquisition. In M. Piattelli-Palmarini (Ed.), *Language and learning: The debate between Jean Piaget and Noam Chomsky.* London: Routledge & Kegan Paul.

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

Hamill, J. F., (1990). *Ethno-logic: The anthropology of human reasoning.* Urbana, IL: University of Illinois Press.

Henle, M. (1978). Foreword. In R. Revlin, & R. E. Mayer (Eds.), *Human reasoning.* Washington, DC: Winston.

Hitch, G. J., & Halliday, M. S. (1983). Working memory in children. *Philosophical Transactions of the Royal Society of London, Ser. B, 302,* 325–340.

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

Inhelder, B., & 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, MA.: Harvard University Press.

Johnson-Laird, P. N. (1990). The development of reasoning ability. In G. Butterworth, & P. Bryant (Eds.), *Causes of development: Interdisciplinary perspectives* (pp. 85–110). New York: Harvester Wheatsheaf.

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

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

Johnson-Laird, P. N., Oakhill, J. V., & Bull, D. (1986). Children's syllogistic reasoning. *Quarterly Journal of Experimental Psychology, 38A,* 35–38.

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

Newell, A. (1990). *Unified theories of cognition.* Cambridge, MA: Harvard University Press.

Oakhill, J. V., & Johnson-Laird, P. N. (1985). Rationality, memory and the search for counterexamples. *Cognition, 20,* 79–94.

Polk, T. A., & Newell, A. (1988). Modeling human syllogistic reasoning in Soar. In *Tenth annual conference of the Cognitive Science Society* (pp. 181–187). Hillsdale, NJ: Erlbaum.

Sells, S. B. (1936). The atmosphere effect: An experimental study of reasoning. *Archives of Psychology, 29,* 3–72.

Stenning, K., & Oberlander, J. (1991). *A cognitive theory of graphical and linguistic reasoning: Logic and implementation* (Res. Rep. No. 20). Edinburgh, Scotland: University of Edinburgh, Human Communication Research Centre.

Sternberg, R. J. (1985). *Beyond I.Q.: A triarchic theory of human intelligence.* New York: Cambridge University Press.

Wason, P. C., & Johnson-Laird, P. N. (1972). *Psychology of reasoning: Structure and content.* Cambridge, MA: Harvard University Press.

Wetherick, N., & Gilhooly, K. (1990). The figural effect in syllogistic reasoning. In K. Gilhooly, M. T. G. Keane, R. Logie, & G. Erdos (Eds.), *Lines of thought: Reflections on the psychology of thinking* (Vol. 1). London: Wiley.

Wilkins, M. C. (1928). The effect of changed material on the ability to do formal syllogistic reasoning. *Archives of Psychology, 16,* 102.

Woodworth, R. S., & Sells, S. B. (1935). An atmosphere effect in formal syllogistic reasoning. *Journal of Experimental Psychology, 18,* 451–460.