AN ANALYSIS OF DIFFICULTIES IN ABSTRACT SYLLOGISTIC REASONING

Dissertation for the Degree of Ph. D. MICHIGAN STATE UNIVERSITY FRED HELSABECK, JR. 1973



This is to certify that the

thesis entitled AN ANALYSIS OF DIFFICULTIES IN ABSTRACT SYLLOGISTIC REASONING

presented by

Fred Helsabeck, Jr.

has been accepted towards fulfillment of the requirements for

Ph.D. degree in Psychology

Doual U. Johnson Major progessor

Date Auguel 9, 1973

O-7639



ABSTRACT

AN ANALYSIS OF DIFFICULTIES IN ABSTRACT SYLLOGISTIC REASONING

By

Fred Helsabeck, Jr.

This study analyzes student performance in solving abstract syllogisms from the perspective of instructional psychology, that is, an attempt is made to determine the initial cognitive state of the typical college student without formal logical training, and then to determine what transforming actions would be effective in causing a change to a desired final state.

An "algebraic substitution" model is proposed as being most typical of the reasoning processes of untrained college students. A given subject is hypothesized to make substitutions of letters in the statements as though they were equations, according to a "substitution hierarchy" consisting of one or more of the four standard sentence forms traditionally used in syllogisms. This model fits the pattern of response frequencies somewhat better than two other models based on Woodworth and Sell's atmosphere effect or Chapman and Chapman's invalid conversion with probabilistic inference. To get further insight into the subject's initial state as well as what transforming actions might be effective, a series of experiments were performed, testing the effect of various modifications of the original syllogism task.

The first experiment tested the effect of changing the word "some" to "at least one" to remove this ambiguity in the English language. Scores of individuals were very similar on the two forms of wording for the same individual and were highly correlated (r = .90). The small difference between groups was interpreted as sampling error. A number of subject protocols showed confirming evidence for the algebraic substitution model.

The second experiment tested the effect of using spatial wording in place of the traditional wording. This alteration produced mixed results. Some subjects made much higher scores on the spatial task while others simply made different types of errors. The deciding factor in whether or not a subject would profit from the use of drawings seemed to be whether or not all of the possibilities inherent in the statements were considered, but it was not clear whether this was due to lack of imagination or simply failing to look for these possibilities.

The third experiment forced consideration of appropriate possibilities by testing the production of counter-examples to invalid syllogisms. The relative effectiveness of three methods of producing counter-examples was tested, one using set diagrams (spatial), one using concrete objects (verbal-specific), and one using general verbal descriptions (verbal-general). Scores for the spatial and verbal-specific methods were significantly better than the verbal-general method, but the scores for all three methods were relatively low even for this reduced task, owing in large measure to the difficulty subjects had in negating statements, especially statements involving "Some are not ."

A fourth experiment removed the difficulty of forming negations of statements from the production of counterexamples by modifying the task to one in which the step of negation was already taken, leaving only the difficulty of example production. Near perfect scores resulted for both of the two methods of example generation used, but the spatial method took less time than the verbal-concrete method, even when a correction was made for writing time.

Several conclusions were drawn from these experiments that relate to the process of instruction: (1) Subjects are capable of considering various alternatives when reasoning with syllogisms but must be induced to do so. (2) Subjects have difficulty forming negations of statements and must be instructed in this. (3) Set diagrams are helpful in generating examples. A simple instructional procedure that was consistent with the above conclusions was proposed.

AN ANALYSIS OF DIFFICULTIES IN ABSTRACT

SYLLOGISTIC REASONING

By

Fred Helsabeck, Jr.

A DISSERTATION

Submitted to Michigan State University in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Department of Psychology

1973

•



I would like to thank Dr. Donald Johnson for spending many hours with me from the early exploratory work to the final draft, supplying advice and encouragement. I would also like to thank the other members of my dissertation committee, Dr. David Wessel, Dr. Richard Hart, and especially Dr. John Hunter who has had a particularly strong influence in clarifying my thinking.

* * * * * *

TABLE OF CONTENTS

Chapter		Page
I.	THE INITIAL COGNITIVE STATE	1
	Some Theoretical Hypotheses Regarding	
	the Initial State	4
	Empirical Evaluation of Models	10
	Summary	18
II.	EXPERIMENTS TESTING THE EFFECTS OF CHANGES	
	OF WORDING	20
	Tomoniment T	20
		20
		21
	Results and Discussion	23
	Experiment II	27
	Method	28
	Results and Discussion	29
	Summary ,	32
	• · ·	
III.	EXPERIMENTS TESTING ABILITY TO GENERATE	
	COUNTER-EXAMPLES TO INVALID SYLLOGISMS	33
	Experiment III	33
	Method	37
	Reculte	39
		43
		45
	Experiment IV	40
	Method	45
	Results	46
	Discussion	48
IV.	IMPLICATIONS FOR INSTRUCTIONAL PROCESS	49
		_
REFERENC	CES	53
Appendiz	ĸ	
>	LOCION DESCONTING MECH	55
Α.	LOGICAL REASONING TEST	55
_		C 1
в.	SPATIAL REASONING TEST	01
_		
С.	INVALID SYLLOGISMS	63
D.	ILLUSTRATIONS	68

LIST OF TABLES

Table			Page
1.	Comparisons of Predictions from Three Models with Data from Chapman and Chapman Study	•	11
2.	Comparison of Observed Responses with Algebraic Model Under Assumption of Equal Distribution of Three Hierarchies Among 80% of Subjects	•	14
3.	Comparison of Observed Responses with Algebraic Model Under Assumption of Equal Distribution of Four Hierarchies Among 80% of Subjects	•	15
4.	Effects of Wording and Order on Number Correct	•	23
5.	Percent of Response of Each Kind to Each Premise Combination	•	25
6.	Comparison of Spatial and Verbal-Specific Methods of Counter-Example Production	•	40
7.	Comparison of Spatial and Verbal-General Methods of Counter-Example Production	•	41
8.	Comparison of Verbal-Specific and Verbal- General Methods of Counter-Example Production	•	42
9.	Effect of Removing Conclusion-Negation Step from Spatial and Verbal-Specific Methods of Counter-Example Production	•	47
10.	Analysis of Variance for Solution Times for Spatial and Verbal-Specific Methods	•	48

CHAPTER I

THE INITIAL COGNITIVE STATE

Anyone who has taught undergraduate mathematics or logic should be well aware that the instructional process requires much more than the instructor's mastery of the subject. He must also take into consideration the psychological processes involved as the student's thinking and behavior undergoes modification to the desired state. Even though he might not think of himself in this way, he functions in a manner not too different from that of a behavior therapist or athletic coach in that he must not only know the desired thinking and behavior but also what is required to facilitate desired change in the thinking or behavior of the student, client, or athlete.

Glaser and Resnick (1972) point out the increasing interest of experimental psychologists in the instructional process and of the need for analyzing complex tasks. They contrast descriptive theories of learning with prescriptive theories of instruction and go on to describe some characteristics of the latter:

Regardless of the kind of descriptive theory with which one works, certain characteristics of prescriptive theory for the optimization of learning seem reasonable to consider. They are: (a) a description of the state of knowledge to be achieved; (b) description of the initial state with which one begins; (c) actions which can be taken, or conditions that can be implemented to transform the initial state; (d) assessment of the transformation of the state that results from each action; and (e) evaluation of the attainment of the terminal state desired (p. 208).

The above will serve as a useful framework to which to relate results of this study, an analysis of the task of solving abstract syllogisms.

Before proceeding with this analysis, it is desirable to define a few terms used to describe syllogisms. A <u>categorical proposition</u> is one of the following sentence forms:

Name	Expression	Symbol
Universal Affirmative	All are	А
Universal Negative	No are	E
Particular Affirmative	Someare	I
Particular Negative	Some are not	0

Into the blanks can be substituted names of categories. These categories, called <u>terms</u>, can be either <u>abstract</u>, such as letters of the alphabet, or <u>concrete</u> as specific nouns or phrases. A <u>syllogism</u> consists of three such propositions, the first two called <u>premises</u> and the third called the <u>conclusion</u>. For example, the following is an abstract syllogism: All A are B; Some B are C. Therefore, Some A are C.

A syllogism is said to be <u>valid</u> if any consistent replacement of its terms by specific categories that results in both premises being true will also result in the conclusion being true. For example, the above abstract syllogism is not valid because it is possible to find at least one replacement of terms that fails to meet the above requirement, namely the replacement A = cats, B = animals, and C = dogs, resulting in the concrete syllogism,

> All cats are animals. (true) Some animals are dogs. (true) Therefore, Some cats are dogs. (false)

The desired final state is the ability to reliably judge any syllogism to be valid or invalid.

The description of final state was fairly easy to describe, at least in behavioral terms. The description of the initial state is more difficult and will be discussed next.

Some Theoretical Hypotheses Regarding the Initial State

For the purposes of this study, abstract syllogisms, that is, those with letters rather than nouns as terms will be used to rule out the more complex effects produced by the subject's attitude toward the content of the material. Research into these effects have been done by Wilkins (1928), Janis and Frick (1964), Luchins and Luchins (1965), Frase (1966a,b), Dillehay, Insko, and Smith (1966), Parrott (1967, 1969), and Stratton (1970). These studies are discussed by Johnson (1972).

Even using abstract syllogisms with the suggestive effect of content neutralized, untrained subjects are remarkably prone to error. In fact, Wilkins (1928) in comparing various types of material, found symbolic material to be one of the more difficult types.

Woodworth and Sells (1935) and Sells (1936) used the concept "atmosphere" in their description of the initial state. This could be thought of as a kind of verbal set or tendency to make conclusions that are similar in verbal form to the premises. For example, the following invalid syllogism could be explained by the atmosphere effect:

> Some A are B; Some B are C. Therefore, Some A are C.

In this case, since both premises are "I" statements, there would be a tendency to conclude an "I" statement. That the above syllogism is invalid can be seen by making the replacement, A = cats, B = black things, and C = dogs, resulting in the syllogism:

> Some cats are black; (true) Some black things are dogs. (true) Therefore, Some cats are dogs. (false)

Woodworth and Sells supplement their atmosphere hypothesis with two others, an interpretation of the word "some" to mean "some but not all" instead of the logically conventional interpretation "some or all," and a principle of caution which says that subjects tend to favor particular or negative conclusions over universal or positive conclusions. Their three principles can be used to account for the data quite well, but because it is not clear just how far to carry each of these principles, it is not possible to make <u>a priori</u> predictions independent of the data. Begg and Denny (1969) interpret the effects of atmosphere and caution in a more explicit way as follows:

The first principle, referring to quality, states that whenever the quality of at least one premise is negative, the quality of the most frequently accepted conclusion will be negative; when neither premise is negative, the conclusion will be affirmative. The second principle, referring to quantity, states that whenever the quantity of at least one premise is particular, the quantity of the most frequently accepted conclusion will be particular; when neither premise is particular, the conclusion will be universal.

Stated another way, subjects would tend to make conclusions having the same quality and quantity as the premises, with mixed types being decided in favor of negatives or partic-To see how a conclusion would be drawn from a mixed ulars. pair of premises, consider the premise pair, AE (all-no). Both premises are universal so the conclusion would also be universal. The A statement is positive but the E statement is negative, so the conclusion would be negative. Therefore the conclusion would be the universal negative proposition, Ε. Applying the model to each of the ten possible premise pairs, the following predictions are made: An A conclusion would be predicted from an AA premise pair, an E from AE and EE, I from AI and II, and an O from AO, EI, EO, IO, and OO. An empirical evaluation of these predictions as well as those from other hypotheses will follow after some alternative hypotheses have been discussed.

Chapman and Chapman (1959) proposed that one source of error was an invalid acceptance of the converse of an A or O statement. For example, the O statement, "Some A are not B" would be converted into "Some B are not A." For those pairs not containing A or O statements, they hypothesize a process of "probabilistic inference." They describe it this way:

By one kind of probabilistic inference, S reasons that things that have common qualities or effects are likely to be the same kinds of things, but things that lack common qualities or effects are not likely to be the same. In the syllogism, the available common characteristic is the middle term.

This idea certainly deserves consideration, but the theory seems to have two drawbacks. (1) The authors in their discussion did not try to apply the probabilistic inference uniformly to all of the premise pairs but only to those remaining that were not successfully covered by the conversion hypothesis. It does not seem likely that subjects would switch methods from item to item in this manner. (2) While their description of how the inference is made in specific cases sounds reasonable, it is not clear exactly how the inference would be made in other In other words, it seems to lack the properties cases. of a model capable of generating a priori predictions. At the risk of misinterpreting the authors' intent, it seems that the essential element of their explanations is that of shared properties. The statement "All A are B" could be phrased "All A's have the property B" or for the purpose of making a priori predictions, "A is linked to B." The order, A, I, O, E would be represented by links of decreasing strength until the E proposition which would be represented by "A has no link with B." When two statements are combined, the resulting link would be as strong as the weaker of the two original links. For example, the premise pair IO would yield the conclusion O, it being represented by the weaker link. Applying this process to each of the ten premise pairs, the following predictions are

made: A from AA; I from AI and II; O from AO, IO, and OO; and E from AE, IE, OE, and EE. Since this model disregards order of premises or terms, it is not necessary to use the hypothesis of invalid conversion, this reinterpretation of probabilistic inference being extendable to all premise pairs.

The above hypotheses are by no means exhaustive. The following alternative hypothesis is proposed, based on some informal observations of students in their attempts to work syllogisms.

Most undergraduates, even though quite unsophisticated with formal logic, have had some experience with ordinary number algebra. The fact that abstract syllogisms use letters instead of nouns would tend to favor an "algebraic set" in untrained subjects. With this set, the subject would see the relations between terms as "degrees of equality" and perform operations appropriate to equations, such as the substitution of a quantity for its equal. For example, consider the following pair of premises:

```
All A are B;
Some B are C.
```

With an algebraic set, the subject would interpret the statement "All A are B" as "A = B" and substitute A for B in the second premise, resulting in the conclusion, "Some A are C."

Another example:

Some A are not B; No B are C.

In this case, neither statement is strongly suggestive of equality, but this would not necessarily keep subjects from trying to treat the statements in this way. The question is which terms will be substituted for which? Will A be substituted for B in the second statement, or will C be substituted for B in the first statement? One way to resolve this ambiguity is to postulate the existence of an "equality hierarchy." This would be a given subject's perception of how similar a given statement is to an equa-It seems reasonable that all subjects would rate the tion. A statement as being most like equality, followed by the I statement. Beyond this subjects would vary. One type of subject would see the O statement as being just another, more indirect way of saying the I statement. This would most likely result from the interpretation of "some" as meaning "some but not all" which would imply "some-not." This type of subject would have the equality hierarchy, A-I-O-E. A second type of subject could be thought of as having two hierarchies, one for degree of equality and one for inequality. Many students mistakenly treat inequalities as having the same algebraic properties as equalities, so it would not be very surprising if these subjects formed two sub-orders placing the stronger statement first, producing

the sub-orders A-I and E-O, which could be combined as A-I-E-O. A third type of subject would use only the abbreviated hierarchy A-I.

To see how this model would be applied to the above premise pair, the A-I-O-E subject would consider the first statement (O) to be higher on the equality hierarchy and substitute A for B in the second statement. The A-I-E-O subject would consider the second statement (E) to be higher and substitute C for B in the first statement. The A-I subject would respond "none of the above," as neither premise is an A or I statement. Predicted conclusions for each of the ten premise pairs for each form of the above model are given in Table 1.

Empirical Evaluation of Models

For the purpose of this evaluation, the data gathered by Chapman and Chapman seem to be the most suitable. The multiple-choice format produces more clear-cut patterns of response than the true-false or extended truefalse scale used by earlier researchers. Also the fact that four possible conclusions are given in a single test item allows for complete coverage of different syllogisms with fewer items. Another factor arguing for the superiority of their data is the unusually large number of subjects used (222) with the resulting reliability of

percentages in their tabulations. A summary of predictions based on each of the three models discussed above, together with the experimental results of Chapman and Chapman are given in Table 1.

TABLE 1

			Alg	ebra	ic*	Percent of Responses				
Premises	Atmosphere	Inference	(1)	(2)	(3)	A	E	I	0	N
AA	A	А	A	A	A	80	5	5	1	9
AE	E	Е	Е	Е	E	2	83	2	5	8
AI	I	I	I	I	I	5	5	76	6	8
AO	0	0	0	0	0	2	5	11	70	12
EE	Е	E	N	E	Е	2	54	4	5	35
EI	0	E	Е	Е	Е	2	55	6	18	19
EO	0	E	N	Е	0	3	31	8	24	34
II	I	I	I	I	I	2	5	65	8	20
IO	0	0	0	0	0	1	6	12	55	26
00	0	0	N	0	ο	0	7	11	52	30
Responses Predicted	58%	62%		74%		<u></u>				

COMPARISONS OF PREDICTIONS FROM THREE MODELS WITH DATA FROM CHAPMAN AND CHAPMAN STUDY

*(1)--Equality hierarchy--A-I

(2)--Equality hierarchy--A-I-O-E

(3)--Equality hierarchy--A-I-E-O

N--None of these.

It should be pointed out that each of the models discussed generates predictions from premise pairs without regard to the order of terms within the statements. These variations of order, traditionally called <u>figures</u>, are not necessarily equivalent logically. For example, both of the test items given below have AA premise pairs, but differ in figure, the second premise of each item being the converse of the other. The correct response to the second item is "None of the above," but all of the models discussed would predict "All A are C" as responses to both items.

	All A are B;		All A are B;
	All B are C.		All C are B.
	Therefore,		Therefore,
(1)	All A are C.	(1)	All A are C.
(2)	No A are C.	(2)	No A are C.
(3)	Some A are C.	(3)	Some A are C.
(4)	Some A are not C.	(4)	Some A are not C
(5)	None of the above.	(5)	None of the above.

There is considerable evidence that most subjects in fact ignore differences of figure (Chapman and Chapman, 1959; Johnson, 1973; Experiment 1 of this study). This finding permits a convenient simplification in the presentation of the data of Chapman and Chapman. The percentages for different figures of the same premise pair were averaged and listed in a single row of Table 1.

Each of the three models fits the data reasonably well. For the premise pair, EI, the atmosphere model accounts for a considerably lower percent of the responses.

The probabilistic model predicts the same responses as the algebraic model with equality hierarchy, A-I-O-E. The algebraic model predicts the highest percent of the responses, but it must be kept in mind that because the model predicts more than one response for some of the items, it would necessarily have a mathematical advantage over the other models. To make the comparison fully leqitimate, it would be necessary to have access to individual protocols and to use some a priori method of assigning equality hierarchies to subjects. In the absence of these protocols, an assumption could be made about the percentage of subjects using various substitution hierarchies. For example, note in Table 1 that 80% of the responses to the premise pair AA are the predicted A and that the predicted responses for the next three rows occur at approximately the same percentage. Note also that in some of the rows where different hierarchies predict different responses, the totals of the corresponding percentages are not far from 80%. Suppose the assumption is made that 80% of the subjects conformed to the algebraic model and that there were about an equal number of these subjects using the three equality hierarchies. This would result in the pattern of responses given in Table 2. Another possibility would be to assume the existence of another class of subjects possessing the hierarchy consisting only of the A

statement and further assume that each of the resulting four hierarchies is used by 20% of the subjects. The predictions following this assumption are given in Table 3. Of course these assumptions are rather speculative and only included to illustrate possibilities. Further evaluation of these models can come from other approaches such as a systematic examination of individual protocols or the use of experimental manipulations.

TABLE 2

	ASSUM	PTIO	N OF	EQU. Al	AL DI MONG	STRIB 80% OI	UTIOI F SUI	N OF BJEC'	THR	EE HI	ERARCHIES
	Percent Predicted				P	erce	nt Ol	Dowistions			
mises	A	Е	I	0	N	A	E	I	0	N	from Predictio

_

COMPARISON OF OBSERVED RESPONSES WITH ALGEBRAIC MODEL UNDER

Pre ns AA 0 80 80 • • AE 80 3 83 4 AI •• 80 . . 76 • • • • 10 AO 80 70 . . • • . . • • • • 8 EE 54 27 54 35 •• • • . . •• . . • • 25 ΕI 80 55 • • • • • • • • ΕO 27 27 31 24 34 14 27 . . • • •• • • 15 II 80 65 . . •• • • • • • • • • • • 25 IO 80 55 • • 5 00 54 27 52 30

^aAverage deviation = 10.9.

	Pe	Percent Predicted					erce	nt O	bser	a	
Premises	A	E	I	0	N	A	E	I	0	N	Deviations from Predictions
AA	80	••	••	•••	••	80	•••	•••		••	0
AE	••	80	••	••	••	••	83	••	••	••	3
AI	••	••	80	••	••	••	••	76	••	••	4
AO	••	••	••	80	••	••	••	••	70	••	10
EE	••	40	••	••	40	••	54	••	••	35	19
EI	••	60	••	••	20	••	55	••	••	19	6
EO	••	20	••	20	40	••	31	••	24	34	21
II	••	••	60	••	20	••	••	65	••	20	5
IO	••	••	••	60	2 0	••	••	••	55	2 6	11
00	••	••	••	40	40	••	••	••	52	30	22

COMPARISON OF OBSERVED RESPONSES WITH ALGEBRAIC MODEL UNDER ASSUMPTION OF EQUAL DISTRIBUTION OF FOUR HIERARCHIES AMONG 80% OF SUBJECTS

^aAverage deviation = 10.1.

The latter approach was used by Simpson and Johnson (1966). They constructed a test containing two scales, an atmosphere scale consisting of syllogisms thought to be subject to atmosphere errors but not to conversion errors, and a conversion scale thought to be subject to conversion errors but not to atmosphere errors. The atmosphere scale consisted of syllogisms, EE-E, II-I, OO-O, IO-I or O, and EO-E or O. They interpreted the atmosphere effect a little differently than the model presented earlier in this paper,

TABLE 3

as can be seen from the fact that they considered either an I or O response to the pair, IO to be the result of an atmosphere effect, similarly for the pair, EO. For the conversion scale, they used only different figures of AO-O syllogisms. Other syllogisms were recognized as possibly resulting from invalid conversion, but they were also capable of resulting from the atmosphere effect, so they were not used. For each of these scales, the correct response was "none of the above." Items of this kind are called indeterminant items. Several determinant items were added to the test to prevent a response set. The scales were fairly reliable and uncorrelated with each other, indicating the existence of two separate kinds of error. They used differential training designed specifically as anti-atmosphere and anti-conversion training. The antiatmosphere training was quite effective in reducing errors on the atmosphere scale, but anti-conversion training was less effective in reducing errors on the conversion scale.

Johnson (1972) modified the above experiment using a reinterpretation of the atmosphere effect. He still used the same five premise combinations as before, but used the responses predicted by the Begg and Denny interpretation of the atmosphere effect given earlier in Table 1. From these syllogisms, he formed two scales, a primary atmosphere scale, consisting of the syllogisms EE-E, II-I, and 00-0,

and a supplementary atmosphere scale consisting of the syllogisms, IO-O and EO-O. His conversion scale was formed as before only from AO-O syllogisms. He also used a larger number of determinant items, so that the effect of the training on these items could be better assessed. The primary atmosphere scale was uncorrelated with either the supplementary atmosphere scale, the conversion scale, or the determinant scale, the latter three scales having moderate positive correlations with each other. The antiatmosphere training clearly reduced the number of errors on the atmosphere, supplementary atmosphere, and conversion scales; but at the same time also reduced the number of correct responses on the determinant scale. These results could be interpreted as the production of a "skeptical set" or a tendency to respond "none of the above" if in doubt. The anti-conversion training produced a similar but weaker effect.

How do these results affect the acceptability of the algebraic model discussed earlier? The only result that really causes trouble is the near zero correlation between the primary atmosphere scale and the other scales. It seems that the best way to deal with this result, which was in evidence in three experiments, is to assume that in addition to the algebraic set, there is a tendency in some subjects to reason in an even more superficial manner than

that of the algebraic substitutions. These subjects look at the items and simply make responses that are verbally similar to the premises, without regard for the structure of the sentences. As the responses predicted for this process would not differ much from those predicted by the algebraic model, probably the best approach to gathering further evidence for a verbal similarity set would be the use of individual protocols. Indirect evidence could also be gotten from such factors as an absence of scratch work and a relatively short time spent taking the test.

Summary

The process of instruction was discussed as a process by which an initial state of knowledge is transformed into a desired state by certain transforming actions. The desired state for the task of judging the validity of syllogisms was defined as a behavioral criterion. Alternative descriptions of possible initial states were made, with two of them, one termed an algebraic set and the other a verbal similarity set, being singled out, at least tentatively, as best typifying untrained subjects. The effect of two transforming actions, anti-atmosphere and anticonversion training, were discussed; and it was concluded that the major effect was to transform whatever was the initial state to an equally erroneous state termed a

skeptical set. Additional attempts to find more effective actions for inducing the desired behavior will be discussed in Chapter II.

CHAPTER II

EXPERIMENTS TESTING THE EFFECTS OF CHANGES OF WORDING

As an alternative to training or instructions, a very convenient and effective way to induce a change in a subject's reasoning process is to simply change the task itself so that the subject will reason as though he has been given appropriate training. Following are two experiments which attempt to do this by changes of wording into logically equivalent, but psychologically more informative forms.

Experiment 1

It is generally recognized that the word "some" can be ambiguous to subjects untrained in formal logic, but the effect of clearing up this ambiguity has not been tested specifically. The purpose of this experiment is to see to what extent this ambiguity contributes to errors in judging the validity of syllogisms.

Method

<u>Subjects</u>.--The subjects were sixteen introductory psychology students who stated that they had not had any training with syllogistic reasoning.

<u>Materials</u>.--A 14-item multiple-choice syllogism test was constructed by random selection of items, subject to the following constraints:

- The word "some" must occur in at least one of the two premises. (At least one premise must be I or 0.)
- 2. Each of the remaining seven combinations of premise pairs must be represented once in each half-test.
- 3. With the mixed pairs, each order is used. For example, if the pair AI occurs in the first seven items, then the pair IA must occur in the second seven items.
- 4. One of the four figures (patterns of assignment of the three terms to positions within the statements) is randomly chosen for each of the items, subject to the conditions that figures for the same premise combination are not repeated, and that if a determinant figure for a given premise combination exists, both a determinant and an indeterminant figure must be used for that premise combination.

- 5. The three determinant items are spaced through the fourteen item test by moving, if necessary, the easy AI or IA determinant item to position #1 on the test and the other two to positions #5 and #10.
- Letters of the alphabet are put into random order and assigned in this order, three at a time, as terms for each of the items.

After the items were selected, two forms of the test were constructed, one with the usual wording and the other with modified wording as shown below:

Usual Wording	Modified Wording								
All A are B.	Every A is a B.								
Some A are B.	At least one A is a B.								
Some A are not B.	At least one A is not a B.								
No A are B.	No A is a B.								

Each form of the test contained the following instructions:

Read each syllogism carefully, then select any conclusion which must follow logically from the premises and circle its number. Feel free to write in the empty space if you find it helpful.

The inclusion of empty space was to encourage subjects to show their work, in case some interesting protocols resulted. The departure from the usual forced-choice instruction was made so that any ambivalence in making responses could be observed. The two forms of the test are given in Appendix A. <u>Procedure</u>.--Subjects were run in groups of from 6 to 10. However each subject was randomly assigned to one of two orders on an individual basis, resulting in two treatment groups of 8 subjects each. The first treatment group received the usual form followed by the modified, and the second group, the two forms in opposite order. For each subject, both tests were scored for number correct. There was no limit on the time allowed for completion of the tests.

Results and Discussion

The means and standard deviations for each of the two test forms with each of the two orders are given in Table 4.

TABLE 4

	Means					
	Original Wording	Modified Wordin				
Original first	1.9	2.1				
Modified first	5.4	4.9				
	Standard	Deviations				
Original first	1.4	1.5				
Modified first	4.3	4.1				

EFFECTS OF WORDING AND ORDER ON NUMBER CORRECT

If the scores of the two groups are combined, counterbalancing any effects of order, the means of the original and modified wording are 3.6 and 3.5, respectively, with correlated t = .325, df = 14, and r = .90. If only the first test score for each subject is considered, then we have an independent groups comparison of the two forms of wording. The mean of the original wording is 1.9 and that of the modified wording is 4.9 (Table 4). A one-tailed t-test is statistically significant at the .05 level (t = 1.96, df = 14).

Two interpretations of the above results seem possible. The first interpretation is that the modification in wording made little difference in subject performance, and that the rather large difference in group means was just a matter of sampling error. That is, it just happened that the three most logically sophisticated subjects were assigned to the same group.

The second interpretation is that subjects tended to form a response set after their first test which caused them to continue with the same thought process on their second test. The modified wording was easier to interpret correctly and when it was given first, subjects were more likely to use valid reasoning processes, not only on the first test but also on the second. The reverse was the case with the opposite order.

Examination of individual test papers seemed to favor the first interpretation, that is that the modified wording was not significantly better than the original wording. For example, the subject making the highest score and contributing the most to the higher mean and variance of his group used set diagrams and it seems unlikely that the use of this aid would be influenced by the modification.

Examination of individual protocols reveals that not only are the totals similar on both forms of the test, but also that the response patterns to individual items are similar. Thus, the results on the two forms could be combined, and response frequencies to each premise combination are given in Table 5.

TABLE	5
-------	---

Premises	A	E	I	0	N	I & O
AI	9	3	58	6	8	16
AO	3	5	25	42	14	9
EI	0	45	5	25	16	5
EO	0	34	20	22	22	0
II	1	1	59	9	18	9
IO	0	3	25	28	28	13
00	2	6	22	36	23	11

PERCENT OF RESPONSE OF EACH KIND TO EACH PREMISE COMBINATION

There are a few things that could be observed from this tabulation and the original test papers:

1. Even with the modified test format allowing the selection of more than one response to a given set of premises, the pattern of errors was very similar to that of earlier studies using the usual forced-choice format.

2. All of the simultaneous I and O responses shown in Table 2 were made by three of the sixteen subjects. These subjects also showed a considerable number of items in which either the I or O response had been erased in favor of the other response, indicating a degree of ambivalence in response between these two sentence forms. Other subjects also showed some inconsistency in making an O response where an I response had been made earlier in the test to the same premise combination and vice versa. All of the above suggests the possibility that some of the subjects were interpreting the I and O statements as meaning the same. The similarity of responses on the two forms of the test by a given subject indicates that a change from "some" to "at least one" is ineffective in dispelling this interpretation.

3. The format of the test allowed for an examination of scratch work done by the subjects. Six of them left their spaces blank, but ten showed considerable use of symbolic work, mostly in the form of algebraic equations

and inequalities, supporting the algebraic model discussed in Chapter I. It is interesting that some of the subjects would write very infrequently and very small, suggesting a desire to try to "do it in their heads" even when supplied with plenty of writing space. Thirteen of the sixteen subjects ranged in score from 1 to 8 out of a possible 28 (14+14), two subjects had intermediate scores of 14 and 17, and one subject had an almost perfect score of 26. The latter subject was the only one to use set diagrams. He claimed not to have had previous training with syllogistic reasoning but said that he had had some exposure to work with sets in mathematics courses.

Experiment II

Set diagrams are often taught as an aid to reasoning in mathematics and logic courses. Frandsen (1969) showed training with diagrams to be especially useful in overcoming difficulties of students with low spatial aptitude. Whimbey and Ryan (1969) found ability to solve syllogisms to be related to short-term memory as measured by a modified digit test, and proposed that the use of diagrams functioned to reduce dependence on short-term memory. Johnson (1972) used diagrams to aid in his anti-atmosphere and anti-conversion training. Each of these researchers used diagrams as a part of a more complete training procedure, but the question
remains, what would be the effect of simply inducing subjects to use diagrams, without additional training? At the very least, the use of diagrams would be expected to break up either of the two cognitive sets proposed in Chapter I, allowing the possibility of a change to a correct process. Whether or not this happens for a particular subject, additional information would be gained about the initial cognitive state of the untrained subject. To test this effect of inducing subjects to use diagrams, the method of modified wording as used in Experiment I was again used, this time to appropriate spatial wording.

Method

<u>Subjects</u>.--This experiment was run as an adjunct to that of Johnson (1973). The subjects were 20 introductory psychology students from his control group.

<u>Materials</u>.--An ll-item subtest of that used by Johnson was constructed by randomly selecting one of each of the ll combinations of premises appearing on the original test. Five items were determinant and six were indeterminant. These syllogisms were reworded in logically equivalent spatial terms as follows:

Original	Reworded			
All A are B.	A is inside B.			
No A are B.	A and B are separate.			
Some A are B.	A and B overlap.			
Some A are not B.	Some of A lies outside B.			

The spatial test had the following instructions:

The following are statements about geographical regions. For each pair of statements circle the number of any (one or more) of the conclusions listed that follow logically. You may use the margins if you find it useful to make notes or drawings of the situations described by the statements.

A copy of the spatial form of the test is given in Appendix B.

<u>Procedure</u>.--Each subject was given the original syllogism test and then given the spatial test. Counterbalancing the order of the two tests, while desirable, was not possible without compromising the Johnson experiment.

Results and Discussion

The mean number correct out of eleven was 5.3 and 5.8 for the original and spatial tests, but the difference between these means was not significant. The relationship between the performances on the two tests was interesting and is given in the scatterplot shown in Figure 1.

There seems to be three types of subject. One type scored low on the original test and as low or lower on the spatial test but with a different pattern of errors. A second type scored low on the original test, but scored considerably higher on the spatial form. A third type scored high on both tests. There were no subjects getting more than 6 out of 11 of the items correct on the original test who got 6 or less of the items on the spatial test.



Figure 1. Number correct (out of 11) on each form of test.

If it can be assumed that subjects would have no difficulty in translating statements with traditional wording into spatial terms, then it would be expected that those subjects represented in the upper left portion of the scatterplot would have profited from the simple suggestion to make this translation into spatial terms. As for those scoring low on both tests, more would be required.

Examination of individual papers suggests that an important variable in determining whether or not a given subject will profit from a change from verbal to spatial representation is whether or not the subject uses drawings that indicate the consideration of more than one case. A count of the number of items for which a given subject drew multiple-case drawings was made and if non-zero, was indicated on the subject's point in Figure 1. A high number was not necessary but seems to have been sufficient for a high score. It seems reasonable that some of the subjects would be able to visualize the cases without actually making the drawings. Assuming that lack of consideration of all the cases implied by a set of premises is an important reason for invalid conclusions, it is not clear whether subjects have difficulty generating different special cases or whether they just fail to look for these cases. This question will be considered in Chapter III.

Summary

Two attempts were made to modify the initial cognitive state as described in Chapter I. The first, a removal of the ambiguity in the word "some," did not yield conclusive results. Subjects showed little change in response with modification of wording, and the marginal difference between groups was interpreted as the result of sampling error. Both the tabulation of responses and informal examination of individual papers gave additional evidence for the verbal and algebraic models described in Chapter I, as well as the tendency for subjects to interpret I and O statements as the same even with modified wording designed to remove the ambiguity.

The second attempt at modification, the use of spatial wording, was differentially effective. Some subjects could function effectively with either kind of problem. Some subjects scored considerably higher on the spatial task, suggesting that if they were taught to represent statements spatially, they would show dramatic improvement on the original task. Other subjects misused the diagrams, often drawing conclusions on the basis of a single special case. Thus these subjects simply made different kinds of errors on the spatial form and were equally ineffective with both tasks.

In the experiment reported in the next chapter, subjects were forced to consider alternative possibilities.

CHAPTER III

EXPERIMENTS TESTING ABILITY TO GENERATE COUNTER-EXAMPLES TO INVALID SYLLOGISMS

Experiment III

In the process of solving syllogisms by diagrams, successful subjects seem to draw or visualize different positions of the sets making up the diagram that are consistent with the premises, in an effort to eliminate potential conclusions, until it is clear whether or not a conclusion must follow. For example, consider the following item:

> All A are B; Some B are C Therefore, 1. All A are C. 2. No A are C. 3. Some A are C. 4. Some A are not C. 5. None of the above.

In this problem, it seems likely that the subject would first draw the first premise, which would be represented by two regions. A typical example is shown by the solid lines in the diagram at right. Then he could represent the second premise by drawing or imagining

alternative positions for the third region, such as are shown by the dotted regions, until it is clear that each of the four alternatives is possible, so that none of the alternatives is necessarily true. Therefore, the correct response would be "None of the above."

Actually, it is possible to be more systematic in the above process. To be certain that none of the alternatives can be deduced, it is enough to find a single example that will refute "Some A are C" and a single example to refute "Some A are not B." Refuting the I statement will automatically refute the stronger A statement, and refuting the O statement will automatically refute the stronger E statement. To test the ability of subjects to generate counter-examples, the subject's time can be better used if instead of the multiple-choice format, test items consisting of single invalid syllogisms are used. If the subject can find counter-examples for a randomly selected list of invalid syllogisms, it is assumed that he could find counter-examples for each of the invalid conclusions given in a list of multiple-choice items.

While set diagrams are in wide use, other methods are possible. Two of these are the use of specific nouns and the use of general verbal descriptions. Each of these methods is illustrated below:

> All A are B; Some B are C. Therefore, Some A are C.

Solutions:

Spatial Verbal Specific		Verbal-General ¹				
B A C	All dogs are animals; Some animals are cats. Therefore, Some dogs are cats.	All A are B, but not all B have to be A. The C's could be just those B's that are not A's, refuting "Some A are C."				

It would seem reasonable to predict that the spatial approach would give the best results with most subjects. The verbal-specific method has the advantage of concreteness and direct relation to the sentence forms but is not conducive to consideration of general structures. This method would then demand a considerable amount of divergent production, requiring more trial-and-error than would be expected with the other two methods. The verbal-general method has the advantage of inducing consideration of general structures but lacks the concreteness of the other two methods. It also seems likely that it would place a greater demand on short-term memory. Moreover, this method might be more easily misunderstood with limited instruction. The spatial method has both generality and concreteness but requires a translation into spatial terms. However, in Experiment II there was little evidence of difficulty in correctly making the translation from spatial wording to diagrams; and since there is little reason to believe that

¹This method was derived from the protocols of a mathematical psychologist, using his customary method.

translation from the original wording to spatial wording would be difficult, it is predicted that this method of generating examples would be the easiest for untrained subjects to learn and apply.

First a pilot study was done to determine the number of items that subjects could handle in the allotted span of time. The number of items was relatively small. In fact, the number of counter-examples that the subjects could generate was much smaller than the number of syllogisms that they could answer in the same length of time. This was true despite the fact that a single counter-example is only part of what is required for the correct solution of the usual multiple-choice item, that is the rejection of one particular incorrect answer. Actually this is not so surprising when one considers that in this task it is impossible for subjects to use the fast but superficial algebraic substitution or verbal similarity processes. They are effectively forced to come to grips with specific cases. Since this is what the logician must do, performance on this task should reflect more clearly the subject's potential to profit from instructions to consider special cases and to eliminate possibilities.

Method

<u>Subjects.--</u>The subjects were 42 introductory psychology students.

<u>Materials</u>.--Ten invalid syllogisms were randomly selected, subject to the following criteria:

- Each of the ten kinds of premise pairs that can be formed from A, E, I, or O statements were ordered randomly.
- 2. For each premise pair, the conclusion was chosen to be either an I or O statement. Five I and five O conclusions were put in random order and assigned to the ten premise pairs.
- 3. The resulting triples of statement types were randomly assigned one of the four figures. If the resulting syllogism was valid, then the conclusion was strengthened from I to A or O to E, whichever applied. If the syllogism was <u>still</u> valid, then another figure was selected and the process repeated. (As it turned out, it was not necessary to apply this rule.) For most items, the premises could be represented by a number of different drawings, requiring some searching by the subject. However for one item there was only one drawing possible, and a different figure was selected.
- 4. The alphabet was randomly ordered and assigned to terms of the syllogisms.

The ten items were divided into three parts. The first two items were sample problems, the next three were to be scored but not timed, and the last five were to be scored and timed. Some reordering was done to produce a "better" test. That is, the first item seemed to be particularly difficult and was shifted to the end of the list. The second and third items seemed to make particularly good illustrative problems, so they were moved to the top of the list. The item in the ninth position seemed particularly easy, so it was moved to the third position, just after the sample problems.

Three forms of the test were constructed, one for each of the three methods described earlier. The forms were identical except for the explanations of the sample solutions. (The three methods of counter-example production were judged to be sufficiently different to prevent undue interference from one task to the other. The pilot study showed the instructions to be effective in causing the required shift in the subject's method.) The timed portions were separate from the untimed portions. These tests are given in Appendix C.

<u>Procedure</u>.--Subjects were tested in groups of from 6 to 15. However each subject was randomly assigned to one of the six treatment conditions on an individual basis. After a brief discussion of the general purpose and design

of the experiment at the beginning of the session, the untimed portion of the test was distributed to the subjects. When a subject finished with this portion, he was given the second portion with the time written on the sheet. When the subject returned, a second time was written on the sheet and then the process repeated with his second method. Each of the eight test items was scored as correct or incorrect and the number correct was recorded.

Results

For each pair of methods, there were two relevant groups. One of the groups got the two methods in the order AB, the other BA. The results for each such pair of groups were analyzed separately. Table 6 gives the results of comparing subjects' scores for spatial counterexamples with those for verbal-specific counter-examples.

Table 7 gives the results of comparing subjects' scores for spatial counter-examples with those for verbalgeneral counter-examples.

Table 8 gives the results of comparing subjects' scores for verbal-specific counter-examples with those of verbal-general counter-examples.

COMPARISON OF SPATIAL AND VERBAL-SPECIFIC METHODS OF COUNTER-EXAMPLE PRODUCTION

_

Mean Correct (out of fo	ur)					
	I Co	nclusi	ons			
Groups	Spat	ial	Verb	al-Speci	fic	
Spatial first	3.3	3		2.9		3.1
Verbal-specific first	2.0	C		2 . 5		2.3
	2.0	5		2.7		2.7
	0 Cc	onclusio	ons			
Spatial first	1.7	7		0.7		1.2
Verbal-specific first	0.4	4		1.0		0.7
	1.2	1.1 0.8			1.0	
Analysis of Variance						
Source	SS	df	MS	F	P	e⁺.a?
Groups (G)	5.785	1	5.8 [°]	1.1	. u G	•••
Ss/G	62.429	12	5 J 2	ه ن ت	u O D	• • •
Methods (M)	0.071	1	0.07	600		
GXM	5.787	1	5.8	4.3		
M X Ss/G	16.142	12	1.3	o u •	•••	c • •
Conclusions (C)	41.143	1	41.1	39.3	.001	.26
GXC	0.286	1	0.3	0 0 6		
C X Ss/G	12.571	12	1.0	∘ • •	• • •	• • •
мхс	0.286	1	0.3			
G X M X C	0.285	1	0 . 3		د ن ه	a u •
M X C X Ss/G	15.429	12	1.3	u • a		•••
Total	160.214	55		9 U 9	• • •	.26

COMPARISON OF SPATIAL AND VERBAL-GENERAL METHODS OF COUNTER-EXAMPLE PRODUCTION

Mean Correct (out of four)

	1 00	onclusio	ons			
Groups	Spat	Spatial		al-Gener	al	
Spatial first	2.	4		0.9		
Verbal-general first	2.	6		1.1		
	2.	5	1.0			1.8
	0 Cc	onclusio	ons			
Spatial first	0.9	9		1.3		1.1
Verbal-general first	0.	7		0.7		
	0.	8		1.0		
Analysis of Variance						
Source	SS	df	MS	F	р	eta ²
Groups (G)	0.055	1	0.05	c o e	D G 0	•••
Ss/G	44.389	12	3.70		•••	• • •
Methods (M)	4.500	1	4.5	2.1		
GXM	0.056	1	0.06		•••	•••
M X Ss/G	25.944	12	2.2	•••	•••	•••
Conclusions (C)	8.000	1	8.0	5.3	.05	.06
GXC	0.889	1	0.9	•••	•••	• • •
C X Ss/G	17.611	12	1.5	•••	• • •	•••
мхс	7. 889 [°]	1	7.9	5.3	.05	.06
G X C X Ss/G	18.278	12	1.5	•••	• • •	• • •
Total	128.033	55	• • •	•••	• • •	.12

I Conclusions

COMPARISON OF VERBAL-SPECIFIC AND VERBAL-GENERAL METHODS OF COUNTER-EXAMPLE PRODUCTION

Mean Correct						
		I Conc	lusions			
Groups	Ve	rbal-Spe	cific	Verbal-Ge	neral	
Verbal-specific first		1.7		0.1		0.9
Verbal-general first		2.3		0.4		1.4
		2.0		0.3		1.1
n an		0 Conc	lusions			
Verbal-specific first		1.0		0.3		0.6
Verbal-general first		0.3		0.3		0.3
		0.6		0.3		0.5
Analysis of Variance				Lee		
Source	SS	df	MS	5 F	q	eta ²
Groups (G)	0.01	B 1	0.0)2	D ¢ U	0 C 0
S s/ G	12.57	1 12	1.0)		
Methods (M)	15.01	в 1	15.0	12.4	.005	. 21
GXM	0.16	0 1	0.1			010
M X Ss/G	14.57	2 12	1.2	214	0 U U	0 V G
Conclusions (C)	6.44	6 1	6.4	12.6	.005	٥٥،
GXC	2.16	1 1	2.2	2 4.2		•••
C X Ss/G	6.14	3 12	0.5	51	с с о	c e e
мхс	6.26	в 1	6.3	3 11.7	.01	.09
G X M X C	0.87	61	0.8	38		
M X C X Ss/G	6.42	8 12	0.5	54	• • •	• ~ 4
Total	70.66	1 55	0 • ·		• • •	.39

To summarize the statistically significant results, scores for the I conclusions were significantly higher than those for the O conclusions. This was true for each of the three analyses. Since this variable and its interactions with method accounted for so much of the significant variance, comparisons between pairs of methods were carried out for I and O conclusions separately. None of the group-bymethod interactions (order effects) were significant, so the two orders were combined for each of the three pairs of methods. For the I conclusions, both the spatial and verbal-specific methods had significantly higher means than the verbal-general method (t = 2.94 and t = 4.17, df = 13) but did not differ significantly from each other. Means for the O conclusions were all low and did not differ significantly. This "floor" effect seems to have been enough to mask the effect of the method variable and produce the significant method-by-conclusion interactions.

Discussion

Subjects evidently found the verbal-general method considerably more difficult than either of the other two methods. This does not mean that this method could not be potentially useful, only that the small amount of training used in this experiment was not sufficient to help the type of student used. Perhaps the results would have been different, had either more logically sophisticated subjects

been used or more extensive training used. However, the emphasis in this study is on the effects of minimal alterations in the cognitive state of untrained subjects, so further investigation will use only the spatial and verbalspecific methods.

It is also evident that subjects had considerably more difficulty with O conclusions than with I conclusions. In negating the O statements, many subjects used I or E statements instead of the A statement, which is the correct negation of the O statement. The pervasiveness of errors traceable to incorrect negation of statement suggests an experimental manipulation that should produce significant reduction of errors and in addition, allow a more sensitive comparison of the remaining two methods of generating exam-The generation of counter-examples can be considered ples. to be a two-step process, the first to negate the conclusion, and the second to generate an appropriate example which is consistent with both premises and the negation of the conclusion. The effect of eliminating this first step was tested in the next experiment.

Experiment IV

This experiment has two purposes. One is to demonstrate a significant increase in correct solutions when subjects are given what are essentially the same problems as in Experiment III, but with the step of negating the conclusions removed. This would give additional verification of the hypothesis that this step is a major source of difficulty. The second purpose is to demonstrate what was hypothesized but not verified in Experiment III, that is that the spatial method of generating examples is superior to the verbal-specific method.

Method

<u>Subjects</u>.--Twenty introductory psychology students were assigned randomly, ten to each of two groups.

<u>Materials</u>.--Each of the ten syllogisms used in Experiment III was modified by negating the conclusion. The instructions were also modified in keeping with the change in items. For example, instead of asking subjects to give a counter-example for the invalid syllogism,

> All A are B; Some B are C. Therefore, Some A are C

they are asked to give an example which satisfies each of the three statements,

All A are B; Some B are C; No A are C.

Note that the modified item is different only in that the third statement has been replaced by its negation. The tests are given in Appendix D.

<u>Procedure</u>.--This was the same as in Experiment III except that only two experimental groups were used, one for each order of the two methods to be compared.

Results

To test the effect of removing the conclusionnegation component from the task, the initial test scores of each of the twenty subjects in this experiment were compared with twenty corresponding scores selected randomly from Experiment III. This permitted a two-by-two independent groups comparison, and the results are given in Table 9. Scores on the reduced task were significantly higher than those on the original task.

To compare the two methods for the reduced task, solution times for the final five problems were used. In an effort to separate the thinking and writing components of the solution process, writing times were compared for three subjects by having them copy randomly selected answer sheets. The mean difference in writing time was 2:19 min. with 7 sec. standard error of the mean. A time, 2:40, which is three standard errors above the mean, was

Mean Correct (out of eigh	t)		1			
Task			Spatial		Verbal-S	pecific	
Negation requi	red	4.9		3.3			
Negation not required			7.1		6.5		
Analysis of Va	riance						
Source	SS	df	MS	F	P	eta ²	
Tasks (T)	72.9	1	72.9	19.5	.001	.33	
Methods (M)	12.1	1	12.1	3.24	•••	•••	
тхм	2.5	1	2.5	0.67	•••	• • •	
Error	134.4	36	3.73	• • •	•••	•••	
Total	221.9	39	•••	• • •	•••	•••	

EFFECT OF REMOVING CONCLUSION-NEGATION STEP FROM SPATIAL AND VERBAL-SPECIFIC METHODS OF COUNTER-EXAMPLE PRODUCTION

subtracted from the times for the verbal-specific method. This would seem to be a liberal correction, especially considering that it would be reasonable to assume that some thinking could take place during writing. With this adjustment, the mean time for the spatial method was 5.1 min. and the corrected mean time for the verbal-specific method was 10.0 min. The analysis of variance for this comparison is given in Table 10.

Source	SS	df	MS	F	р	eta ²
Groups (G)	10.1	1	10.1	•••	•••	•••
Ss/G	261.1	18	14.5	•••	• • •	•••
Methods (M)	245.8	1	245.8	19.1	.001	.32
GXM	27.5	1	27.5	2.1	• • •	•••
M X Ss/G	232.75	18	12.9		• • •	• • •
Total	777.3	39		• • 0	• • •	•••

ANALYSIS OF VARIANCE FOR SOLUTION TIMES FOR SPATIAL AND VERBAL-SPECIFIC METHODS

Discussion

The results of this experiment, together with those of Experiment III, indicate that for the task of generating counter-examples to invalid syllogisms, a major source of difficulty is the first step of forming the negations of the conclusions. For the second step of generating the examples, the spatial method takes less time than the verbal-specific method. Both are more accurate than the verbal-general method, at least for untrained subjects. Implications for instruction in producing counter-examples are that students must be given some training in forming statement negations, and that the conventional tendency to use set diagrams in the instruction of syllogisms has experimental backing.

CHAPTER IV

IMPLICATIONS FOR INSTRUCTIONAL PROCESS

In Chapter I, descriptions of the initial cognitive state were proposed that were relevant to subject behavior when misinterpreting traditional syllogism items, but these processes are easily changed and seem to have little to do with what happens when attempts are made to induce subjects to use an effective process. As a result of the experiments of this study, the following elements of the initial cognitive state are proposed as more relevant to the instructional process:

- Ability to translate statements into diagrams and to work with these diagrams, when shown by a couple of examples.
- 2. Tendency to overlook alternatives.
- Ability to generate specific examples when tendency to overlook alternatives is corrected. Diagrams are not necessary, but facilitate in this process.
- Tendency to negate propositions incorrectly, especially the 0 statement.

Using the above as descriptive of untrained subjects, it is now possible to propose a simple training procedure

that could be tested experimentally. The subject could be shown a listing of the diagrams representing the five possible relations between two categories and then the A, E, I, and O statements could be defined as collections of these relations by simply indicating them as in Figure 2.



Figure 2. Set of diagrams for use in instruction.

This procedure should induce consideration of alternatives by the fact that they are explicitly listed. The negation of a statement could be explained as simply all other cases. For example, it is clear from the drawings the complementary nature of the A and O statements and the E and I statements.

An experiment testing the effect of showing subjects the information given in Figure 2 would complement a study done by Ceraso and Provitera (1971). They used syllogisms

constructed from simpler statements corresponding to a single drawing of Figure 2. For example, they used the statement, "Whenever I have a yellow block it is striped, but there are some striped blocks which are not yellow." They presented the premises to their subjects by holding up objects and orally pointing out the information given in the premises. This task was significantly easier for subjects than the corresponding task using traditional statements. Unfortunately, it is not possible to tell from these results to what extent the better performance on the modified syllogisms resulted from the fact that the meaning of the statements was made more explicit or from the fact that they were logically simpler. If the example mentioned above is put into abstract form, the statement "All A are B, but some B are not A" was compared with the traditional statement "All A are B," but the latter statement is logically equivalent to the statement "All A are B," but some B might not be A," a logically more complex statement. A replication of the Ceraso and Provitera study using the latter forms should produce an interesting supplement to the study. This replication in verbal form could then be compared with the proposed study using diagrams. It would be predicted that the diagram method would be superior in that not only would the alternatives be made explicit, but in handling these alternatives it would not place as much burden on the short-term memory as the verbal method.

The elements of the cognitive state listed above were arrived at by a sequence of modifications of the original task of judging the validity of a syllogism. At first modifications of wording were used. Then it was necessary to look at the subtask of counter-example production and finally at an even smaller subtask of this subtask. From this analysis, inferences were drawn about what should be effective instruction for the original task, but the process is not complete until some training such as that mentioned earlier in this chapter has actually been tried. REFERENCES

REFERENCES

- Begg, I., and Denny, J. P. Empirical reconciliation of atmosphere and conversion interpretations of syllogistic reasoning errors. <u>Journal of Experimental Psychology</u>, 1969, 81, 351-354.
- Ceraso, J., and Provitera, A. Sources of error in syllogistic reasoning. <u>Cognitive Psychology</u>, <u>2</u>, 1971, 400-410.
- Chapman, L. J., and Chapman, J. P. Atmosphere effect reexamined. Journal of Experimental Psychology, 1959, 58, 220-226.
- Frandsen, A. N., and Holder, J. R. Spatial visualization in solving complex verbal problems. Journal of Psychology, 1969, 73, 229-233.
- Dillehay, R. C., Insko, C. A., Smith, M. B. Logical consistency and attitude change. Journal of Personality and Social Psychology, 1966, 3, 646-654.
- Frase, L. T. Validity judgment of syllogisms in relation to two sets of terms. Journal of Educational Psychology, 1966, 57, 239-245. (a)
- Glaser, R., and Resnick, L. B. Instructional psychology. Annual Review of Psychology, 1972, 23, 207-276.
- Janis, I. L., and Frick, F. The relationship between attitudes toward conclusions and errors in judging logical validity of syllogisms. <u>Journal of Experimental</u> Psychology, 1943, 33, 73-77.
- Johnson, D. M. <u>Systematic introduction to the psychology of</u> thinking. New York: Harper and Row, 1972, 236-240.

Johnson, D. M. 1973 (unpublished to date).

- Luchins, A. S., and Luchins, E. H. Reactions to inconsistencies: Phenominal versus logical contradictions. Journal of General Psychology, 1965, 73, 47-65.
- Parrott, G. L. The effects of instructions, transfer, and content on reasoning time. Unpublished Ph.D. thesis, Michigan State University, 1969.
- Parrott, G. L. The effects of premise content on accuracy and solution time in syllogistic reasoning. Unpublished Master's thesis, Michigan State University, 1967.
- Sells, S. B. The atmosphere effect: An experimental study of reasoning. Archives of Psychology, 1936, No. 200.
- Simpson, M. E., and Johnson, D. M. Atmosphere and conversion errors in syllogistic reasoning. Journal of Experimental Psychology, 1966, 72, 197-200.
- Stratton, R. P. Atmosphere and conversion errors in syllogistic reasoning with contextual material and the effect of differential training. Unpublished Master's thesis, Michigan State University, 1970.
- Whimbey, A. E., and Ryan, S. F. Role of short term memory and training in solving reasoning problems mentally. Journal of Educational Psychology, 1969, 60, 361-364.
- Wilkins, M. C. The effect of changed material on ability to do formal syllogistic reasoning. <u>Archives of</u> Psychology, 1928, No. 102.
- Woodworth, R. S., and Sells, S. B. An atmosphere effect in formal syllogistic reasoning. <u>Journal of Experi-</u> mental Psychology, 1935, 18, 451-460.

APPENDICES

. | |

APPENDIX A

LOGICAL REASONING TEST (Test 1) Logical Reasoning Test (Test 1)

Read each syllogism carefully, then select any conclusion which <u>must</u> follow logically from the premises and circle its number. Feel free to write in the empty space if you find it helpful.

- All V are X; 1. Some V are Z. Therefore, 1. All Z are X. No Z are X. 2. 3. Some Z are X. 4. Some Z are not X. 5. None of the above. 2. Some K are not E; All S are E. Therefore, All S are K. 1. 2. No S are K. 3. Some S are K. 4. Some S are not K. 5. None of the above. 3. No I are P; Some U are not I. Therefore, 1. All U are P. No U are P. 2. 3. Some U are P. 4. Some U are not P. 5. None of the above. 4. Some B are T; Some M are B. Therefore, All M are T. 1. 2. No M are T. 3. Some M are T. 4. Some M are not T. 5. None of the above. No A are G; 5.
- Some G are W. Therefore, 1. All W are A. 2. No W are A. 3. Some W are A. 4. Some W are not A. 5.
 - None of the above.

Some N are Q; 6. Some N are J. Therefore, All J are Q. 1. 2. No J are Q. 3. Some J are Q. Some J are not Q. 4. 5. None of the above. 7. Some H are not D; Some D are not O; Therefore, All H are O. 1. 2. No H are O. 3. Some H are O. 4. Some H are not O. 5. None of the above. 8. Some R are F; No C are F. Therefore, 1. All C are R. No C are R. 2. 3. Some C are R. Some C are not R. 4. 5. None of the above. 9. Some L are Y; All V are Y. Therefore, All V are L. 1. 2. No V are L. 3. Some V are L. 4. Some V are not L. None of the above. 5. 10. All X are Z: Some K are not Z. Therefore, All K are X. 1. 2. No K are X. 3. Some K are X. 4. Some K are not X. 5. None of the above. 11. Some E are not S; Some E are I. Therefore, 1. All I are S. 2. No I are S. 3. Some I are S. 4. Some I are not S. 5. None of the above.

- 12. Some P are U; Some U are B. Therefore, 1. All B are P. 2. No B are P. 3. Some B are P. 4. Some B are not P. 5. None of the above. 13. Some T are not M; Some A are not T. Therefore, 1. All A are M. No A are M. 2. 3. Some A are M. 4. Some A are not M. 5. None of the above. 14. Some G are not W;
- No N are W. Therefore, 1. All N are G.

 - 2. No N are G.
 - 3. Some N are G.
 - 4. Some N are not G.
 - 5. None of the above.

Logical Reasoning Test (Test 2)

Read each syllogism carefully, then select any conclusion which <u>must</u> follow logically from the premises and circle its number. Feel free to write in the empty space if you find it helpful.

- 1. Every V is an X; At least one V is a Z. Therefore, Every Z is an X. 1. 2. No Z is an X. 3. At least one Z is an X. 4. At least one Z is not an X. 5. None of the above. 2. At least one K is not an E; Every S is an E. Therefore, 1. Every S is a K. No S is a K. 2. At least one S is a K. 3. 4. At least one S is not a K. 5. None of the above. No I is a P; 3. At least one U is not an I. Therefore, Every U is a P. 1. No U is a P. 2. 3. At least one U is a P. At least one U is not a P. 4. 5. None of the above. 4. At least one B is a T; At least one M is a B. Therefore, Every M is a T. 1. 2. No M is a T. At least one M is a T. 3. 4. At least one M is not a T. 5. None of the above. 5. No A is a G. At least one G is a W. Therefore, 1. Every W is an A. No W is an A. 2. 3. At least one W is an A. At least one W is not an A. 4.
 - 5. None of the above.
6. At least one N is a Q. At least one N is a J. Therefore, 1. Every J is a Q. No J is a Q. 2. At least one J is a Q. 3. 4. At least one J is not a Q. 5. None of the above. At least one H is not a D; 7. At least one D is not an O. Therefore, 1. Every H is an O. No H is an O. 2. 3. At least one H is an O. At least one H is not an O. 4. 5. None of the above. At least one R is an F; 8. No C is an F. Therefore, 1. Every C is an R. 2. No C is an R. 3. At least one C is an R. 4. At least one C is not an R. None of the above. 5. 9. At least one L is a Y. Every V is a Y. Therefore, Every V is an L. 1. 2. No V is an L. 3. At least one V is an L. At least one V is not an L. 4. 5. None of the above. 10. Every X is a Z; At least one K is not a Z. Therefore, Every K is an X. 1. 2. No K is an X. At least one K is an X. 3. 4. At least one K is not an X. 5. None of the above. 11. At least one E is not an S; At least one E is an I. Therefore, 1. Every I is an S. No I is an S. 2. 3. At least one I is an S. At least one I is not an S. 4. 5. None of the above.

59

- 12. At least one P is a U; At least one U is a B. Therefore, 1. Every B is a P. 2. No B is a P. 3. At least one B is a P. 4. At least one B is a B. 5. None of the above. 13. At least one T is not an M; At least one A is not a T. Therefore, Every A is an M. 1. 2. No A is an M. 3. At least one A is an M. 4. At least one A is not an M. 5. None of the above. 14. At least one G is not a W; No N is a W. Therefore, 1. Every N is a G. 2. No Nⁱs a G. 3. At least one N is a G.
 - 4. At least one N is not a G.
 - 5. None of the above.

APPENDIX B

SPATIAL REASONING TEST (Brief Form)



. .

SPATIAL REASONING TEST (Brief Form)

The following are statements about geographical regions. For each pair of statements, circle the number of any (one or more) of the conclusions listed that follow logically. You may use the margins if you find it useful to make notes or drawings of the situations described by the statements.

- M lies inside N; N lies inside O. Therefore,
 M lies inside O.
 M and O are separate.
 M and O overlap.
 - 4. Some of M lies outside O.
 - 5. None of the above.
- 2. T and X are separate; X and B are separate. Therefore,
 - 1. T lies inside B.
 - 2. T and B are separate.
 - 3. T and B overlap.
 - 4. Some of T lies outside B.
 - 5. None of the above.
- 3. A lies inside B; Some of B lies outside C. Therefore,
 - 1. A lies inside C.
 - 2. A and C are separate.
 - 3. A and C overlap.
 - 4. Some of A lies outside C.
 - 5. None of the above
- P lies inside Q;
 Q and R are separate.
 Therefore,
 - 1. P lies inside R.
 - 2. P and R are separate.
 - 3. P and R overlap.
 - 4. Some of P lies outside R.
 - 5. None of the above.
- 5. V and W overlap; U and V overlap. Therefore
 - 1. U lies inside W.
 - 2. U and W are separate.
 - 3. U and W overlap.
 - 4. Some of U lies outside W.
 - 5. None of the above.

- 6. X and Y overlap; Some of Y lies outside Z. Therefore, X lies inside Z. 1. 2. X and Z are separate. 3. X and Z overlap. Some of X lies outside Z. 4. 5. None of the above. 7. V lies inside T; Some of S lies outside T. Therefore, S lies inside V. 1. 2. S and V are separate. 3. S and V overlap.
 - 4. Some of S lies outside V.
 - 5. None of the above.
- Some of J lies outside K; Some of K lies outside L. Therefore,
 - 1. J lies inside L.
 - 2. J and L are separate.
 - 3. J and L overlap.
 - 4. Some of J lies outside L.
 - 5. None of the above.
- 9. K and L are separate; Some of L lies outside M. Therefore,
 - 1. K lies inside M.
 - 2. K and M are separate.
 - 3. K and M overlap.
 - 4. Some of K lies outside M.
 - 5. None of the above.
- 10. S and M are separate; M and P overlap. Therefore,
 - 1. S lies inside P.
 - 2. S and P are separate.
 - 3. S and P overlap.
 - 4. Some of S lies outside P.
 - 5. None of the above.
- 11. S and T overlap; T lies inside V. Therefore, 1. S lies inside V. 2. S and V are separate.
 - 3. S and V overlap.
 - 4. Some of S lies outside V.
 - 5. None of the above.

APPENDIX C

INVALID SYLLOGISMS

INVALID SYLLOGISMS

A syllogism consists of two statements called premises, which are assumed to be true, together with a third statement, called the conclusion, which may or may not logically follow from the premises. For example, consider the following syllogism:

> No K are V; Some K are R. Therefore, Some R are V.



Statements of this type can be represented by diagrams as shown above. Note that regions K and V are separate, representing

"No K are V," and that regions K ane R have points in common, representing "Some K are R," but that "Some R are V" is false as indicated by their separation in the diagram. Therefore, the conclusion does not necessarily follow from the two premises.

Let us look at another example:

Solution

2. All C are Q; Some C are not S. Therefore, Some S are not Q.



Note that the word "some" is given the interpretation "some or all" in the second premise. This wider interpretation is customary in formal logic.

As an aid to finding an appropriate counter-example, it is often useful to form the negation of the conclusion ahead of time. The above problem would then reduce to finding an example for which each of the following statements is true:

> All C are Q; Some C are not S; All S are Q (negation of "some S are not Q").

> > Go to next page.

INVALID SYLLOGISMS

A syllogism consists of two statements called premises, which are assumed to be true, together with a third statement, called the conclusion, which may or may not logically follow from the premises. For example, consider the following syllogism:

 No K are V; Some K are R. Therefore, Some R are V.

Now consider the following counter-example:

No animals are plants; Some animals are cats. Therefore, Some cats are plants.

In this example, the two premises are true, but the conclusion is clearly false, showing that in general the conclusion of the above syllogism does not <u>necessarily</u> follow from the two premises.

Let us look at another example:

Solution

2.	All C are Q;	All Mexicans are North Americans
	Some C are not S.	Some Mexicans are not Canadians.
	Therefore,	Therefore,
	Some S are not Q.	Some Canadians are not North Americans.

Note that the word "some" is given the interpretation "some or all" in the second premise. This wider interpretation is customary in formal logic.

As an aid to finding an appropriate counter-example, it is often useful to form the negation of the conclusion ahead of time. The above problem would then reduce to finding an example for which each of the following statements is true:

All C are Q; Some C are not S. All S are Q (negation of "some S are not Q).

Go to next page.

INVALID SYLLOGISMS

A syllogism consists of two statements called premises, which are assumed to be true, together with a third statement, called the conclusion, which may or may not logically follow from the premises. For example, consider the following syllogisms:

> No K are V; Some K are R. Therefore, Some R are V.

Some K are R, and in particular, K could be the same as R (K = R). Then in this case, "No K are V" would imply "No R are V," refuting the conclusion given.

Let us look at another example:

2. All C are Q; Some C are not S. Therefore, Some S are not Q.

Solution: All C's are Q's but not all Q's have to be C's. Then the S's could be just those Q's that are not C's. (Q = C + S). Then the two premises would be true, but the conclusion false.

Note that the word "some" is given the interpretation "some or all" in the second premise. This wider interpretation is customary in formal logic.

As an aid to finding an appropriate counter-example, it is often useful to form the negation of the conclusion ahead of time. The above problem would then reduce to finding an example for which each of the following statements is true:

> All C are Q; Some C are not S; All S are Q (negation of "some S are not Q").

> > Go to next page.

Refute the following invalid syllogisms, using the method illustrated in the first page. Feel free to refer back to the first page if you find it helpful.

 All J are W; No J are D. Therefore, Some D are W.

4. Some P are M; Some M are not S. Therefore, Some S are P.

5. Some L are E; Some E are O. Therefore, Some O are not L.

Return this part to experimenter when finished to get the remainder of the test.

Refute the following syllogisms. Length of time of work will be used as a secondary measure of the test difficulty, but take time to be careful. Return this sheet to the experimenter when finished.

- 6. Some Y are not I; No Z are Y. Therefore, Some Z are not I.
- 7. Some U are not B; Some P are not U. Therefore, Some P are B.
- 8. All M are X; All G are X. Therefore, Some G are not M.
- 9. No T are N; No F are T. Therefore, Some F are N.
- 10. Some F are H; All H are A. Therefore, Some A are not F.

APPENDIX D

ILLUSTRATIONS

Name

ILLUSTRATIONS

Your task here is to illustrate a number of sets of statements by drawing appropriate diagrams. For example, consider the following set of statements:

1. No K are V.
Some K are R.
No R are V.



Note that in the above diagram, regions K and V are separate, representing "No K are V";

regions K and R have points in common representing "Some K are R"; and that regions R and V are separate representing "No R are V."

Let us look at another example:

Illustration



All C are Q.
 Some C are not S.
 All S are Q.

Note that the word "some" is given the interpretation "some or all" in the second statement above. This wider interpretation is customary in formal logic.

Illustrate the following sets of statements, using the method shown above.

3. All J are W. No J are D. No D are W.

- Some P are M.
 Some M are not S.
 No S are P.
- 5. Some L are E. Some E are O. All O are L.

Return to experimenter and receive continuation of test.

Name_____

ILLUSTRATIONS

Your task here is to illustrate a number of sets of statements by giving appropriate examples. For example, consider the following set of statements:

Illustration

1. No K are V.	No animals are plants.
Some K are R.	Some animals are cats.
No R are V.	No cats are plants.

Note that in the above example words were substituted consistently, that is, each letter stands for one and only one word. In addition, effort was made to form statements that are as clearly true as possible.

Let us look at another example:

2.	All C are Q.	All Mexicans are North Americans
	Some C are not S.	Some Mexicans are not Canadians
	All S are Q.	All Canadians are North Americans.

Note that the word "some" is given the interpretation "some or all" in the second statement above. This wider interpretation is customary in formal logic.

Illustrate the following sets of statements, using the method shown above.

- All J are W. No J are D. No D are W.
- Some P are M. Some M are not S. No S are P.
- 5. Some L are E. Some E are O. All O are L.

Return to experimenter and receive continuation of test.

Continue to illustrate the following by the method shown on the preceding sheet (the one just handed in).

70

- 6. Some Y are not I. No Z are Y. All Z are I.
- 7. Some U are not B. Some P are not U. No P are B.
- 8. All M are X. All G are X. All G are M.
- 9. No T are N. No F are T. No F are N.
- 10. Some F are H. All H are A. All A are F.

