Journal of Youth and Adolescence, Vol. 18, No. 1, 1989

Correlates of Syllogistic Reasoning Skills in Middle Childhood and Early Adolescence

Kathleen M. Galotti[1] and Lloyd K. Komatsu[2]

Received May 8, 1987; accepted May 25, 1988

Sixth, eighth, tenth, and twelfth graders, and college students, were given a preliminary test of categorical syllogistic reasoning ability. In a subsequent session, subjects were given other categorical syllogisms and asked to depict as many of the possible relationships between the A, B, and C terms of the syllogism as they could. The number of possible relationships, and the time it took to decide if other relationships were possible, did not differ among the noncollege groups. The results indicated, however, that the correlates of reasoning proficiency differed for those subjects younger and those older than about age 13.

INTRODUCTION

Reasoning with categorical syllogisms is regarded as a skill central to thinking and intelligence (e.g., Guilford, 1959; Thurstone, 1938). It is one of the most frequently investigated, at least with adult subjects (e.g., Erickson, 1978; Galotti *et al.*, 1986; Guyote and Sternberg, 1981; Johnson-Laird, 1982,

Support for this research was provided by the Spencer Foundation, whose generosity is greatfully acknowledged.

[1]Assistant Professor of Psychology at Carleton College. Received the Ph.D. in Psychology and the M.S.E. in Computer Science from the University of Pennsylvania, and has research interests in the areas of formal and everyday reasoning, the development of reasoning skills, decision making, and conceptual development. To whom reprint requests should be addressed at Department of Psychology, Carleton College, One North College Street, Northfield, Minnesota 55057.

[2]Assistant Professor of Psychology at Carleton College. Received the Ph.D. in Psychology from the University of Pennsylvania, and has research interests in the areas of language acquisition, conceptual development, and the development of reasoning skills.

0047-2891/89/0200-0085$06.00/0

1983; Johnson-Laird and Bara, 1984; Revlin and Leirer, 1978; Revlis, 1975a, b). Typically, subjects are presented with two premises of the form "Some A are B, All B are C." They are then asked to supply a conclusion relating the A and the C terms that is always true, given the premises, or to decide whether a particular conclusion (for example, "Some A are C") is always true. Even when the task and the meaning of logical operators and quantifiers are carefully explained, subjects often make errors on this task (Begg and Denny, 1969; Begg and Harris, 1982; Ceraso and Provitera, 1971; Chapman and Chapman, 1959; Dickstein, 1975, 1976).

Several models have been proposed to account for these findings. Galotti *et al.* (1986) suggested that these models can be divided into two classes, and found evidence that both types of models are needed to explain individual differences among college students. Models of the "deduction rules" class explain reasoning with syllogisms in terms of an explicit or implicit reliance on deduction rules (cf. Braine and Rumain, 1983; Osherson, 1975; Rips, 1983). For example, when given the premises "Some A are B, Some B are C," a subject might use the "two somes" rule: if both premises contain the quantifier "some," then there is no logically valid conclusion. Errors in reasoning arise if a subject lacks knowledge of a relevant rule or set of rules, or possesses the relevant knowledge but for some reason fails to apply it.

On the other hand, models of the "mental models" class (e.g., Johnson-Laird, 1982, 1983) attribute poor syllogistic reasoning to the failure to consider the different relationships that may hold between the set of A things, the set of B things, and the set of C things mentioned in the premises. In brief, these models propose that a reasoner who considers fewer of the possible interpretations and combinations of the premises will perform less well than a reasoner who considers more.

For ease of reference, we refer to a possible combination of the A, B, and C terms of a categorical syllogism as an *instance*. An instance, therefore, is one possibility allowed by two premises. The different "mental models" accounts of syllogistic reasoning performance propose different ways of representing premises and instances. One influential proposal was that of Erickson (1978), who assumed that the representation is isomorphic to Euler diagrams. Figure 1 presents an example of this form of representation. It shows Euler diagram depictions of possible interpretations of single premises in the top panel, and possible instances in the bottom panel.

A "mental models" description links syllogistic reasoning to other thinking and reasoning tasks in adults, such as tests of creativity or hypothesis construction, in which the generation and evaluation of possibilities or instances have been seen as fundamental processes (Perkins, 1981; Wason, 1960, 1977). It also links syllogistic reasoning to theories of cognitive development. The systematic consideration of possibilities is often identified as a major intellectual achievement in the thought processes of adults and preadolescent children (Inhelder and Piaget, 1964; Keating, 1980).

Because the consideration of alternative possibilities is thought to be central to thinking, because of the significance of systemic instance generation in cognitive developmental theory, and because the categorical syllogistic reasoning task is useful for investigating the generation of alternatives, the study of categorical syllogistic reasoning by children and adolescents is important. Few such studies have been performed, however, and none have directly examined the role of generation of instances in the task (Bickersteth and Das, 1981; Lane, 1983; Lane *et al.,* 1983; Linn, 1983).

The goal of the present study was to examine the correlates of proficiency at the syllogistic reasoning task in subjects of different ages. The question asked, therefore, was, would younger "good reasoners" enumerate more of the alternative possibilities allowed by the premises of a categorical syllogism than age-equivalent "poor reasoners," just as adult good reasoners appear to enumerate more alternatives than adult poor reasoners? If younger subjects perform differently from older subjects, what are the differences and when do they emerge?

METHOD

Subjects

Twenty-five sixth graders (mean age = 12.2 years), 22 eighth graders (mean age = 14.1 years), 23 tenth graders (mean age = 16.1 years), 21 twelfth

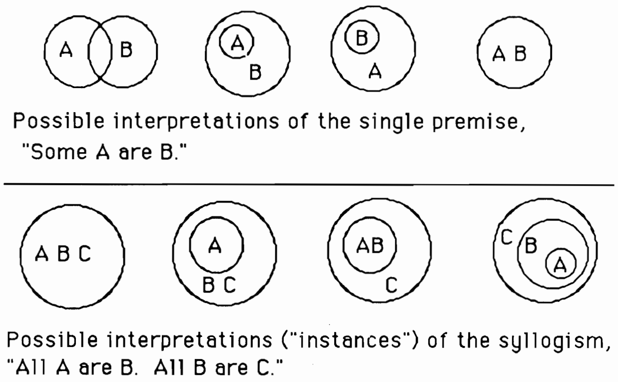

Fig. 1. Examples of Euler diagram depictions of single interpretations (top panel) and instances (bottom panel).

graders (mean age = 18.0 years), and 22 college students (mean age = 20.4 years) participated. One additional sixth grader was dropped from the study for failure to cooperate. There were 11 sixth-grade boys, 15 eighth-grade boys, 12 tenth-grade boys, 7 twelfth-grade boys, and 13 males in the college sample. College students were enrolled at Carleton College; all other subjects were students in Northfield Public Schools.

Apparatus

The training, experimental, and recall problems were presented on an Apple Macintosh 128K computer programmed to present the problems, collect responses, and record response times.

Materials and Procedure

There were four parts to the experimental prodedure. First, subjects' reasoning ability with categorical syllogisms was assessed in a preliminary paper and pencil test. Next, in the training phase, subjects were trained in the possible interpretation of single premises, depicted with Euler diagrams. In brief, they were shown in this phase how to draw diagrams similar to those in the top panel of Fig. 1. We call such diagrams, which consist of either one or two circles carrying two labels, *premise interpretation diagrams.* We distinguish these from *instance diagrams,* which depict possible combinations of two premises, and consist of one, two, or three circles with three labels, similar to those in the bottom panel of Fig. 1.

The third part of the procedure was the experimental phase proper. Subjects read a pair of premises and drew as many instance diagrams as they could think of that were consistent with the premises. Finally, the fourth part of the procedure was the recall phase. It tested the subjects' recall and speed of drawing single-premise interpretation diagrams.

Preliminary Test

The test of categorical syllogistic reasoning proficiency was a paper and pencil task used previously with college students (Galotti *et al.,* 1986). It consisted of eight categorical syllogisms in the form A-B, C-B, using familiar, concrete terms (e.g., "All green figures are striped. No large figures are striped"). The subjects' task was to provide, if possible, a statement relating the A and C term that was always true, assuming the premises given were true. If no statement were always true, the problem was to be marked with an X. Three of the problems had no logically necessary conclusion.

The precollege subjects were given the preliminary test in one of their regularly scheduled classes, after a brief discussion of categorical syllogisms and logically necessary conclusions. College students filled out the preliminary test individually, after reading through a written introduction that covered the same information presented orally to the precollege subjects.

Experimental Session

The precollege students who had parental consent and the college student volunteers were run in groups of one to three in an experimental session that lasted 50 minutes. The training phase, the experimental phase proper, and the recall phase were all conducted during that session.

In the training phase, subjects were presented with one-sentence premises that used familiar, concrete terms with one of the four standard quantifiers for categorical syllogisms: *all, no, some,* and *some not* (e.g., "Some blue blocks are not small"). Subjects were then shown how the different interpretations of each single premise could be drawn on the computer screen using Euler diagrams. When a subject was able to draw every possible interpretation of each of the four premises, the experimental phase proper began.

The experimental phase was similar to the training phase, except for the following: (1) The subjects were given two premises to work with instead of one, and were told that their diagrams (of instances) would have one, two, or three circles instead of one or two as in the training session. (2) The subjects were not told ahead of time how many different interpretations each problem could have, but were asked to draw as many as they could think of. (3) The subjects were asked to work as quickly as possible, because the computer would be timing them.

The computer presented one problem at a time, recording each instance diagram drawn, the amount of time the subject spent drawing it, and the amount of time it took the subject to decide whether another instance diagram existed. The subjects had the opportunity to work on up to six experimental problems. All were in the form A-B, B-C, and used familiar, concrete terms different from those used in the preliminary test or training phase (e.g., "All astronomy books are grey. No grey books are heavy"). Subjects were allowed to complete as many problems as possible while leaving the last 10 minutes of the session for the recall problems.

The recall phase was similar to the training session in that only one premise (using a different set of familiar, concrete terms) was presented at a time. As in the experimental phase, subjects were not told how many alternative interpretations each premise had, and were asked to work as quickly as possible. Subjects' performance on this task gave us a measure of their fluency with the mechanics of the experimental procedure (use of the computer to draw Euler diagrams, recall of all possible interpretations of the premises) apart from the process of generating instances. The session ended when the subject completed the four recall problems, or at the end of 50 minutes.

RESULTS

Measures

In addition to the scores on the preliminary test of reasoning ability, four aspects of performance were obtained: (1) mean log proportion of the total possible instance diagrams generated on the experimental problems (henceforth INSTANCES); (2) mean log proportion of possible premise interpretation diagrams generated on recall problems (henceforth RECALL); (3) mean log time, in seconds, to decide whether additional instance diagrams were possible for a given experimental problem (henceforth DECTIME); and (4) mean log time to draw interpretation diagrams of single premises on recall problems (henceforth DRAW TIME). This last measure was included to take account of differential familiarity and agility with using Macintosh computers and drawing Euler diagrams—in other words, with the mechanics of responding. Logarithmic transformations were used to correct for the non-normality of both reaction time distributions and distributions of proportions.

Age and Sex Effects

The four measures described above along with scores on the preliminary test were subjected to 5 (grade) × 2 (sex) analyses of variance (ANOVAs). None of the ANOVAs revealed any main effect for sex. Significant effects of grade were found for preliminary test score ($F[4, 103] = 12.55$, $p < .001$), mean log proportion of possible instance diagrams generated on experimental problems (INSTANCES; $F[4, 102] = 9.04$, $p < .001$), and mean log proportion of premise interpretation diagrams generated on the recall problems (RECALL; $F[4, 99] = 16.65$, $p < .001$). Table I reports the means of the five measures by grade.

Specific comparisons revealed that for the measures of preliminary test score and INSTANCES, college students had significantly higher scores than all other groups ($p < .01$); no other differences were reliable. For RECALL, college students had significantly higher scores than all other groups ($p < .05$), and sixth graders had significantly lower scores than eighth and tenth graders and college students ($p < .01$).

Table I. Means and Standard Deviations of Dependent Measures, by Age and Sex

Measure[a]	Grade 6	8	10	12	College
Preliminary test (out of 8)	1.72(1.37)	2.86(1.55)	3.09(1.44)	2.05(1.60)	4.73(1.55)
INSTANCES	−1.73(.69)	−1.27(.32)	−1.41(.61)	−1.65(.97)	−.72(.35)
RECALL	−.48(.28)	−.22(.16)	−.26(.14)	−.37(.25)	−.07(.08)
DEC TIME	.85(.33)	.77(.27)	.71(.23)	.70(.20)	.91(.30)
DRAW TIME	1.46(.15)	1.36(.14)	1.41(.12)	1.37(.19)	1.37(.13)

[a]INSTANCES: mean log proportion of possible instance diagrams generated. RECALL: mean log proportion of premise interpretation diagrams generated on recall problems. DEC TIME: mean log time (seconds) to decide if additional instance diagrams could be generated. DRAW TIME: mean log time (seconds) to draw premise interpretation diagrams on recall problems.

Significant interactions between sex and grade emerged for RECALL ($F[4, 99] = 5.241, p < .05$). Specific comparisons showed that sixth-grade males recalled significantly fewer premise interpretations than did any other group of males ($p < .01$), while groups of females did not differ. No clear interpretation of this interaction is apparent.

Individual Differences Within Different Age Groups

In general, the data above suggest few age-related differences, save for those between college aged vs. younger students. That difference, however, can plausibly be attributed to a selection effect. To assess actual developmental differences, it is necessary to ask another question: Do the correlates of superior performance on a syllogistic reasoning task change with age?

The four measures described above were correlated with preliminary test score, first for all subjects, and then separately for each age group. Table II presents these correlations. The correlations between preliminary test score and INSTANCES (mean log proportion of possible instance diagrams generated) for sixth and eighth graders were significantly lower from those correlations in older age groups (Fisher's Z; $p < .01$, except for the difference between eighth and tenth graders, where $p < .05$), being moderate and negative in the former cases and moderate and positive in the latter. The correlation between preliminary test score and DRAW TIME (mean time to draw interpretation diagrams on the recall test) for eighth graders was significantly higher than the corresponding correlations for tenth graders and twelfth graders (Fisher's Z; $p < .05$).

To what extent are the significant correlations between preliminary test scores and INSTANCES a function of drawing skill, familiarity with Euler diagrams, or memory of premise interpretations? To answer this question, we calculated partial correlations between preliminary test score and IN-

Table II. Correlations of Pretest Score with Other Variables, by Age[a]

	INST	RECALL	DEC	DRAW
Overall	.40[b]	.49[b]	.07	−.28[c]
Sixth grade	−.51[c]	.23	.26	−.13
Eighth grade	−.38[d]	−.01	.01	.16
Tenth grade	.31[e]	.43[d]	−.22	−.54[d]
Twelfth grade	.71[b]	.35[e]	−.16	−.48[d]
College	.46[d]	.28	−.05	−.23

[a]INST (INSTANCES): mean log proportion of possible instance diagrams generated. RECALL: mean log proportion of premise interpretation diagrams generated on recall problems. DEC TIME: mean log time (seconds) to decide if additional instance diagrams could be generated. DRAW TIME: mean log time (seconds) to draw premise interpretation diagrams on recall problems.
[b]$p < .001$.
[c]$p < .01$.
[d]$p < .05$.
[e]$p < .10$.

STANCES, controlling for RECALL (mean log proportion of interpretation diagrams of single premises generated on recall problems, presumably in part a measure of familiarity with Euler diagrams), and controlling for DRAW TIME (mean log time to draw single premise interpretation diagrams, presumably a measure of drawing skill and proficiency with the computer). We also calculated a partial correlation controlling for both. Table III presents these correlations, and shows that the pattern of correlations across the age groups remains unchanged, even when controlling for these variables.

DISCUSSION

The lack of sex differences is not surprising, given the findings of Meehan (1984) and Johnson (1984). Johnson, for example, reports that sex differences in problem solving "[do not hold for] closed-context reasoning problems or measures of general intelligence" (p. 1359). The one interaction that emerged with sex as a factor showed no discernible pattern.

It is surprising, on the other hand, that few overall age differences were found. In particular, there were no differences among the precollege groups in the number of instance diagrams generated (INSTANCES). If the generation of possibilities is a hallmark of formal operations, and if formal operations typically emerge and consolidate during this age range (Keating, 1980), then an effect of age among these subjects should have been found.

Differences between the college and the precollege groups did exist for INSTANCES. However, that college students score significantly higher on

Table III. Partial Correlations of Preliminary Test Score with INSTANCES, Controlling for RECALL and DRAW TIME, by Age[a]

		First-order partial controlling for		Second-order partial controlling for
	Zero order	RECALL	DRAW TIME	RECALL and DRAW TIME
Overall	.40[e]	.20[c]	.36[e]	.19[c]
Sixth grade	−.51[d]	−.54[d]	−.50[d]	−.53[d]
Eighth grade	−.38[c]	−.39[b]	−.39[b]	−.39[b]
Tenth grade	.31[b]	.09	.25	−.10
Twelfth grade	.71[e]	.65[d]	.69[d]	.66[d]
College	.46[c]	.38[b]	.64[d]	.58[d]

[a]INSTANCES: mean log proportion of possible instance diagrams generated. RECALL: mean log proportion of premise interpretation diagrams generated on recall problems. DRAW TIME: mean log time (seconds) to draw premise interpretation diagrams on recall problems.
[b]$p < .10$.
[c]$p < .05$.
[d]$p < .01$.
[e]$p < .001$.

the preliminary test than do younger groups suggests that this is a selection effect rather than an age effect. Furthermore, if the college/noncollege differences represent a developmental effect, the effect takes place considerably later than Piagetian (or other current) theory would predict.

There is an age effect, however, in the correlations between preliminary test scores and the instance generation measure (INSTANCES). The proportion of total instance diagrams correlates significantly and *negatively* with preliminary test score for sixth and eighth graders, but significantly and *positively* for tenth and twelfth graders, and college students. This suggests that what makes for good syllogistic reasoning in subjects older than age 13 (namely, at least in part, the ability to consider multiple instance, as measured by the ability to generate Euler diagrams of those instances) is not what makes for good reasoning in subjects younger than age 13.

Notice that our claim is not that our younger subjects were unable to reason with categorical syllogisms. The performance of the eighth graders in particular suggests that they are able to successfully solve such problems. It appears, however, that preadolescent syllogistic reasoning does not involve instance generation. One possibility is that our younger subjects approached the task in a different way from the way our older subjects approached it. Perhaps sixth and eighth graders first solved the syllogism, then generated instances consistent with their final conclusion, whereas older subjects approached syllogisms by first generating instances, then arriving at a conclusion. Such a proposal awaits future research. A related question is whether younger good reasoners later become the older good reasoners, or whether the strategies that prove useful for younger subjects prove maladaptive at a later period.

A second possible interpretation of our results is that the more competent younger subjects are capable of generating more instances than the less competent ones, but do not use Euler diagrams to represent instances. What changes with age, then, may be the type of representation used. Although many researchers have used Euler diagrams to represent the possible combinations of terms in syllogistic reasoning, as we did here (e.g., Erickson, 1978; Lane, 1983), others propose different representational systems (e.g., Johnson-Laird, 1982, 1983; Johnson-Laird and Bara, 1984). One important difference between Johnson-Laird's representation and Euler diagrams is that the former depicts individual members of a set, but not the set boundaries, while the latter depicts only the set boundaries, but no individual members. It may be that children at different ages find different representations more useful. A third possibility to explain our results is that none of our subjects use Euler diagrams, but that the more competent older subjects were better able to adapt to that representational system.

A fourth issue one might raise is that the different age groups had had previous differential exposure to computers, and that this somehow affected the results. We point out in response that the pattern of correlations across different age groups is relatively unchanged, even when controlling for drawing speed, presumably at least a partial measure of agility with and efficiency at using this computer. Thus, although this issue cannot be ruled out, we do not see how it provides a clear account of the results.

We believe, therefore, that the most profitable future work must be aimed at determining whether cognitive development in adolescence involves a change in representation, and/or greater flexibility in representation, or instead involves a change in process or strategy. A related question is whether the age differences seen here occur in other kinds of reasoning, either of the formal, self-contained type, or of the type called "everyday" reasoning (see Perkins, 1985, for a review). In either case, the specific differences between the "middle-aged" child and the adolescent, and how the changes occur, remain to be charted.

ACKNOWLEDGMENTS

We thank Paul Jorgensen, Diane Cirksena, Arnie Giebel, John Jarvis, and Lu Lindstrom for providing access to subjects, and Beth Wiggins, Janet Levitt, Peter Gerhardstein, and Andrea Matchett for their assistance in running the study. We also thank Jon Baron, John Bare, and an anonymous reviewer for comments on an earlier version. A preliminary report of this

study was made at the 1986 meeting of the Jean Piaget Society, Philadelphia, Pennsylvania.

REFERENCES

Begg, I., and Denny, J. P. (1969). Empirical reconcilation of atmosphere and conversion interpretations of syllogistic reasoning errors. *J. Exp. Psychol.* 81: 351-354.
Begg, I., and Harris, G. (1982). On the interpretation of syllogisms. *J. Verbal Learn. Verbal Behav.* 21, 595-620.
Bickersteth, P., and Das, J. P. (1981). Syllogistic reasoning among school children from Canada and Sierra Leone. *Int. J. Psychol.* 16: 1-11.
Braine, M. D. S., and Rumain, B. (1983). Logical reasoning. In Flavell, J., and Markman, E. (eds.), *Handbook of Child Psychology (4th ed., Vol. 3): Cognitive Development.* Wiley, New York.
Ceraso, J., and Provitera, A. (1971). Sources of error in syllogistic reasoning. *Cog. Psychol.* 2: 400-410.
Chapman, L. J., and Chapman, J. P. (1959). Atmosphere effect reexamined. *J. Exp. Psychol.* 58: 220-226.
Dickstein, L. S. (1975). Effects of instruction and premise order on errors in syllogistic reasoning. *J. Experimental Psychol. Human Learn. Memory.* 1: 376-384.
Dickstein, L. S. (1976). Differential difficulty of categorical syllogisms. *Bull. Psychonom. Society* 8: 330-332.
Erickson, J. R. (1978). Research on syllogistic reasoning. In Revlin, R., and Mayer, R. E. (eds.), *Human Reasoning.* V. H. Winston and Sons, Washington, DC.
Galotti, K. M., Baron, J., and Sabini, J. (1986). Individual differences in syllogistic reasoning: Deduction rules or mental models? *J. Exp. Psychol. Gen.* 115: 16-25.
Guilford, J. P. (1959). Three faces of intellect. *Am. Psychol.* 14: 469-479.
Guyote, M. J., and Sternberg, R. J. (1981). A transitive-chain theory of syllogistic reasoning. *Cog. Psychol.* 13: 461-525.
Inhelder, B., and Piaget, J. (1964). *The Early Growth of Logic in the Child.* Harper & Row, New York.
Johnson, E. S. (1984). Sex differences in problem solving. *J. Educ. Psychol.* 76: 1359-1371.
Johnson-Laird, P. N. (1982). Ninth Bartlett Memorial Lecture. Thinking as a skill. *Quart. J. Exp. Psychol.* 34A: 1-29.
Johnson-Laird, P. N. (1983). *Mental Models.* Harvard University Press, Cambridge, MA.
Johnson-Laird, P. N., and Bara, B. G. (1984). Syllogistic inference. *Cognition* 17: 1-61.
Keating, D. P. (1980). Thinking processes in adolescence. In Adelson, J. (ed.), *Handbook of Adolescent Psychology.* Wiley, New York.
Lane, D. S., Jr. (1983). Using observation and action instruction to facilitate conditional reasoning performance in early adolescents. *J. Early Adoles.* 3: 335-347.
Lane, D. S., Fletcher, D. N., and Fletcher, H. J. (1983). Improving conditional syllogism performance of young normal and gifted students with discovery and rule instruction. *J. Educ. Psychol.* 75: 441-449.
Linn, M. (1983). Content, context and process in reasoning during adolescence: Selecting a model. *J. Early Adoles.* 3: 63-82.
Meehan, A. (1984). A meta-analysis of sex differences in formal operational thought. *Child Develop.* 55: 1110-1124.
Osherson, D. (1975). Logic and models of logical thinking. In Falmagne, R. (ed.), *Reasoning: Representation and Process in Children and Adults.* Lawrence Erlbaum Associates, Hillsdale, NJ.
Perkins, D. N. (1981). *The Mind's Best Work.* Harvard University Press, Cambridge, MA.
Perkins, D. N. (1985). Postprimary education has little impact on informal reasoning. *J. Educ. Psychol.* 77: 562-571.

Revlin, R., and Leirer, V. O. (1978). The effect of personal biases on syllogistic reasoning: Rational decisions from personalized representations. In Revlin, R., and Mayer, R. E. (eds.), *Human Reasoning*. V. H. Winston and Sons, Washington, DC.

Revlis, R. (1957a). Syllogistic reasoning: Logical decisions from a complex data base. In Falmagne, R. (ed.), *Reasoning: Representation and Process in Children and Adults*. Lawrence Erlbaum Associates, Hillsdale, NJ.

Revlis, R. (1975b). Two models of syllogistic reasoning. *J. Verbal Learn. Verbal Behav.* 14: 180-195.

Rips, L. J. (1983). Cognitive processes in propositional reasoning. *Psychol. Rev.* 90: 38-71.

Thurston, L. L. (1938). *Psychometric Monograph* (No. 1 ed.). University of Chicago Press, Chicago, IL.

Wason, P. C. (1960). On the failure to eliminate hypotheses in a conceptual task. *Quart. J. Exp. Psychol.* 12: 129-140.

Wason, P. C. (1977). On the failure to eliminate hypotheses in a conceptual task: A second look. In Johnson-Laird, P. N., and Wason, P. C. (eds.), *Thinking: Readings in Cognitive Science*. Cambridge University Press, Cambridge.