A CHRONOMETRIC ANALYSIS OF LOGICAL INFERENCE

DISSERTATION FOR THE DEGREE OF PH. D.

MICHIGAN STATE UNIVERSITY

DAVID WILLIAM CARROLL

1976

This is to certify that the

thesis entitled

A Chronometric Analysis of Logical Inference

presented by

David William Carroll

has been accepted towards fulfillment
of the requirements for

Ph.D. degree in Psychology

Gordon Wood

Gordon Wood

Major professor

Date November 10, 1976

O-7639

ABSTRACT

A CHRONOMETRIC ANALYSIS OF LOGICAL INFERENCE

By

David William Carroll

A series of studies examined the time needed to generate and then test syllogistic inferences. Major theoretical interest centered on the nature of the internal representation of the inference. The time required to read and integrate the premises (storage time) was measured along with the time needed to compare the inference with a single conclusion (verification time).

The independent variables were the form and content of the syllogistic propositions. Statements were either presented in affirmative ("All A are B") or double negative ("No A are not B") form. It was predicted that the syntactically more complex negatives would take longer to store than the affirmative propositions. If the inference were stored in an abstract semantic format, however, the syntactic complexity of the premises should not affect verification time.

A second variable was the logical complexity of the premises, defined in terms of the number of set relationships that could be generated from the premises. It was anticipated that more errors would occur on the more complex problems. Two models that generated opposite predictions regarding the reaction time effects of logical

complexity were developed. One model assumes that multiple set relations are generated and tested against the single conclusion, and predicts that logical complexity affects both storage and verification stages. An alternate model assumes that the inference representation is unitary: only a single set relation is stored. This model predicts a storage but not a verification effect for logical complexity.

The third variable was truth value. The correct response for any problem was either true or false. It was predicted that true problems would be easier than false problems in both time and errors, but that the truth value factor would not interact with the other factors of interest.

The major findings were:

(1) Syntactically complex negative sentences took longer to store than affirmative sentences but the effects of separate negation factors were not additive. Premises with two negatives were, in fact, slightly easier to process than those with single negatives.

(2) The syntactically complex premises took no more time to verify than syntactically simple premises.

(3) The logical complexity of a problem increased storage time only when the task precluded common errors of interpretation. This form of complexity did not affect verification time.

(4) Problems whose correct response is true were verified faster than false problems. While consistent across experiments, this difference was not influenced by any other factor.

Two conclusions were drawn:

(1) Syllogistic inferences are internally represented as abstract semantic structures. While the form of the premises exerts a large effect on processing time, the inference representation is unaffected by syntactic complexity;

(2) The representation is unitary; single set relations are stored and then tested against the conclusion.

A CHRONOMETRIC ANALYSIS OF LOGICAL INFERENCE

By

David William Carroll

A DISSERTATION

Submitted to

Michigan State University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

Department of Psychology

1976

ACKNOWLEDGMENTS

I gratefully acknowledge the substantial assistance of Dr. Gary M. Olson, particularly for his patience and intelligent advice during the course of this work. I also wish to thank Dr. Gordon Wood, Dr. Lauren Harris, Dr. Lester Hyman, Dr. Donald Johnson, and Dr. Rose Zacks for serving as committee members. Their comments and criticisms have significantly improved the quality of this thesis.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

Page

Chapter 1

INTRODUCTION

This thesis is an exploration of some psychological processes underlying performance in an abstract logical task. In particular, it is an attempt to develop the outlines of a process model of syllogistic inference. The basic strategy is to take a subset of syllogisms and explore performance under different experimental conditions. This strategy contrasts with the practice of many investigators (e.g., Revlis, 1975a; Erickson, 1974) of presenting a large number of different types of problems and developing models that generate predictions for all such problems. In the current work, though a small subset of problems will be used, the investigation will be more intensive than in previous studies.

The fundamental innovation is the use of reaction time (RT) in conjunction with the traditional dependent measure, error rates. RT has been used occasionally, but not analytically, in previous research in syllogistic inference. It has, however, been employed successfully in other areas, such as memory scanning (Sternberg, 1969) and sentence verification (Clark & Chase, 1972). Errors are almost negligible in these experiments, and RT can be a sensitive index of the processes by which such errorless performances occur.

By contrast, investigators of logical inference have no difficulty in discovering errors. Yet it is difficult to discriminate between theories using only error rates since, while errors provide valuable information about the final product of the reasoning process, they are relatively uninformative about the process itself. The present research strategy is to combine RT and error rates as dependent measures. Different experiments measure the time taken for different parts of the total inference process. Thus, RT ought to vary greatly from one experiment to the next. But since the stimuli remain the same, it is assumed that the same inference process is being tapped in each experiment. Hence, error rates should be quite stable.

The organization of this work is as follows. The remainder of this chapter is devoted to an explication of the general structure of an inference model, and one theory is discussed in detail. Attention is then given to the rationale behind the use of RT to infer properties of processing stages, and to some other matters peculiar to logical inference tasks. Chapter 2 presents four experiments that demonstrate the utility of an RT approach, and Chapter 3 builds on these findings to provide evidence for the role of logical complexity in syllogistic inference. A final chapter summarizes the research and its implications.

A Model of Logical Inference: General Structure

As a working assumption, it is reasonable to think of the inference process as a set of steps performed sequentially. The model outlined below is specifically designed to apply to an experiment in which the two premises of a problem are to be compared with a single conclusion, and a single true (valid) or false (invalid) response is required. Extension to situations in which multiple conclusions are presented would not be difficult.

Stage 1: Encoding of first premise

Stage 2: Encoding of second premise

Stage 3: Integration of premises

Stage 4: Encoding of conclusion

Stage 5: Comparison processes

Stage 6: Response execution

Several points are noteworthy. First, the value of this kind of model is in the questions it leads one to ask. The general form of the model (i.e., the order of the stages) is not controversial, but theoretically significant questions may be posed regarding the substance of each of these stages and the extent to which stages are performed separately of one another.

Stages 1, 2, and 4 entail stimulus processing. The term "encoding" is taken to mean a process by which one transforms a physical event into a meaning-based representation: one has encoded a term when one has understood its underlying relations. For syllogisms, these are set relations.

There is not a one-to-one correspondence between proposition type and set relations, but there are constraints on how each of the basic propositions may be encoded. The four standard propositions, together with their names and relations are shown below.

Proposition	Name	Relation(s) of A to B
All A are B	Universal Affirmative	Subset
No A are B	Universal Negative	Disjoint
Some A are B	Particular Affirmative	Overlap, Subset, Superset
Some A are not B	Particular Negative	Overlap, Disjoint Superset

Encoding time differs for the four propositions. Meyer (1973) has found that RTs for universal negatives are longer than for particular affirmatives in a semantic memory experiment. The negation effect was also found in a comparison of universal affirmatives and particular negatives (Meyer, 1975).

Stage 3 combines the information from separate representations into a single representation. The ease with which this process occurs evidently depends on the complexity of the set relations produced by integrating the premises. Ceraso and Provitera (1971) found that problems for which there are multiple and incompatible combinations were among the most difficult problems. Two specific models of integration will be discussed later.

Stage 5 evaluates the fit between the conclusion and the integrated representation stored in memory. The comparison process ultimately depends on how the premise information is encoded. If the information is represented analogically as, for example, a visual image of set relationships, then the comparison process could involve scanning, rotating, or otherwise manipulating such images. Propositional formats similar to the predicate calculus suggest a process such as pairwise comparison of constituents. The present research cannot decide between these alternatives, and emphasis will be on the information contained in a representation rather than its particular form.

Once a decision is reached, a response must be made (Stage 6). This output process is presumably affected by the difficulty and frequency of the response, but not by the semantic content of the information processed in preceding stages.

A Set-Theoretic Model

One example of a stage theory is Erickson's (1974) set-theoretic analysis of syllogistic inference. He assumes that propositions are encoded as set relations. There is no explicit assumption here that Stages 1 and 2 are separate events. Other models (e.g., Revlis, 1975a) explicitly postulate operations made on each premise and thus implicitly assume that these stages are additive. In Erickson's model, integration of premises can occur in at least two different ways. A complete or exhaustive integration model would store as many distinct relations as are computable given the two premises.

Consider the two problems shown in Figure 1. For universal affirmatives, the order of the terms is a critical property since conversion of premises is not logically acceptable (i.e., "All A are B" does not mean the same as "All B are A"). Thus, different orders may be associated with different levels of logical complexity. Transitive orders (those for which the predicate of the first premise is the subject of the second premise) are simpler than non-transitive orders (those for which the predicate of the first premise is the predicate of the second premise). This is because transitive orders generate fewer set relations (see Figure 1). For non-transitive orders, numerous and incompatible relations are possible at

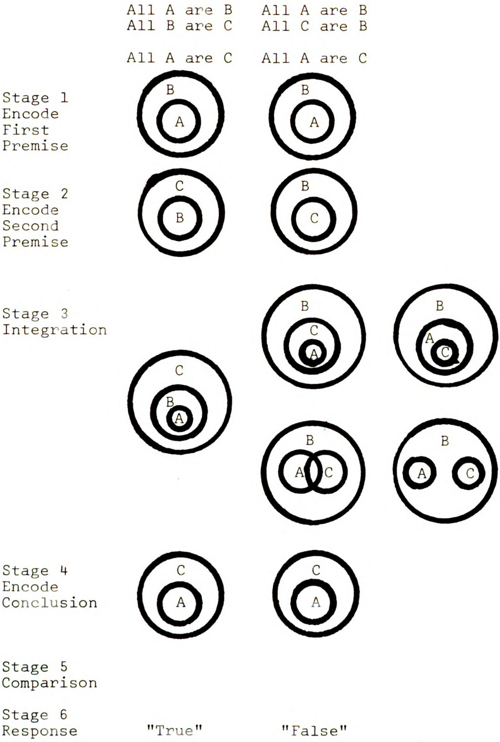


Figure 1. Examples of set-theoretic processing

Stage 3, and a complete integration model would compute and store all of them.

After the conclusion has been encoded, it must be compared with the multiple relations stored at Stage 3. The comparison procedure discussed by Erickson was specifically designed for multiple-choice tests, but the essence of the proposal is that one accepts a conclusion that is consistent with all of the available information. One could either generate a "common denominator" of all the set relations, or, alternatively, compare the given conclusion successively with each of the stored relations. The exact nature of the comparison routine (e.g., self-terminating versus exhaustive) could be empirically determined.

There is considerable evidence that, of the two problems shown in Figure 1, the transitive problem on the left is decidedly easier than the non-transitive problem on the right. In general, problems with a valid conclusion are much easier than problems without a valid conclusion (Begg & Denny, 1969; Ceraso & Provitera, 1971; Chapman & Chapman, 1959; Erickson, 1974; Revlis, 1975b; Roberge, 1970). In studies in which single conclusions are evaluated, the former problem is solved about 90% of the time, while the latter is solved only about 50% of the time (Sells, 1936; Traub & Erickson, 1975; Wilkins, 1928).

With one modification, Erickson's original version of the complete integration model can account for these error effects. Erickson assumed that the conclusion accepted at Stage 5 had to be consistent with _all_ of the combinations generated during Stage 3. This model, however, would predict no errors on the False Non-Transitive, since there is no single conclusion that is consistent with all relations. If, instead, one assumes that the comparison process is more difficult (i.e., more likely to be erroneous) when there are numerous set relations, then the complete integration model can handle the results noted above. This is a reasonable assumption given Ceraso & Provitera's (1971) finding that set complexity is a determinant of syllogistic performance.

In contrast, an _incomplete_ integration model would store only one set relation and compare this directly to a conclusion or set of conclusions. This one relation could be either chosen at random (Erickson, 1974), or so as to be a prototype of all possible relations. In the False Non-Transitive shown in Figure 1, a suitable prototype would be the disjoint relation in which A and C are distinct subsets of B. With regard to the conclusion presented, this one relation can serve as the basis for (correctly) rejecting the conclusion "All A are C." The other relations may be _generated_, but it is the essence of the incomplete integration model that only one relation is ultimately stored. For the incomplete model, the errors made on the False Non-Transitive problem would be due to an incorrect inference drawn during Stage 3.

In sum, these two models differ in their characterization of how an inference is internally represented. The complete integration model assumes that this representation is complex; the incomplete integration model assumes that is is unitary.

The data collected by Erickson and others do not permit a secure assessment of the relative merits of these two proposals. For example, Erickson (1974) re-examined the data of Ceraso and Provitera and discovered that the complete model predicted 89% of judgments against 85% for the incomplete model. It will be shown below that a reaction time method permits a direct test of the two models.

Reaction Time Analysis of Logical Inference

Time is infrequently used as a dependent measure in problem solving research, though there are noteworthy exceptions (e.g., Trabasso, Rollins, & Shaughnessy, 1971; Johnson & Jennings, 1963). The specific use of the additive factor method described below has been limited to simple tasks, such as addition (Parkman & Groen, 1971), search of well-learned lists (Shanteau & McClelland, 1975), or three-term series problems (Clark, 1969). In syllogistic inference, studies that have used time as a dependent measure have used total response time to complex problems as the unit of analysis rather than attempting to decompose RT into components (Erickson, 1972; Frase, 1966, 1968; Lippmann, 1972; Parrott, 1967, 1969; Traub & Erickson, 1975). The present research attempts to use an analytic technique found to be useful in studies of simpler cognitive acts.

Additive Factor Analysis. A method for inferring the existence and characteristics of processing stages has been developed by Sternberg (1969). The rationale is to conceive of processing in any task as a series of stages, each of which takes time. One seeks independent variables or factors likely to increase the duration of a stage. Hence, if two experimental factors each exert a significant effect on RT, but do not interact, one may infer that these factors influence different stages. An example by Sternberg (1969) is a short-term memory task in which subjects memorize a series of digits and then are later presented with test digits, one at a time. Their task is to decide whether each digit is one of those originally presented. The time taken to make this decision is recorded. Sternberg has found that RT varies with the size of the original set of digits and he attributes this effect to a process by which presented digits are compared with memory items. If the stimulus materials are physically degraded, RT is lengthened but this factor does not interact statistically with the set size factor. Thus, Sternberg concludes that at least two separate stages operate here, and that their effects are additive. There is no empirical basis for ordering the stages, but logical considerations dictate that stimulus processing must precede comparison.

The rationale of the additive factors method thus is to search for experimental factors that exert additive effects on RT. Additive factors are assumed to affect different processing stages, while interacting factors must influence at least one stage in common.

The interpretation of RT in a task that produces high error rates must be cautious. Two specific points may be raised here. First, as with most RT work, only latencies for correct responses will be examined. Such responses will be less frequent in a harder task, so there will be more missing observations. This forces some changes in the usual methods of data treatment, to be discussed below. Secondly, it is important to show that the high error rates reflect something more than speed-accuracy tradeoffs (discussed by Pachella, 1974). Errors must be shown to be related to the structural complexity of the materials, and it must be demonstrated that RT differences between experimental conditions are not merely due to such tradeoffs.

Present analysis and hypotheses. Five experimental factors are used in the present series of studies. Three are concerned with the form of the syllogistic propositions, while two are content variables. The form variables are whether the first premise, second premise, and conclusion are stated affirmatively or negatively. Negation is expressed by the double negative ("No A are not B"). Negation is purely syntactic in that the propositions stated in affirmative and (double) negative terms have identical underlying set relations. Despite the synonymity of the terms, negatives are likely to take longer to encode than affirmatives (Clark, 1974) and, furthermore, more words need to be read. This becomes our first hypothesis.

> Hypothesis 1. Negative premises take longer to encode than affirmative premises.

This hypothesis is not controversial. The use of three negation factors, however, allows one to consider whether the encoding of each proposition is performed in additive fashion.

A separate question about the processing of form information deals with the effect of syntactic complexity on inference representations. Examination of verification latencies is made for problems originally presented in simple (affirmative) or complex (negative) terms. According to the principle of congruence (Clark, 1974), representations that are more congruent in form take less time to compare than less congruent forms. If the stored representation preserves syntactic information, then negative premises should be most easily compared with negative conclusions and affirmative premises most readily with affirmative conclusions. Alternatively, if the memory structures are more abstract, then how the premises are expressed should not affect verification latencies. While there is no literature in syllogistic inference relevant to this question, pertinent data from other tasks suggest that abstract representations are the rule in several domains. Kintsch and Monk (1972) gave subjects short texts to read, and later timed their latencies to answer questions based on the texts. No response differences were found with syntactically simple and complex texts. Kintsch and Monk (1972) argued that the passages were stored in a form more abstract than the actual wording. Supportive findings have been found in the problem solving domain by King and Greeno (1974), who found no response differences for simple and complex problems, and Potts and Scholtz (1975), who found no verification effects for marked and unmarked terms in a three-term series study.

One qualification on these conclusions concerns the time elapsed since an event is first encountered. In each of these studies, the time between initial registration and later verification was relatively long. For example, in the Potts and Scholtz study, the time taken to read and understand statements such as "A is better than B, B is better than C" was about ten seconds. If either the inspection time for the initial information or the interval between study and test is shortened, the memory trace for initial events may be more "verbatim" than "propositional" (Anderson, 1974; Kintsch, 1974).

In the present series of studies, subjects were either given a long time to read and understand the premises or were in control of this interval. Under each of these circumstances, the effect of premise negation on verification time should provide evidence as to the abstractness of the inference representation. Thus, based on previous work we can generate a second hypothesis.

> Hypothesis 2. The time needed to verify a conclusion is unrelated to the syntactic form of the original premises.

While error patterns related to syntactic form will be discussed briefly below, they are not of central theoretical interest.

One of the two content factors is truth value. Conclusions are said to be true if they are valid inferences from the premises, and false otherwise. It seems reasonable that true responses will be faster than false responses, since response type often affects RT. Such an effect could be due to either comparison processes or response processes. Thus, we have Hypothesis 3.

<u>Hypothesis 3</u>. Response time for true problems is less than for false problems.

Main effects of truth value, however, are of less interest than interactions between truth value and other factors. Failure to find such interactions means that truth value affects stages unaffected by negation and transitivity.

The content factor of central interest is logical complexity, defined in terms of the complexity of the set relationships. This study uses only universal affirmative propositions, though their form may be either affirmative or negative. For universal affirmatives, transitive orders are logically simpler than non-transitive orders. Since non-transitive problems generate more possible set relations, the complete integration model predicts that non-transitive problems would take longer to integrate than transitive problems. The incomplete integration model assumes only one relation is stored, but non-transitive problems would still require more integration time since there are more relations to choose from. Thus, both models are compatible with the following hypothesis.

> Hypothesis 4. Non-transitive problems take longer to integrate than transitive problems.

If a number of relations are stored in memory, then each has to be considered in the verification process. Thus, the complete integration model would predict longer verification times for non-transitive problems.

> Hypothesis 5a. Non-transitive problems require more time during the comparison process than transitive problems.

It is important to note that the incomplete integration model does not make this prediction. If Stage 3 representations are unitary rather than complex, then transitive and non-transitive problems should be equivalent in comparison time. Thus, whether logical complexity affects Stage 5 is a crucial test of the two positions.

To insure that Stage 5 differences, if found, are due to logical complexity rather than to true-false response time differences, the additional problems shown in Figure 2 will be employed. The problem on the left (False Transitive) is similar to the True Transitive, but the inference here is not valid. The True Non-Transitive (on the right) requires extensive Stage 3 processing, but the conclusion merely repeats one of the two premises and thus should cause no special difficulty. Since the True Non-Transitive does not

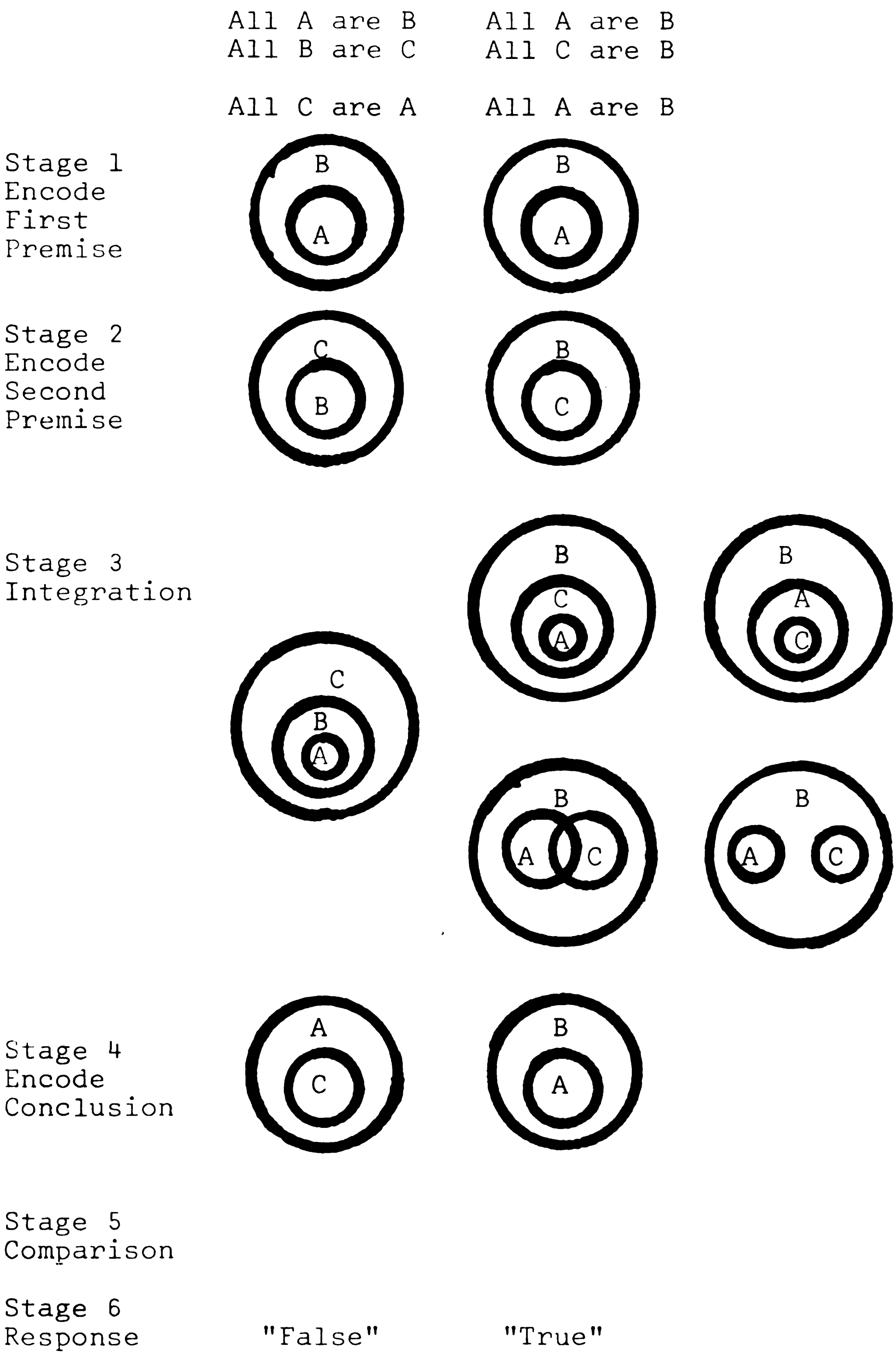


Figure 2. More examples of set-theoretic processing

require a correct encoding of the A-C relationship, the set relations, while numerous, are fully consistent with one another. The complete integration model therefore predicts that the comparison process would not be so difficult as when, with the False Non-Transitive, the relations are inconsistent. This becomes the second part of Hypothesis 5.

> Hypothesis 5b. The increment for non-transitive problems during the comparison process, relative to transitive problems, should be greater for false than true items. Truth value and transitivity should interact.

Once again, the incomplete integration model does not make this prediction.

Throughout this discussion, the assumption has been made that transitivity RT effects may be attributed to logical complexity differences between transitive and non-transitive problems. An alternate hypothesis is that these effects are surface phenomena: some orders are easier to process than others. Huttenlocher (1968) has considered an order-based strategy for the solution of three-term series problems. In the syllogistic literature, much discussion has centered on the role of the order of the terms in the premises (technically known as problem "figure"). Problems with the transitive order or first figure (A-B, A-C) are generally easier than those with the non-transitive order or second figure (A-B, C-B). However, the results are generally small and somewhat inconsistent when letters (as opposed to words) serve as categories (Frase, 1968; Pezzoli & Frase, 1968; Roberge, 1971).

In any event, it is important to distinguish between order *per se* and the logical complexity that is signified by order as explanations for RT differences. The approach taken to this problem is to compare RT patterns when order is logically significant and when it is not. This control is developed more fully in Chapter 3.

A final hypothesis is concerned with error rates. As noted above, false problems are generally harder than true problems. Further, the data of Ceraso and Provitera (1971) and Erickson (1974) demonstrate that logical complexity affects error patterns. Taken together, these considerations lead to Hypothesis 6.

> *Hypothesis 6*. Errors should be most frequent for the False Non-Transitive, least frequent for the True Transitive, and intermediate for the False Transitive.

The True Non-Transitive should be easy, since it merely repeats a premise. No explicit comparisons will be made between this problem and the more difficult inference problems.

Non-Logical Processes

Most people do not spend their spare time solving logic problems. When asked to use their logical abilities in a psychological experiment, it may be difficult to immediately adopt the proper mode, which involves restricted meanings of certain terms as well as stringent criteria for the acceptability of arguments. One's logical knowledge may be thought of as a well-structured but seldom-used system of concepts that is internally consistent yet difficult to access. If so, a cursory examination of logical performance may seriously underestimate an individual's logical competence because too little attention has been given to the difficulties one has in orienting to complex, abstract, and unfamiliar tasks.

These considerations assume greater importance when one realizes that logical factors such as set complexity are partially confounded with other, non-logical factors. In Ceraso and Provitera's (1971) experiment, for instance, the most difficult type of problem was one for which multiple and inconsistent set relations could be generated by combining the premises. For such premise pairs, all possible inferences are false or invalid. This creates a difficulty in disentangling errors due to logical structure (e.g., coping with inconsistent input) versus those resulting from response factors (e.g., a tendency to assume a conclusion is true unless proven false). Attention must therefore be paid to the ways one can identify and control various non-logical factors that may contribute to error patterns.

General Response Bias. Subjects have a general bias to verify rather than falsify the conclusions they are asked to evaluate (e.g., see Wason & Johnson-Laird, 1972). Since most studies in the syllogistic literature have used a large percentage of false problems (e.g., Begg & Denny, 1969; Chapman & Chapman, 1959; Roberge, 1970; Sells, 1936), this kind of bias is likely to have flourished in previous work. A related difficulty is the tendency to use "probabilistic" reasoning, in which conclusions are accepted on the basis of being probable rather than necessary (Chapman & Chapman, 1959). These general biases have the effect of inflating observed error differences between true and false problems.

The present approach to this problem is threefold. First, of the problems presented, half of the conclusions are true and half are false. Secondly, subjects are informed of this fact. Thirdly, subjects are told to reserve the true response to conclusions that must be true. As noted by some other investigators (e.g., Revlis, 1975a), these conditions are the exception, not the rule, in studies of syllogistic inference.

Matching Bias. The atmosphere hypothesis of Woodworth and Sells (1936) contends that responses in logical tasks are made on the basis of superficial and logically irrelevant stimulus features. Conclusions such as the following are often accepted.

All A are B
<u>All C are B</u>
All A are C

Woodworth and Sells argued that the conclusion is incorrectly accepted because it is expressed in the same affirmative form as the premises. Similar findings in propositional reasoning have been extensively documented (Evans, 1972a, b, c; Evans & Lynch, 1973). Since previous studies have routinely used affirmative sentences to express inclusion relations, the relative contributions of logical and surface factors are not clear.

One way to deal with this problem is to formulate a more precise matching bias model and test its predictions (Revlis, 1975a). The general procedure is to devise rules for combining features (e.g., + Universal, + Affirmative) from two premises into a single representation. One such rule states that the composite structure (Stage 3) is chosen simply to match the features of the two premises.

A problem with this type of approach is that by dealing with standard syllogistic propositions, one cannot manipulate "features" without also altering set relationships. A feature is not simply a surface factor but is also (imperfectly) correlated with the underlying logical structure. While features correspond in some way to logical concepts, their exact conceptual status is unclear. For example, the differences in set structure between universal affirmatives and universal negatives is distinct: one is encoded as a set inclusion structure, and the other as two disjoint sets. In contrast, particular affirmatives and particular negatives have, under one interpretation, the identical set diagram, one of overlapping sets. The meaning of a "feature" may depend on context.

A less ambiguous way of dealing with matching biases is to make the form of the proposition independent of its content. This is the approach taken here. As noted above, propositions will be expressed in either affirmative or double negative form. The strength of these biases can be determined independently of the logical structure of the propositions, and differences between logical structures, when collapsed over form variations, will not be contaminated by non-logical strategies.

Premise Misinterpretation. Recently, many investigators have concluded that virtually all errors in logical tasks stem from a misunderstanding of the premises, with correct reasoning from that point on. In propositional reasoning, for instance, the tendency to misinterpret conditional statements is well documented (Taplin, 1971; Taplin & Staudenmayer, 1973; Staudenmayer, 1975). Syllogistic propositions are ambiguous as to their set relations, and this has been argued as a source of difficulty by several authors (Ceraso & Provitera, 1971; Erickson, 1974; Neimark & Chapman, 1975).

A different kind of misinterpretation that can occur with universal affirmatives is to view them as symmetrical rather than asymmetrical. Thus, "All A are B" can be converted into "All B are A." If the statements are interpreted as equivalences rather than as set relations (Chapman & Chapman, 1959), then it is natural to convert propositions (Revlis, 1975a, b) and to give true responses to false conclusions. Since the standard propositions allow for such misinterpretations, disambiguation can be achieved by altering the propositions. Direct manipulation of interpretation is done in Experiments 5 and 6, which compare performance with statements of the form "All A are included in B" and "All A are equal to B."

Statement of the Problem

Prior to analyzing RT patterns, it is necessary to provide evidence that errors are more frequent for logically complex problems than for logically simple problems. This entails a replication of the basic results of Ceraso and Provitera (1971) in a context in which the contributions of various non-logical processes may be assessed. Once it has been determined that logical structure affects error patterns, analysis of the time course of syllogistic inference will be performed using the additive factors method. The present research attempts to determine: (1) What factors are responsible for observed error differences between problems; and (2) To the extent that the differences are attributable to structural factors, what processes mediate these differences? Here, interest is centered on (a) the nature of the internal representation of the inference, and (b) the nature of the processes that generate and operate on this representation.

Chapter 2

REPRESENTATION AND PROCESSING OF SYNTACTIC COMPLEXITY

The results of four experiments will be presented in this chapter. Together, these studies identify several sources of error effects but indicate that at least some differences are due to logical structure. These differences are sensitive to the level of practice. Further, the results show that syntactic and truth value factors affect RT in predictable ways, but differences in logical complexity are absent. Before discussing the findings in detail, the general procedures will be described.

General Method

Materials

The problem set used for Experiments 1-4 is shown in Table 1. Each of the four problems is expressed in eight different forms. In all of the studies, a constant set of alphabetical letters (A, B, C) served as category names. Propositions were typed, photographed, and mounted on slides.

Problems were defined in the following ways. Transitive problems are those in which the predicate of the first premise is the subject of the second premise. Non-transitive problems are those in which the predicate of the first premise is also the predicate of the second premise. True problems are either valid

Table 1

Problems Used in Experiments 1-4

Form	Premises: Transitive	Premises: Non-Transitive
Affirmative Affirmative	All A are B All B are C	All A are B All C are B
Affirmative Negative	All A are B No B are not C	All A are B No C are not B
Negative Affirmative	No A are not B All B are C	No A are not B All C are B
Negative Negative	No A are not B No B are not C	No A are not B No C are not B

Form	Conclusions: True	Conclusions: False	Conclusions: True	Conclusions: False
Affirmative	All A are C	All C are A	All A are B	All A are C
Negative	No A are not C	No C are not A	No A are not B	No A are not C

inferences from the two premises or simple repetitions of one of the premises. False problems are those in which the conclusion cannot be validly derived from the premises. Truth value is thus a relation between premises and conclusion, not a type of premise combination. By treating logical complexity and truth value as separate factors, one can present problems that have complex premises yet simple answers (e.g., True Non-Transitive). This allows consideration of the role of logical complexity on integration processes separately from the effect of truth value on comparison and response processes.

Problems were chosen to permit simple and direct tests of the hypotheses of interest. The logical distinction investigated, simple versus complex set relations, would seem to be well within the grasp of most college students. More elaborate notions of logical complexity will ultimately be necessary, but since there is no available evidence of the role of this kind of factor in reaction time tasks, it seemed more reasonable to start with a basic distinction.

Another consideration was how the logical distinction was realized. With universal affirmative propositions, non-transitive premise orders serve to introduce a more complex relationship between the premises, as compared to transitive orders. The value of this simple arrangement is that it becomes possible later on (see Chapter 3) to control for the order of the terms of the problem. This is accomplished by using transitive and non-transitive orders in equivalence problems. Here, no RT or error effects are expected.

Procedure and Design

Unless otherwise indicated, the following details apply to all experiments except the first one.

Subjects (Ss) were tested individually in a semi-darkened room. Slides were projected onto a white screen. The displays subtended 2° 21' horizontal angle and 1° vertical angle at a distance of five feet.

A control panel with one button and two telegraph keys was directly in front of S. S started a trial by pressing the button on the control panel, and made decisions by depressing one of two telegraph keys marked as "true" or "false." One index finger rested on each key. Labelling of true and false was done so that half of the subjects received true as their preferred hand, and half received false as their preferred hand. Hand assignments were constant across experimental sessions.

Two Hunter timers controlled the time sequence of events in the experiments, and RTs were measured in milliseconds by a Hunter Klockcounter and recorded manually by E. S was given verbal feedback on both time and accuracy after each trial. During the practice sessions, problems in which S was erroneous were reshown but no explanation of the problem was given. Then the apparatus was reset, indicating that the next trial could be initiated. E was present during the entire session.

The instructions emphasized accuracy rather than speed. Ss were told to proceed as rapidly as possible without sacrificing accuracy. They were instructed to reserve the true response for those problems that were necessarily true and respond false otherwise. Ss were told that half the problems were true and half were false.

At the end of the experimental session, Ss were asked to fill out a short questionaire, indicating the strategies used during the course of the experiment. If the answers given were unsatisfactory, E probed for more revealing statements. The experimental session lasted approximately 45 minutes.

Predictions

RT predictions will be made for the model outlined earlier, and then applied to each successive situation. A negative form should take longer to read and encode than its affirmative counterpart, adding an increment at Stages 1, 2, and 4. Transitive problems should take less time during Stage 3 than non-transitive problems. If the incomplete integration model is correct, no other transitive effects are expected, but the complete integration model also predicts an increment during Stage 5 for non-transitive problems, particularly for the False Non-Transitive. Either model would predict a response time difference between true and false problems.

Experiment 1 is a preliminary experiment, using only error data. Experiment 2 employs a sequential presentation procedure, in which the two premises are presented together for ten seconds, followed later by the conclusion. A main effect of conclusion negation and a non-interacting truth value effect are expected here. Experiment 3 uses a simultaneous presentation of all three propositions. Main effects for each of the three negation factors are expected, along with a main effect for transitivity, and one for truth value. Experiment 4 is a strong test of the additivity of stages of the proposed model. The two premises are simultaneously presented, in a self-paced task, for as long as the subjects need, followed by the conclusion, to which a response is made. The time to store the two premises ought to reflect the two premise negation effects found in Experiment 3, along with the transitivity effect. The time to verify the conclusions should replicate Experiment 2; there should be strong effects for conclusion negation and truth value. In addition to these effects, the complete integration model also predicts both a transitivity main effect, and a transitivity by truth value interaction in the verification times of Experiments 2-4.

Experiment 1

This experiment was designed as an exploratory study. The intent was primarily to see whether prominent differences in performance were related to the form and content variables employed. Only two problems were used, True Transitive and False Non-Transitive. The purpose was to compare true and false problems, and the transitivity factor did not become apparent until after the experiment was completed.

Method

Subjects. Eighty-eight introductory psychology students were given course credit for their participation. Subjects were tested individually or in small groups.

Materials. Of the problems shown in Table 1, only the True Transitive and False Non-Transitive were used. Sixteen other problems were given to the subjects, but these are not germane to the present discussion. Problems appeared on a printed sheet. A separate answer sheet was provided.

Procedure. Two lists differing only in the order of the problems were constructed and *S*s were randomly assigned to one of the two lists. *S*s were given as much time as they needed to solve the problems.

Results

The dependent measure of interest was the percentage of correct responses for each condition, presented in Table 2. A four way analysis of variance including the three negation factors (first premise, second premise, and conclusion) and problem type (True Transitive, False Non-Transitive) was carried out. Several significant effects were found and they fall into three main categories.

First, there was a significant main effect for problem type, $\underline{F}$ (1, 87) = 5.5, $\underline{p}$ < .025, with more correct responses for the True Transitive (.77) than for the False Non-Transitive (.71).

A second category of findings indicates matching biases. Here, three results were significant. First, the first premise x conclusion x problem type interaction was significant, $\underline{F}$ (1, 87) = 11.5, $\underline{p}$ < .001. The critical comparison is between problems in which the first premise and conclusion match in form (either affirmative or negative) versus those in which there is a mismatch. For true items, match problems were easier than mismatch problems; for false problems, the reverse was true. Taken together, these results imply a single principle: match problems receive more true responses than mismatch problems. Since this strategy produces opposite results (in terms of accuracy) for true and false problems, one finds an

Table 2

Per Cent Correct Response for Eight Variations of Each Problem in Experiment 1

	Premise Combination								
Problem	AAA	ANA	NAA	NNA	AAN	ANN	NAN	NNN	Total
True	.91	.73	.67	.70	.77	.76	.81	.78	.77
False	.63	.81	.78	.76	.74	.64	.64	.68	.71
Total	.77	.77	.72	.73	.75	.70	.72	.73	.74

with problem type, but it is clear that similar processes occurred in the true and false problems. This latter point is shown in Table 3. Whether the match is defined over the first premise and conclusion, or second premise and conclusion, or all three propositions, the result is the same and equally valid for true and false problems. The latter two interactions were also significant. However, when the match was between the first premise and the second premise, no effect was found.

A final category of results proved less easy to interpret. A significant first premise x problem type interaction, F (1, 87) = 3.9, $p < .05$, indicated that the true-false difference was somewhat larger for affirmative first premises than for negative first premises. It was also larger for negative conclusions than affirmative conclusions, $\underline{F}$ (1, 87) = 5.4, $\underline{p} < .025$. The explanation of these results is not known.

Discussion

This experiment established several conclusions of interest. First, experimentally separating logically relevant and irrelevant dimensions showed that each influences the difficulty of syllogistic inference. The latter results indicate the existence of matching biases similar to those discussed by Woodworth and Sells (1935) and Evans (1972a). Apparently, Ss resort to a strategy of the form, "If problem components match, respond 'true'." Such a strategy distorts the true difference between the two problems. If one considers only affirmatively-stated problems, the true-false difference is 28%; overall, the difference is much smaller (6%) but still significant.

To what extent are these remaining truth value differences due to a general response bias? With one study, it is difficult to say, but in combination with another very similar experiment, we can conclude that these effects are minimal. A study identical to Experiment 1 in materials but differing in instructions was performed. Ss were not told that half of the problems were true and half were false. The true-false

Table 3

Interactions Found in Experiment 1

Match	Condition	True	False	p
First Premise and Second Premise	AA	.84	.68	< .25
	AN	.74	.72	
	NA	.74	.71	
	NN	.74	.72	
	Match	.79	.70	
	Mismatch	.74	.72	
First Premise and Conclusion	AA	.82	.72	< .001
	AN	.77	.69	
	NA	.69	.77	
	NN	.80	.66	
	Match	.81	.69	
	Mismatch	.73	.73	
Second Premise and Conclusion	AA	.79	.70	< .05
	AN	.79	.69	
	NA	.72	.78	
	NN	.77	.66	
	Match	.78	.68	
	Mismatch	.75	.74	
First Premise and Second Premise and Conclusion	AAA	.91	.63	< .025
	ANA	.73	.81	
	NAA	.67	.78	
	NNA	.70	.76	
	AAN	.77	.74	
	ANN	.76	.64	
	NAN	.81	.64	
	NNN	.78	.68	
	Match	.85	.65	
	Mismatch	.74	.73	

differences observed, when collapsed over form, were large (true = .85, false = .33). Other aspects of the data were comparable to Experiment 1. Thus, the instructions used in Experiment 1 seemed to have eliminated most of the general truth value effects. The remaining differences between the True Transitive and False Non-Transitive are likely due to specific content factors--transitivity, truth value, or both--and the remaining studies thus probed the effects of these content factors on error patterns.

Experiment 2

This experiment was designed to confirm the results of Experiment 1, and to assess performance when another factor (transitivity) and another dependent variable (RT) are introduced. The paradigm was a sequential one in which the first and second premises appear visually on a screen for ten seconds and a single conclusion is subsequently evaluated as true or false, with response time being recorded.

Method

Subjects. Thirty-one introductory psychology students were given course credit for their participation. Subjects were tested in three experimental sessions, usually two days apart.

Materials. The problem set is shown in Table 1. In addition to these 32 problems, 32 filler problems dealing with universal negatives were presented and served to keep Ss honest in processing the double negatives.

Two lists differing only in the order of the problems were constructed, and each list consisted of four blocks of 16 problems. Ss were randomly assigned to lists each session and to block orders within each session. One block served as practice.

Procedure. Procedural details were identical to those discussed under the heading "General Method", with the following exceptions. Premises and conclusions were presented sequentially. Premises appeared simultaneously on the screen for ten seconds, and one and one half seconds intervened between slides. The clock-counter began with the appearance of the conclusion and was terminated by S's response. Each S saw each one of the 64 problems once in each session.

Results

Error and RT data for correct decisions are presented in Table 4.

Errors. Error patterns on form factors resembled those found in Experiment 1. Since primary interest was concerned with the role of truth value and transitivity effects on error rates, data were collapsed on form and subjected to a 3 (Problems) x 3 (Days) analysis of variance. The True Non-Transitive,

Table 4

Mean Reaction Times[a] and Percentage of Errors[b] in Experiment 2

	Premise Combination								
Problem	AAA	ANA	NAA	NNA	AAN	ANN	NAN	NNN	Total
True Transitive	1.13 (17)	1.39 (25)	1.30 (25)	1.39 (30)	1.72 (27)	1.65 (24)	1.77 (30)	1.84 (26)	1.53 (25)
False Transitive	1.50 (40)	1.52 (36)	1.57 (27)	1.50 (34)	2.05 (24)	2.13 (31)	2.09 (31)	2.12 (26)	1.81 (31)
True Non-Transitive	1.30 (0)	1.58 (11)	1.52 (13)	1.62 (16)	1.84 (18)	2.05 (16)	2.19 (13)	1.95 (15)	1.76 (13)
False Non-Transitive	1.67 (34)	1.59 (37)	1.91 (32)	1.59 (27)	2.24 (29)	2.10 (40)	2.16 (39)	2.02 (41)	1.91 (35)

[a] In seconds

[b] In parentheses

obviously very easy, was not included in the analysis.

The Problems effect was marginally significant, $\underline{F}$ (2, 60) = 2.39, $\underline{p}$ < .10. The ordering of problems in difficulty conformed to predictions--True Transitive, False Transitive, False Non-Transitive, from easier to harder--though the differences fell short of significance.

The Days effect was significant, $\underline{F}$ (2, 60) = 17.76, $\underline{p}$ < .001. Performance improved consistently over

Table 5

Error Scores for Each Problem for Each Day

Problem	Day 1	Day 2	Day 3
True Transitive	.29	.25	.21
False Transitive	.45	.30	.18
False Non-Transitive	.40	.35	.29

sessions. Days 3 (.77) and 2 (.70) were superior to Day 1 (.62), and Day 3 was also superior to Day 2. All Newman-Keuls comparisons were significant ($p < .05$).

In addition, the Problems x Days interaction was marginally significant, $F(4, 120) = 2.39$, $p < .06$. Table 5 shows the data for each problem as a function of session. The major point of interest is that the rate of performance rise is sharper for the False Transitive than for the other two problems. At Day 1, the two false problems are roughly equal in difficulty, but by Day 3 it is the two transitive problems that are most comparable.

Reaction Time. A considerable amount of variability existed in this paradigm. To obtain reliable RT results, several steps were taken. First, scores on Day 1 were discarded. Second, all scores more than three standard deviations greater than the mean for that condition were discarded. This step, in combination with some subjects' failure to solve certain problems on either of the last two sessions, created many missing observations, 120 of a possible 992. The mean of each subject's overall scores was inserted into each of his missing observations. Twenty-one of 31 subjects had at least one missing data point. Since this method of data treatment tended to deflate within-subject variability, a stringent significance level (.01) was adopted.

The mean correct RT results are shown in Figures 3 and 4. For each of the four graphs, the four premise combinations are plotted separately for affirmative and negative conclusions. A 2 (First Premise) x 2 (Second Premise) x 2 (Conclusion) x 2 (Truth Value) x 2 (Transitivity) within-subject analysis of variance disclosed a strong effect for the polarity of the conclusion (affirmative = 1.51 ± 0.10 seconds, negative = 2.00 ± 0.13 seconds; F (1, 30) = 64.08, $p < .001$). Here and elsewhere, results are presented as means plus or minus one standard error. The negation factor did not interact with any other factor.

Truth value also affected RT (true = 1.64 ± 0.10 seconds, false = 1.86 ± 0.13 seconds; F (1, 30) = 12.54, $p < .001$). A trend, not significant by stringent standards, indicated that transitive latencies were somewhat faster than non-transitive latencies,

Figure 3. Reaction times for true transitive (top) and true non-transitive problems (bottom) as a function of premise and conclusion form in Experiment 2

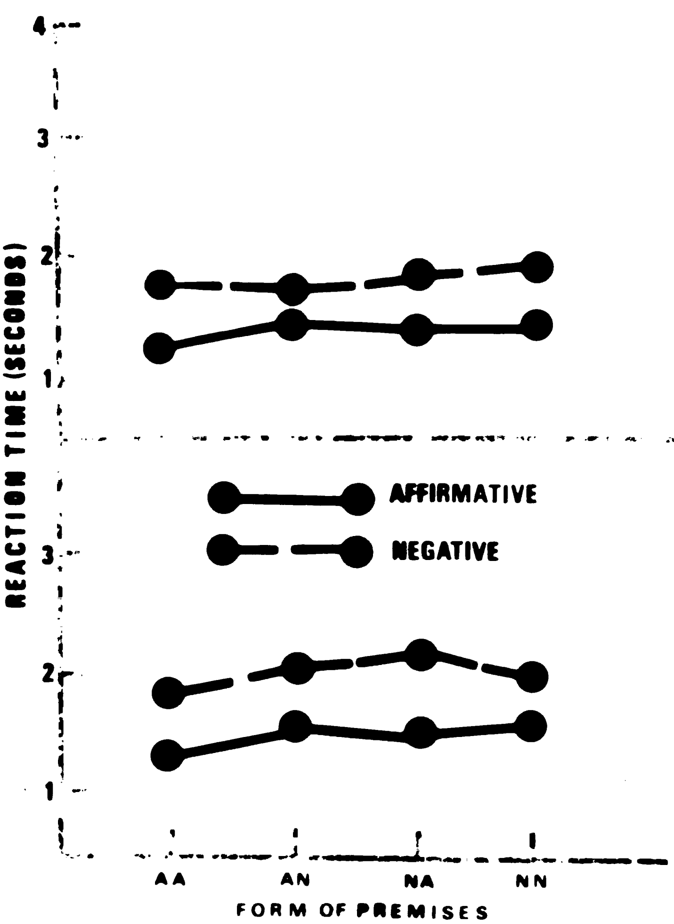

REACTION TIME (SECONDS)
4
3
2
1
3
2
1
AFFIRMATIVE
NEGATIVE
AA
AN
NA
NN
FORM OF PREMISES

FIGURE 3

Figure 4. Reaction times for false transitive (top) and false non-transitive problems (bottom) as a function of premise and conclusion form in Experiment 2

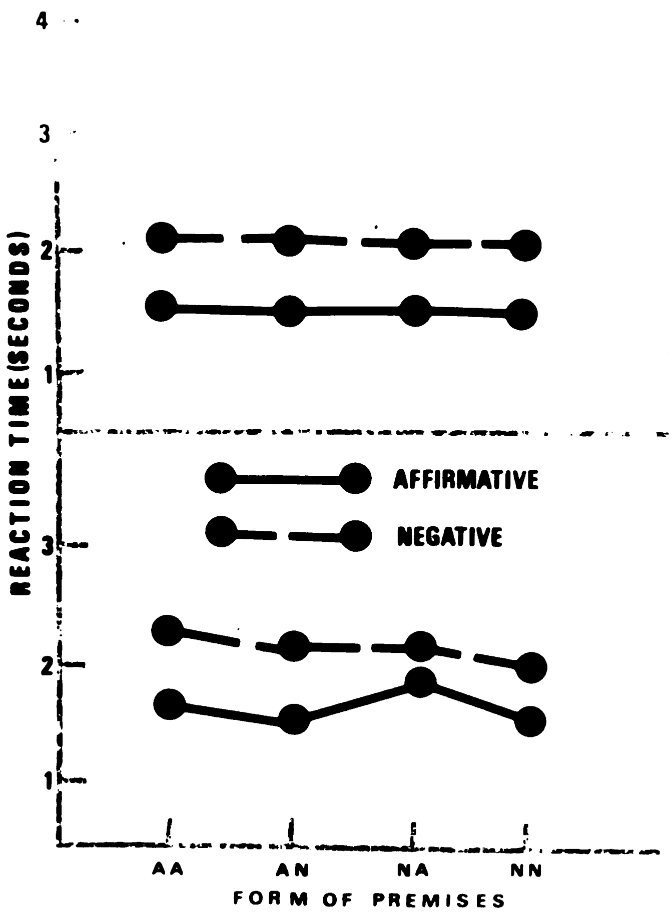

4
3
2
1
3
2
1
REACTION TIME (SECONDS)
AFFIRMATIVE
NEGATIVE
AA
AN
NA
NN
FORM OF PREMISES

FIGURE 4

F (1, 30) = 7.13, $p < .025$. The truth value by transitivity interaction, however, did not approach significance.

Discussion

Additivity of Factors. Stages 4, 5, and 6 are presumably tapped in this experiment. The evidence is that two strong effects and one weak one occurred, yet no interactions were obtained. The simplest conclusion is that the negation factor increased the duration of the conclusion encoding processes (Stage 4), while the truth value effect is due to response processes (Stage 6). The transitive effect, if reliable, would be attributed to comparison processes (Stage 5). This latter finding, however, should be interpreted with some caution, since the effect is weak and not consistent with either of the proposed models. The incomplete integration model predicts no verification differences attributable to the transitivity factor, while the complete integration model predicts an interaction with truth value as well as a main effect. No interaction was found. In general, these results are consistent with a model that assumes three separate processing stages as components of verification latencies.

Stage 3 Representation. Experiment 2 provides clear evidence that the syntactic complexity of the two premises does not affect the manner in which the inference is stored, confirming several earlier reports (Kintsch & Monk, 1972; King & Greeno, 1974; Potts & Scholtz, 1975). If the syntactic information formed an important part of the integrated Stage 3 representation, then one would expect to find interactions between conclusion form and premise form, yet none were found. This does not necessarily mean that no form information has been retained, but only that the representation upon which a decision is based is not different for different form presentations.

Error Results. The error patterns suggest that there are two major components of error differences, truth value and transitivity, though these differences are small. Interestingly, the truth value factor seems to be the more important factor in Ss with little practice, but the transitivity factor seems to dominate with more practiced Ss. This result suggests that early in the experiment subjects are likely to misinterpret the propositions as symmetrical statements. This error would make the false items especially difficult, and minimize the predicted difference between the two false items.

Experiment 3

Experiment 2 was unable to provide any direct evidence regarding the manner in which the independent variables affect Stages 1, 2, and 3 of the proposed model. Experiment 3 was designed to provide this information. The paradigm was a simultaneous RT task, in which all three propositions are shown at the same time and a single true or false response is required.

Method

Subjects. Thirty-nine introductory psychology students received course credit for their participation. They were tested in three experimental sessions.

Materials. The problems used were identical to those of Experiment 2.

Procedure. The procedure was identical to that of Experiment 2, except that premises and conclusion were presented simultaneously.

Results

Error data and RT data for correct responses are shown in Table 6.

Errors. Error data were scrutinized with a 3 (Problems) x 3 (Days) within-subject analysis of variance. The Problems effect was highly significant, F (2, 76) = 10.73, $p < .001$. All individual comparisons showed reliable differences. Once again, Problems and Days interacted, F (4, 152) = 3.32, $p < .02$, as shown in Table 7. The pattern of results is similar to that of Experiment 2, showing large truth value differences on Day 1 and large transitivity differences on Days 2 and 3.

Reaction Time. The details of the data treatment were identical to those of Experiment 2. In this experiment, 180 of 1248 observations were missing, and 26 of 39 subjects had at least one missing observation.

The mean RT data for correct responses are shown in Figures 5 and 6. It is clear that the ease of the True Non-Transitive is primarily due to a tendency to simply search for identical statements, and thus selective decrements in RT occur when first premise and conclusion match (see Figure 5). This problem was thus excluded from the analysis, a 2 (First Premise) x 2 (Second Premise) x 2 (Conclusion) x 3 (Problems) within-subject analysis of variance.

All three negation effects were significant. The effects were stronger for first premises (affirmative = 5.18 ± 0.38 seconds, negative = 5.81 ± 0.42 seconds; F (1, 38) = 21.80, $p < .001$) and conclusions (affirmative = 5.17 ± 0.39 seconds, negative = 5.82 ± 0.41 seconds; F (1, 38) = 26.43, $p < .001$) than for second premises (affirmative = 5.34 ± 0.39 seconds, negative = 5.66 ± 0.42 seconds; F (1, 38) = 6.19, $p < .02$). The latter result is only marginally significant given the stringent level of significance adopted.

Table 6

Mean Reaction Times[a] and Percentage of Errors[b]

in Experiment 3

Problem	Premise Combination								
	AAA	ANA	NAA	NNA	AAN	ANN	NAN	NNN	Total
True Transitive	4.15 (14)	5.57 (38)	5.81 (27)	6.11 (34)	5.60 (22)	6.19 (18)	6.08 (21)	6.04 (17)	5.69 (24)
False Transitive	4.55 (47)	5.96 (25)	6.32 (29)	5.62 (28)	5.74 (32)	6.22 (43)	6.51 (39)	5.82 (38)	5.84 (35)
True Non-Transitive	2.54 (6)	3.46 (3)	5.38 (22)	4.93 (27)	4.54 (11)	5.61 (19)	4.75 (4)	4.10 (5)	4.41 (12)
False Non-Transitive	4.99 (46)	5.49 (40)	6.04 (33)	5.87 (35)	5.86 (39)	6.46 (48)	6.83 (40)	6.30 (40)	5.98 (40)

[a] In seconds

[b] In parentheses

Table 7

Error Scores for Each Problem for Each Day

Problem	Day		
	1	2	3
True Transitive	.24	.28	.20
False Transitive	.47	.32	.26
False Non-Transitive	.45	.41	.35

Two interactions related to form variables were found. A strong interaction existed between first and second premises, $\underline{F}$ (1, 38) = 23.23, $\underline{p} < .001$. Second premise negation exerts a larger effect on RT when the first premise is affirmative than when it is negative. Comparison of affirmative and negative second premises reveals a large effect with affirmative first premises (AA = 4.75 ± 0.36 seconds, AN = 5.62 ± 0.41 seconds). With negative first premises, however, the effect goes in the opposite direction, and two negatives are easier than one (NA = 5.93 ± 0.43 seconds, NN = 5.70 ± 0.43 seconds).

A similar but weaker interaction, not significant by stringent standards, existed between first premise and conclusion, $\underline{F}$ (1, 38) = 5.62, $\underline{p} < .025$. The direction of this effect is similar to the other effect, but it appears to be restricted to the True Transitive.

The Problems effect failed to reach significance, $\underline{F}$ (2, 38) = 1.98, $\underline{p} < .15$. As seen in Table 6, the results are in the predicted direction.

Discussion

Additivity of Factors. Strong negation effects for each proposition were found, consistent with Hypothesis 1. Of more interest is that at least one strong interaction was found, between the two premises. It appears that Stages 1 and 2 are not additive. Models

Figure 5. Reaction times for true transitive (top) and true non-transitive problems (bottom) as a function of premise and conclusion form in Experiment 3

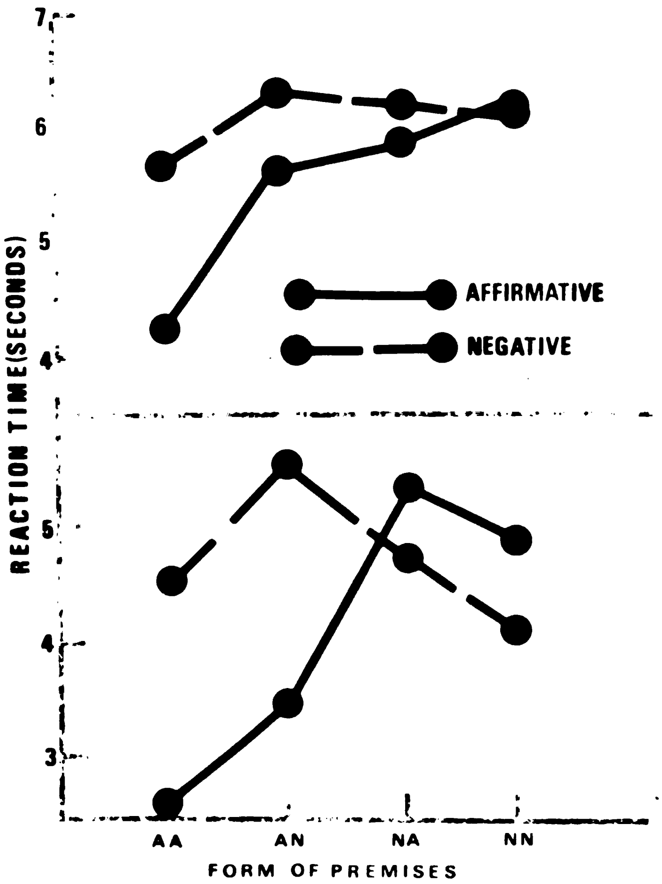
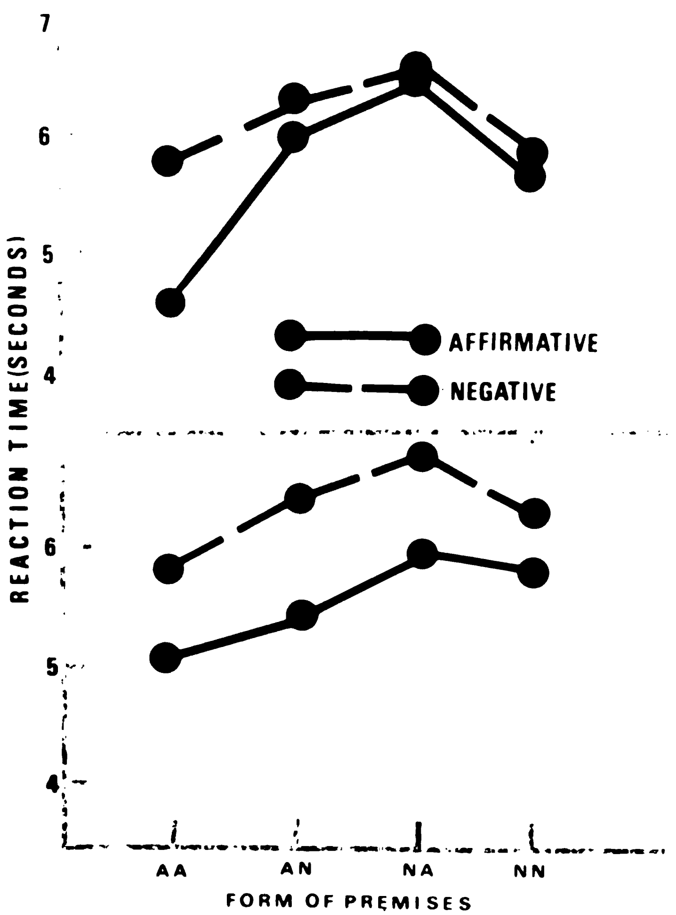

7
6
5
4
5
4
3
REACTION TIME(SECONDS)
AFFIRMATIVE
NEGATIVE
AA
AN
NA
NN
FORM OF PREMISES

FIGURE 5

Figure 6. Reaction times for false transitive (top) and false non-transitive problems (bottom) as a function of premise and conclusion form in Experiment 3



7
6
5
4
6
5
4
REACTION TIME (SECONDS)
AFFIRMATIVE
NEGATIVE
AA
AN
NA
NN
FORM OF PREMISES

FIGURE 6

that assume that operations are sometimes made after only the first premise is registered would not predict this result. For example, Revlis (1975a) has argued that "for syllogisms with abstract propositions, the conversion operation is obligatorily present in the initial encoding of all propositions" (p. 185). Such a model would be hard pressed to explain the obtained interaction.

Stage 3 Representation. The large conclusion negation effect in the present study replicates Experiment 2, and no other factor interacted strongly with conclusion negation (another replication). Thus, even when premises and conclusion are presented simultaneously, Ss appear to derive inferences that are not tied to the syntactic form of the premises. This provides further evidence for Hypothesis 2.

Integration Processes. A major prediction was not confirmed in the comparison of response times for transitive and non-transitive problems. Contrary to prediction, the times were quite similar. One possible reason for this result is that the subjects saw all of the propositions at once. It might be easy to scan the sentences rather than fully processing the two premises (i.e., generating and testing an inference). Evidence in support of Hypothesis 4 might be more likely found in a task in which storage time and verification time are measured separately. Experiment 4 examines this possibility.

Error Results. Error results were consistent with but stronger than those of Experiment 2. Both transitivity and truth value exert influences on error patterns. The significant Problems x Days interaction argues for the interpretation made earlier--that Ss make incorrect interpretations of the materials on Day 1, but these errors decline over time. An equivalence encoding makes the False Transitive as hard as the False Non-Transitive on Day 1. That errors decline faster over time for the False Transitive indicates an abandonment of this strategy in favor of one based on set relations.

Experiment 4

Error results in Experiments 2 and 3 were similar in showing that transitive inferences were easier than non-transitive inferences. In Chapter 1, two hypotheses were advanced regarding the locus of these differences. Whereas an incomplete integration model predicts that logical complexity affects only integration time, a complete integration model assumes that comparison processes are also influenced. The data of Experiments 2 and 3 were equivocal: small transitivity effects were found in both studies. Considerable clarity might be achieved by the use of a storage and verification paradigm (Trabasso, Rollins, & Shaughnessy, 1971) in which subjects are free to take as long as they wish to read the two premises and store an inference. Measurement of storage time independently of verification time thus allows a more direct test of the two models.

Method

Subjects. Twenty-nine introductory psychology students received course credit for their participation. Each subject participated in two experimental sessions.

Materials. Problems employed in previous studies were again used, but no filler problems were included in the study. Two lists of 32 problems were constructed, each consisting of two blocks of 16 items. One block served as practice. Ss were randomly assigned to lists and block orders within lists. In previous studies, Ss saw each problem once per session. In this experiment, each problem was presented twice per session, and only two sessions were needed to acquire comparable amounts of data per S.

Procedure. Ss viewed one slide containing two premises and then verified a single conclusion on a second slide. The time taken to store the two premises as well as the time needed to respond to the conclusion was recorded. This procedure differed from the simple verification task used in Experiment 2 since the initial storage phase was under S's, not E's control and was recorded.

An "advance" button was placed to the left of the control panel directly in front of S. To start a trial S pressed the middle button of the control panel with his right index finger. To bring on the conclusion, he pressed the advance button with his left index finger. This act immediately removed the first slide, which contained the two premises, and left the screen blank for one and one half seconds. Pilot work indicated that this was sufficient time for S to return his left index finger to the telegraph key to prepare to respond. After the blank period, the final slide, containing the conclusion, appeared and a response was made.

The instructions informed S that he should read the first two sentences carefully and then press the advance button when he is ready to see the conclusion.

Results

The results for Experiment 4 are shown in Table 8.

Errors. Once again, an analysis of variance with Problems and Days as factors was performed. The Problems effect was significant, $F\ (2, 56) = 3.44$, $p < .04$, but the ordering of the problems differed from previous findings. The False Transitive was significantly easier than the True Transitive or the

Table 8

Mean Storage and Verification Times[a] and Percentage of Errors[b] in Experiment 4

Problem	Premise Combination AAA	ANA	NAA	NNA	AAN	ANN	NAN	NNN	Total
True	5.06	6.33	5.43	5.72	4.68	6.59	6.18	5.91	5.74
	2.00	1.93	2.05	2.37	2.90	2.97	2.64	2.60	2.43
Transitive	(20)	(32)	(34)	(28)	(24)	(21)	(22)	(24)	(25)
False	5.00	6.18	6.46	5.83	4.58	6.56	6.07	6.90	5.95
	2.02	2.35	2.40	2.13	2.83	2.96	2.91	2.73	2.54
Transitive	(24)	(11)	(18)	(19)	(33)	(19)	(22)	(20)	(21)
True Non-	4.90	6.96	6.22	5.71	5.27	6.58	7.10	5.99	6.00
	1.61	1.78	2.29	2.01	2.60	2.89	2.17	2.01	2.17
Transitive	(0)	(3)	(15)	(20)	(18)	(10)	(3)	(3)	(9)
False Non-	4.91	6.81	6.37	5.74	4.27	6.54	6.58	5.39	5.83
	2.32	2.54	2.54	2.40	2.85	3.23	3.05	2.88	2.73
Transitive	(37)	(23)	(27)	(25)	(24)	(28)	(38)	(23)	(28)

[a] In seconds

[b] In parentheses

False Non-Transitive, as indicated by Newman-Keuls tests (p < .05), but the latter two problems did not differ.

The Days effect was also reliable, F (1, 28) = 28.34, p < .001, and the Problems x Days interaction approached significance, F (2, 56) = 2.42, p < .10. The practice data for each problem are shown in Table 9. While the ordering of the two transitive problems is inconsistent with Experiments 2 and 3, the distinct

Table 9

Error Scores for Each Problem for Each Day

Problem	Day 1	Day 2
True Transitive	.28	.22
False Transitive	.29	.12
False Non-Transitive	.34	.22

trend for practice effects to be most pronounced for the False Transitive is once again apparent. The major difference with previous studies is the lack of a large true-false difference on Day 1.

Storage Time. Treatment of both storage and verification data was identical to that of Experiments 2-3, except that Day 1 data were retained in the statistical analysis. Thirty-seven of 928 observations were missing, and ten of 29 subjects had at least one missing data point.

Mean storage times for trials with correct responses are shown in Figures 7 and 8. A 2 (First Premise) x 2 (Second Premise) x 2 (Conclusion) x 2 (Transitivity) x 2 (Truth Value) within-subject analysis of variance indicated that these times are strongly affected by first premise negation (affirmative = 5.70 ± 0.63 seconds, negative = 6.10 ± 0.66 seconds;

F (1, 28) = 11.19, $p < .002$) and second premise negation (affirmative = 5.57 ± 0.63 seconds, negative = 6.23 ± 0.66 seconds; F (1, 28) = 30.29, $p < .001$).

In agreement with Experiment 3, these two negation factors produced a strong interaction. Storage time is shortest when both premises are affirmative (AA = 4.84 ± 0.55 seconds) or negative (NN = 5.95 ± 0.63 seconds) and greatest when the premises are formally incongruent (AN = 6.56 ± 0.72 seconds, NA = 6.30 ± 0.72 seconds). The result is statistically reliable, F (1, 28) = 26.99, $p < .001$.

One other result was found, a significant first premise by conclusion interaction, F (1, 28) = 7.78, $p < .01$. It is not easy to see how the polarity of the conclusion could affect storage times.

The major prediction about storage times was not confirmed. Transitive problems took no less time to store than non-transitive problems ($F < 1$).

Verification Time. Verification latencies appear in Figures 9 and 10, and they conform to predictions in all but one respect. Robust and non-interactive effects are found for conclusion negation (affirmative = 2.17 ± 0.25 seconds, negative = 2.76 ± 0.28 seconds; F (1, 28) = 48.09, $p < .001$) and truth value (true = 2.29 ± 0.25 seconds, false = 2.63 ± 0.28 seconds; F (1, 28) = 14.98, $p < .001$; interaction $F < 1$).

Figure 7. Storage times for true transitive (top) and true non-transitive problems (bottom) as a function of premise and conclusion form in Experiment 4

true transitive (t
lems (bottom) as a
lusion form in Exp-

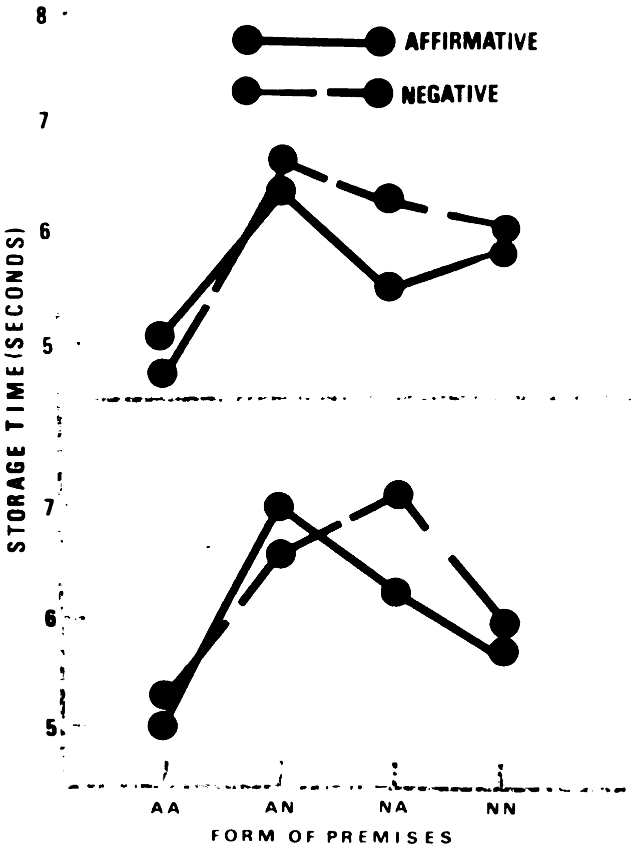
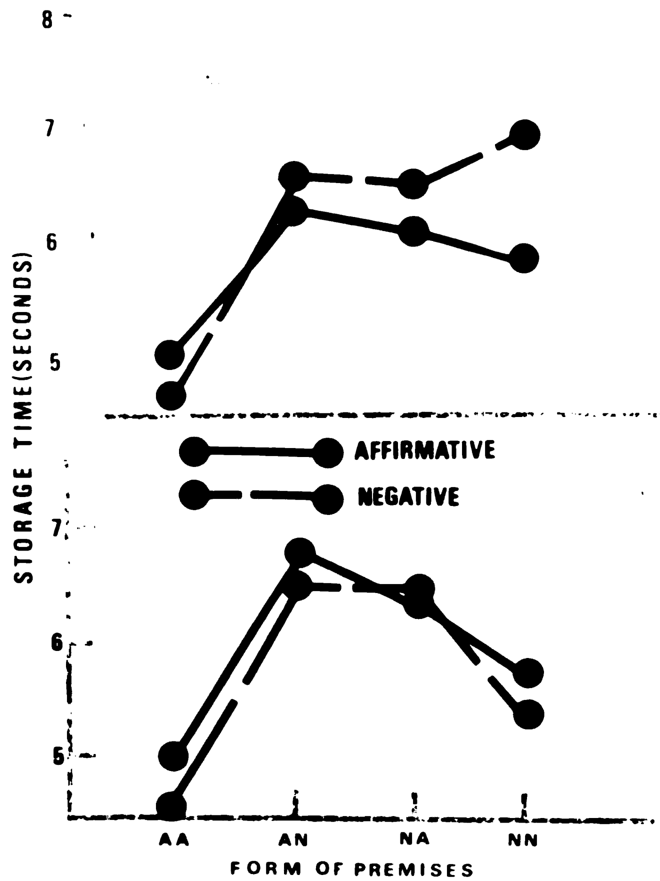


FIGURE 7

Figure 8. Storage times for false transitive (top) and false non-transitive problems (bottom) as a function of premise and conclusion form in Experiment 4



8
7
6
5
STORAGE TIME (SECONDS)
AFFIRMATIVE
NEGATIVE
7
6
5
AA
AN
NA
NN
FORM OF PREMISES

FIGURE 8

In addition, first premise negation and conclusion negation were found to interact, F (1, 28) = 9.46, $p < .005$. Inspection of Figures 9 and 10 indicates that, in contrast to previous results, congruent premise-conclusion combinations tended to have shorter latencies. Figures 9-10, however, show that this congruence pattern is far more striking in the True Non-Transitive than in any of the other problems. To check on the reliability of the interaction for the other three problems, a separate 2 (First Premise) x 2 (Second Premise) x 2 (Conclusion) x 3 (Problems) analysis was undertaken. While a similar trend was observed, the interaction effect was not significant by stringent standards, F (1, 28) = 4.32, $p > .01$. Thus, the results for inference conditions are in general agreement with previous findings.

Finally, no main effect for transitivity was observed ($F < 1$), though there was a trend of a transitive by truth value interaction, F (1, 28) = 6.47, $p < .025$. The major difference lies between the two true problems, as the True Non-Transitive was verified faster than the True Transitive.

Discussion

The patterns observed in this study were complex and not in accord either with predictions or previous

Figure 9. Verification times for true transitive (top) and true non-transitive problems (bottom) as a function of premise and conclusion form in Experiment 4

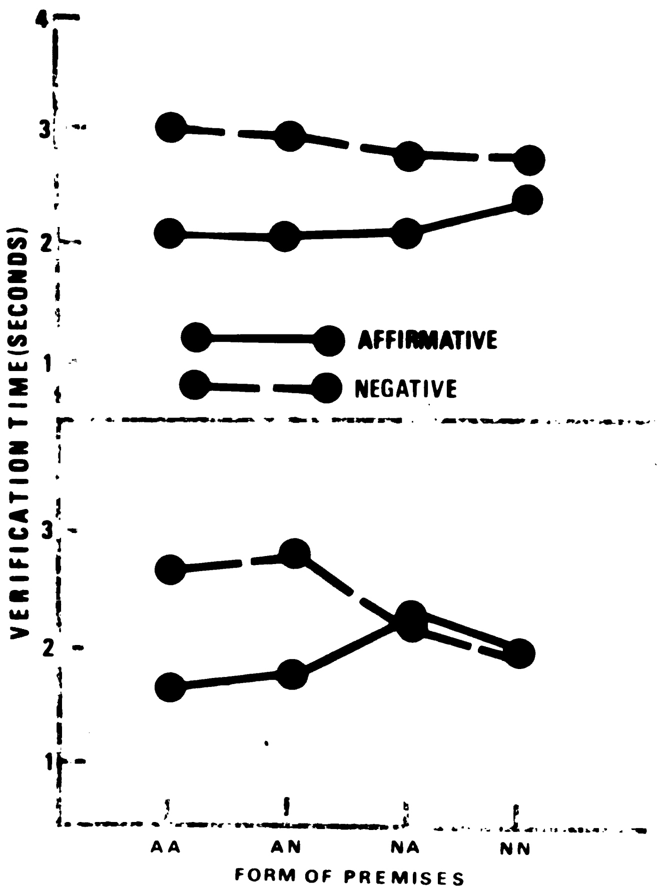
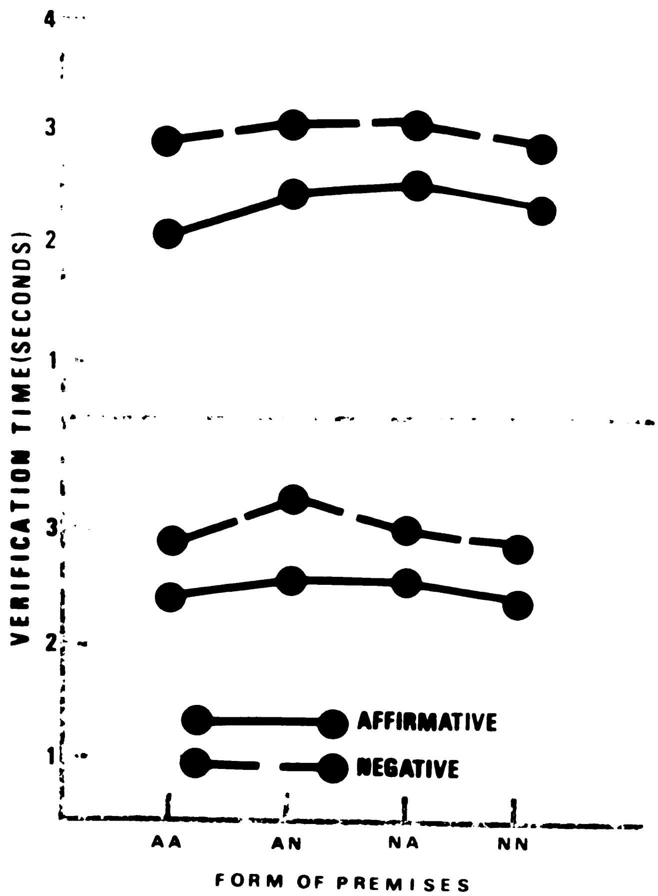

4
3
2
1
3
2
1
VERIFICATION TIME (SECONDS)
AFFIRMATIVE
NEGATIVE
AA
AN
NA
NN
FORM OF PREMISES

FIGURE 9

Figure 10. Verification times for false transitive (top) and false non-transitive problems (bottom) as a function of premise and conclusion form in Experiment 4



VERIFICATION TIME (SECONDS)
4
3
2
1
3
2
1
AFFIRMATIVE
NEGATIVE
AA
AN
NA
NN
FORM OF PREMISES

FIGURE 10

results in several instances.

Inference Representation. The results of Experiment 2 provided strong evidence in support of Hypothesis 2, that inference representations are unaffected by the syntactic complexity of the premises. The interaction between first premise and conclusion negation found in Experiment 4 challenges this claim. Most of the effect, however, is due to the True Non-Transitive, in which one of the premises is presented as the conclusion. It thus appears that while inferences are stored in abstract terms, the premises are available in their original syntactic form.

Transitivity effects. In previous studies, the True Transitive was easier than both the False Transitive and the False Non-Transitive, with the latter being the most difficult. In Experiment 4, the True Transitive was harder than the False Transitive and did not differ from the False Non-Transitive. These departures from previous findings (and predictions) make it difficult to interpret the failure to find either storage or verification effects related to transitivity. One explanation of the error results is that the elimination of filler items made the task too easy and allowed Ss to learn the answers by rote. This view is supported by the observation that error rates were lower than in previous studies even though Ss were less practiced.

One final inconsistent result needs an explanation. Recall that there was a trend for the True Non-Transitive verification latencies to be shorter than those for the True Transitive, thus creating a marginally significant truth value by transitivity interaction not found previously. The result appears to be an artifact of another finding. If the premise information is held in its original syntactic form, then the True Non-Transitive problems can be responded to on the basis of a "verbatim" code, rather than the abstract code used in the inference conditions. If this assertion is correct, then one would expect the True Non-Transitive problem to be fast relative to the other conditions. Why this information is held in verbatim form here, but not in Experiment 2, is unclear.

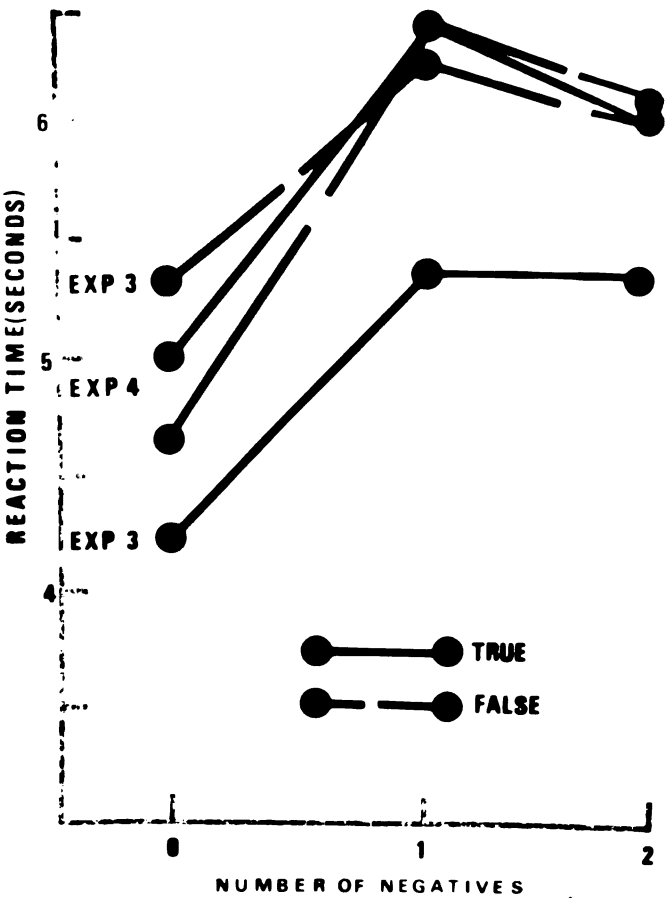
Non-additivity of Premise Negation. The results of both Experiments 3 and 4 are incompatible with the view that Stages 1 and 2 are additive. An additive model predicts that increments in storage or reaction time to two negatives would be the sum of the time increments to each of the single negatives. The results summarized in Figures 5-8 provide ample evidence against such a view. While the exact curves depicted in these figures vary somewhat, the underlying commonality is the lack of additivity.

In particular, the NN condition is less than would be predicted by an additive factors model, a situation that has been termed "underadditivity" (Pachella, 1976) or a negative interaction (Sternberg, 1969).

Interpretation of these negation effects becomes more difficult with the presence of error patterns on negative and affirmative premises. In each of the four experiments discussed, matching biases of the form discussed earlier--more true responses when propositions are congruent--were present. In general, match problems are easy when they are true and hard when they are false. Therefore, correlations between time and errors are generally positive for the true problems and negative for false problems. Could the non-additive premise effects have been contaminated by these error patterns?

An attempt to show the generality of the non-additivity was made by plotting reaction time or storage time as a function of the number of negatives for both true and false problems. The data for Experiments 3 and 4 are shown in Figure 11. Though error profiles are markedly different for true and false problems, the time data are similar in all cases. The non-interactive effect of truth value is

Figure 11. Reaction times (Experiment 3) and storage times (Experiment 4) as a function of the number of negatives in the premises for true and false problems



6
5
4
REACTION TIME(SECONDS)
EXP 3
EXP 4
EXP 3
TRUE
FALSE
0
1
2
NUMBER OF NEGATIVES

FIGURE 11

present in the Experiment 3 data since this study measured total response time.

Sternberg (1969) has noted that experimental artifacts are more likely to obscure true additivity than true interaction, since there are many ways in which interactions can occur but only one way in which an additive relationship may be found. He cautions that interactions are more likely to be "real" if they occur in experiments in which two other factors are shown to be non-interactive. In both Experiments 2 and 4, conclusion encoding and truth value effects were additive, and similar trends can be discerned in Experiment 3.

Interpretations of Underadditivity. Pachella (1976) has discussed several interpretations of underadditive effects, and this discussion is based partly on his presentation. The underadditive result challenges a view that holds that experimental factors affect separate processing stages in simple ways. Two assumptions are present here: (1) The assumption of simple effects, and (2) The assumption of strict seriality of stages. One of the two must be discarded, since this view predicts additivity of negation effects.

The overlap hypothesis retains the simple effects assumption in a parallel processing system. Stages 1 and 2 would be assumed to overlap in time, so that the effects of each negation--operating separately to increase encoding times--would not add. While plausible, this hypothesis would predict that the second premise negation effect should be smaller when the first premise is negative than when it is affirmative (i.e., AA-AN > NA-NN), but that both effects should be positive (i.e., negation harder than affirmation). The latter prediction does not hold: NN is easier than NA. While a parallel processing model could explain an attenuated second negation effect, it cannot account for an effect in the opposite direction.

A second hypothesis is the complex effects hypothesis. Seriality is retained, but factors are assumed to have more than a simple effect on a given stage or process. One complex effects model assumes that both first and second premise negation affect integration, the notion being that more congruent premises are more easily integrated. This position would predict interactions between each premise and transitivity as well as a three-way interaction. In Experiment 4, there is some evidence of the triple interaction, but no evidence of the two-way interactions. While this evidence does not support this model, in fairness it must be noted that transitivity produced no main effect in Experiment 4.

The favored model simply states that first premise negation affects Stage 2 as well as Stage 1. The assumption is that it is easier to encode a proposition that is congruent with a previously-encoded item than to encode an incongruent proposition. This view predicts only the two main effects and the single two-way interaction, and these are the most reliable findings. This model may be falsified by the use of a method that allows the premises to be processed sequentially. Forcing the premises to be encoded in a sequential manner should not affect the context effects found in Experiments 3 and 4, and the interaction between premises should persist. In contrast, an overlap model would predict additivity, since the parallel processing that caused the interaction would be eliminated.

Summary and Conclusions

Four experiments have been presented in this chapter. In the last three, both RTs and errors were recorded, and the independent variables were the form of the propositions (affirmative versus negative) and two content variables (transitivity and truth value). Aside from showing in Experiment 1 that form factors influence error patterns, attention was paid to RT patterns due to form and content, and error patterns attributable to content factors.

The RT predictions regarding negation, or syntactic complexity, were that negatives would (a) increase processing time, but (b) not affect the product of the inferential process. The latter hypothesis was supported by the general absence of interactions between conclusion negation and the two premise negation factors. The former prediction was also upheld, but the negation effects of the two premises were not additive. The interpretation of this finding was that first premise negation affects both Stages 1 and 2. Second premises congruent with the first premises are easier to encode.

Main effects, for both time and errors, were also predicted for the transitivity and truth value factors. It was necessary, first, to demonstrate that transitivity affected error rates in ways that could not be attributable to non-logical processes. This was accomplished by comparing the results for False Transitive and False Non-Transitive problems. The former problem was easier in each study. Unlike truth value differences, this result cannot be explained by response bias hypotheses, since the correct response is constant.

RT effects for transitivity were negligible in all studies. One reason for the negative findings is that Ss apparently tended to misinterpret the universal affirmative propositions as symmetrical statements. This error may have attenuated RT effects; Experiments 5 and 6 in the next chapter investigate this possibility.

Finally, truth value effects for both errors and RTs were found consistently, though the error patterns in Experiment 4 differed somewhat from earlier studies. Truth value RT effects did not interact with negation, suggesting that truth value affects processing subsequent to conclusion encoding (either comparison or response processes).

In sum, syntactically complex problems take longer to process than syntactically simple problems but are ultimately stored in the same abstract format.

Chapter 3

REPRESENTATION AND PROCESSING OF LOGICAL COMPLEXITY

The most striking feature of Experiments 2, 3, and 4 is the absence of any strong effects for the logical complexity of the problems. Since the prediction of storage and (perhaps) verification latency differences as a function of set relationships seems to be a direct implication of existing models, the failures are puzzling. Rather than abandoning the ideas that led to these predictions, the approach taken was to question whether the studies performed were appropriate tests of the original hypotheses. It appears that they were not.

One result that is consistent in all but one of the first four studies is that true problems are easier than false problems, even when the effects of non-logical biases are controlled. This result can be explained within a set-inclusion theory, but it is instructive to note that subjects' self-reports after these experiments were often directly in conflict with such a theory. Many subjects reported that they thought of "All A are B" as "A are equal to B" rather than "A are included in B."

There are a number of reasons to have expected such errors. First, conversions have often been discussed in the context of syllogistic inference (Chapman & Chapman, 1959; Revlis, 1975a, b). One can convert "All A are B" into "All B are A", though it is not logically correct to do so. The assumption is that Ss code propositions as equivalences, in which case such conversions are entirely proper. Secondly, similar misinterpretations occur frequently in other reasoning situations, such as propositional reasoning (e.g., Taplin, 1971; Taplin & Staudenmayer, 1973). Thirdly, evidence on the processing of set-inclusion statements in paragraph materials indicates that subjects often treat such statements as symmetrical unless otherwise instructed (Griggs, 1974). Finally, the equivalence interpretation intuitively seems to make the task easier, and thus is appealing on the grounds of the least effort principle.

If many subjects adopted this interpretation--and an informal tabulation of post-experimental reports indicated this to be so--then the small error and RT results for logical complexity are due to coding rather than processing factors. In particular, these misinterpretations would be likely to attenuate error differences between False Transitive and False Non-Transitive problems, and RT effects due to transitivity. On the other hand, the equivalence coding would be expected to magnify the true-false error differences, since the false items would be true under this interpretation.

Experiment 5

The purpose of Experiment 5 is to experimentally assess the role of interpretational factors in previous results. Subjects were either presented with set-inclusion or equivalence problems and storage time, verification time, and errors were recorded.

The predictions were that transitivity would affect times and errors in the set task but not the equivalence task. In particular, the predictions were that: (1) Transitive problems will take less time to store than non-transitive problems; (2) If the complete integration model is correct, transitive problems will also be easier to verify than non-transitive problems, and this difference will be greatest for the false items; and (3) The False Non-Transitive problem should be more difficult than the False Transitive.

Method

Subjects. Forty-four introductory psychology students participated in one experimental session and were given course credit.

Materials. Standard syllogistic propositions were altered in two ways. One set of problems substituted "are included in" for "are." The other set contained the wording "are equal to." While equal in number of words, the two new versions differ semantically, since the former constrains one to a set-inclusion format, and the latter forces an equivalence interpretation.

The problem set, illustrated with the set problems, is shown in Table 10. The equivalence problems are identical except for the relational phrase. Unlike previous studies, each problem was expressed in only one form. No double negatives were used.

In addition to the true and false items, a third class of items was used in Experiment 5. Contradictory items are a class of items that have the property of being false for both the set and the equivalence tasks. These items are logically incompatible with either a valid inference or a premise. Thus, "No A are C" contradicts the inference "All A are C", and "No C are B" contradicts the premise "All C are B." These items receive the false response in both tasks. Similarly, the true items have the same truth value in each task. The false items, however, are only false for the set group. Nevertheless, the term false will be retained to be consistent with the discussion of Experiments 1-4 and to distinquish these items from the contradictory set.

For the equivalence task, contradiction items serve the purpose of keeping subjects honest, since there is little reason to give a true-false test with all of the answers true. For the set task,

Table 10

Problem Set for Experiment 5

	Transitive	Non-Transitive
True	All A are included in B All B are included in C All A are included in C	All A are included in B All C are included in B All C are included in B
False	All A are included in B All B are included in C All C are included in A	All A are included in B All C are included in B All A are included in C
Contradictory	All A are included in B All B are included in C No A are included in C	All A are included in B All C are included in B No C are included in B

these items provide another demonstration that true-false differences are not merely due to a general response bias toward verification. To insure that subjects do not automatically respond false to any problem beginning with a negative, four filler problems, two of which had true negative conclusions, were added to the problem set. It is important to note that these were universal negatives (No A are B), unlike the double negatives used in Experiments 1-4.

Finally, any statement beginning with A or C was equally likely to fall into each of three response categories. To achieve this, the True Non-Transitive now repeats the second premise rather than the first.

Procedure. Ss were randomly assigned to either the Equivalence group or the Set group. The storage-and-verification paradigm of Experiment 4 was employed.

Seven blocks of the ten problems were constructed, with the blocks differing only in the order of the problems. Order of problems within blocks was random, with the stipulation that no runs of more than two occurred for any problem dimension. Block order was randomized for each subject, with the first block treated as practice. Unlike earlier studies, problems were not re-presented during the practice session.

Feedback was given in accordance with the task: Equivalence subjects were correct in responding true to the false items, but this would be an error for the Set group.

Results

The results of Experiment 5 are shown in Table 11. Several points are noteworthy about the data. First, errors are generally lower than in previous studies, even with just one session of practice. Apparently, subjects had little difficulty in sticking to the task provided for them. Further, the Equivalence task appears the easier of the two tasks by a wide margin. This difference appears in storage times (3.20 versus 5.53 seconds), verification times (1.20 versus 2.27 seconds) and errors (2.5% versus 15.5%).

Errors. Error results were analyzed by a simple one way analysis of variance that probed differences between problems. For the Set group, these differences were reliable, F (5, 105) = 9.32, p < .001. Newman-Keuls tests indicated that error rates were greater (p < .05) for false problems than for true or contradictory problems, which did not differ. Further, the False Transitive (.23) was significantly easier than the False Non-Transitive (.39).

No error differences were found for the Equivalence group scores. Error rates for different problems ranged from 0% to 5%. While there was some

Table 11

Mean Reaction Times[a] and Percentage of Errors in Experiment 5

	Set		
Problems	Errors	Storage	Verification
Transitive			
True	11	5.42	1.93
False	23	5.01	2.51
Contradictory	5	5.39	1.95
Total	13	5.27	2.13
Non-Transitive			
True	8	5.71	2.28
False	39	5.63	2.75
Contradictory	7	5.99	2.21
Total	18	5.78	2.41

	Equivalence		
Problems	Errors	Storage	Verification
Transitive			
True	0	3.14	0.96
False	2	3.20	1.11
Contradictory	4	3.21	1.25
Total	2	3.18	1.11
Non-Transitive			
True	3	3.32	1.31
False	1	3.24	1.07
Contradictory	5	3.10	1.48
Total	3	3.22	1.28

[a] In seconds

trend for the contradictory problems to be more difficult, this was not established statistically, F (5, 105) = 1.92, $p < .10$.

Storage Time. Mean storage times for correct responses for each group are shown in Figure 12. Separate analyses were performed for the two groups. This was done because the main prediction was that transitivity would exert a significant effect on storage times only for the Set group. Use of an interaction term in a combined analysis would be imprecise in that a significant storage effect for the Equivalence group that is smaller in magnitude than the Set effect could also account for such a finding. Separate analyses thus provide a more rigorous test of the hypotheses of interest. All individual times greater than 20 seconds were counted as 20 seconds. This rule was used six times. Since only two of 132 observations were missing from individual subject means, normal (.05) levels of significance were adopted, and 2 (Truth Value) x 2 (Transitivity) within-subject analyses of variance were performed.

As can be seen in Figure 12, transitive problems took less time to store than non-transitive problems in the set task (transitive = 5.26 ± 0.52 seconds, non-transitive = 5.78 ± 0.68 seconds; F (1, 121) = 5.29, $p < .031$). The interaction between truth value and transitivity was not significant ($F < 1$), but

Figure 12. Storage times as a function of transitivity and truth value in set and equivalence tasks in Experiment 5

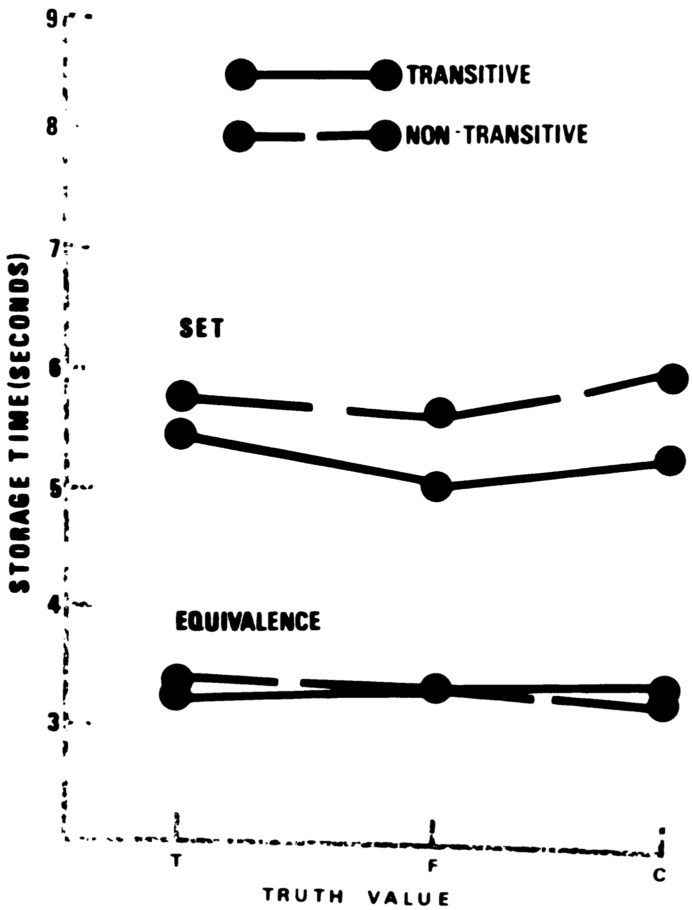

TRANSITIVE
NON-TRANSITIVE
SET
EQUIVALENCE
STORAGE TIME (SECONDS)
9
8
7
6
5
4
3
T
F
C
TRUTH VALUE

FIGURE 12

truth value had a marginally significant influence on storage times, F (2, 42) = 2.94, $p < .06$. This appears to be an artifact. False problems are harder than true or contradictory problems, especially early in the session. Since only correct RTs are considered and since RT improves with practice (see below), the false data may have reflected a different stage of practice.

Storage times for Equivalence subjects, as expected, did not vary as a function of transitivity (transitive = 3.18 ± 0.29 seconds, non-transitive = 3.22 ± 0.29 seconds; F (1, 121) = 0.32). Truth value and interaction effects also failed to approach significance.

Verification Time. The verification data are shown in Figure 13, and the results for the Set group are quite surprising. Transitive problems were verified more rapidly than non-transitive problems (transitive = 2.13 ± 0.16 seconds, non-transitive = 2.42 ± 0.19 seconds; F (2, 42) = 5.08, $p < .035$), but there was no interaction between truth value and transitivity.

Truth value also affects verification scores for Set subjects (true = 2.11 ± 0.18 seconds, false = 2.63 ± 0.27 seconds, contradictory = 2.08 ± 0.10 seconds; F (2, 42) = 6.17, $p < .004$). Newman-Keuls analyses showed that the times for false problems were

Figure 13. Verification times as a function of transitivity and truth value in set and equivalence tasks in Experiment 5

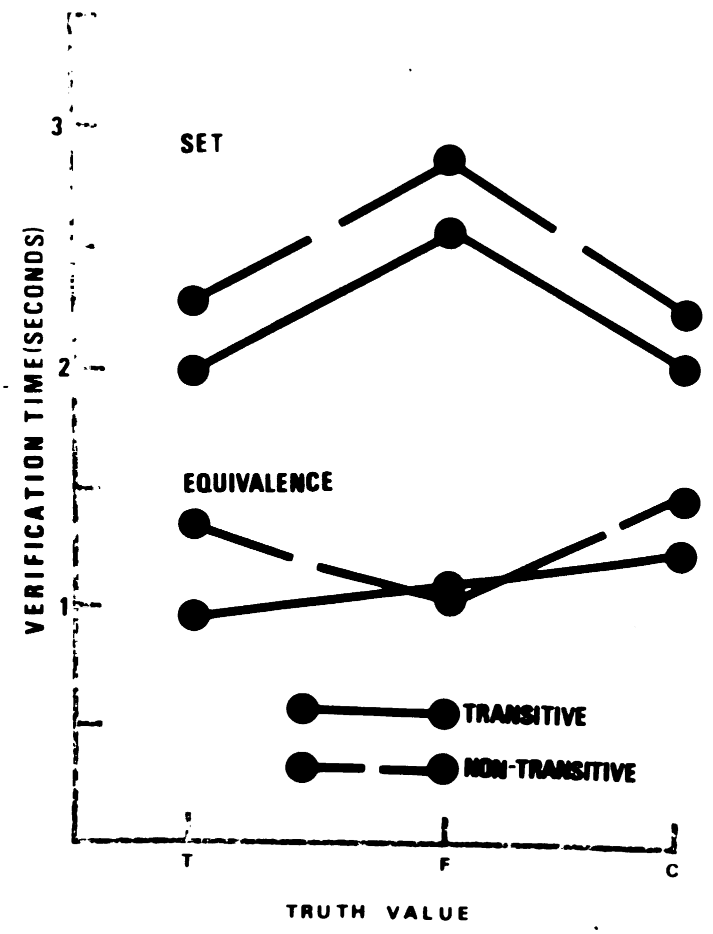

SET
EQUIVALENCE
TRANSITIVE
NON-TRANSITIVE
3
2
1
VERIFICATION TIME (SECONDS)
T
F
C
TRUTH VALUE

FIGURE 13

greater than those for either the true or contradictory items, which did not differ ($p < .05$).

Both main effects and the interaction were significant in the Equivalence analysis, but the effects were somewhat different from those for the Set group. Transitive problems were easier to verify than non-transitive problems (transitive = 1.11 ± 0.05 seconds, non-transitive = 1.28 ± 0.08 seconds; F (1, 21) = 16.27, $p < .001$). Truth value also affected verification latencies (true = 1.14 ± 0.06 seconds, false = 1.09 ± 0.08 seconds, contradictory = 1.34 ± 0.09 seconds; F (2, 42) = 11.30, $p < .001$), but in a different way than with the Set group. False and true problems were easier to verify than contradictory problems (Newman-Keuls, $p < .05$). The significant interaction between transitivity and truth value is best described by stating that while transitive problems are easier to verify than non-transitive problems when the conclusion is true or contradictory, the opposite results occur when the conclusion is false.

Discussion

Storage Effect. The results of this study are clear and in marked contrast to previous findings. Non-transitive problems are harder and take longer to integrate than transitive problems only in the Set task. We wish to conclude from this that logical complexity affects integration time, but at least two alternate explanations of the transitivity findings exist and must be discussed.

One is simply that it is the transitivity per se and not the complexity of the set relations that is the source of the effect. It could be argued that transitive orders are simply easier to process. There is a simple way to discredit this hypothesis. Note that Ss in the Equivalence group did not demonstrate the storage effect; here the difference between transitive and non-transitive problems is a mere 36 milliseconds. This is to be expected since no additional complexity is introduced by non-transitive orders in the equivalence task. Thus, the data argue quite persuasively that it is not order per se, but what the order signifies, that is the crucial factor in the observed storage effect.

A second alternate explanation is that there is no direct relationship between logical structure and storage time. Rather, structural complexity leads to greater errors, thus causing subjects to become overly cautious in dealing with problems they recall having missed. The effect, under this interpretation, is quite general; the extra time is devoted to a generalized cautiousness rather than to analysis of set relations. Two pieces of data argue against this view.

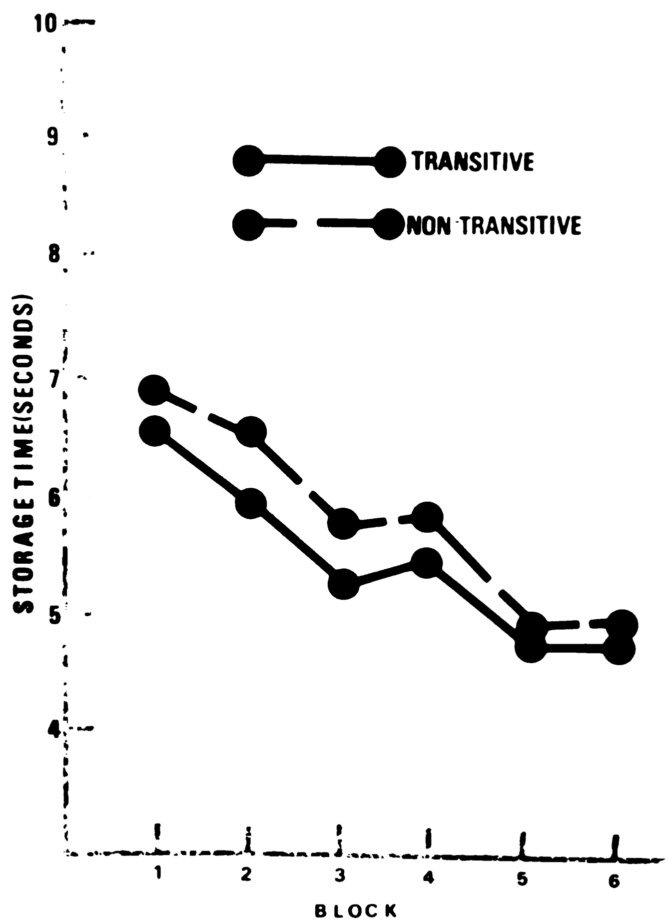
First, it is possible to plot the transitive effect as a function of block. The cautiousness hypothesis would predict that the effect would be negligible at first, but gain in strength as the session progresses and errors are made on the False Non-Transitive. The relevant data, shown in Figure 14, clearly contradict this position. The effect is consistent across all levels of practice.

A second implication of the cautiousness position is that a significant correlation should exist between the number of errors one commits on the False Non-Transitive and the size of the transitive storage effect. Those subjects who had little or no trouble with the False Non-Transitive--8 of 22 subjects missed it once or less--should show little or no storage effect. The product-moment correlation between the two variables, however, is quite low ($r = .01$) and thus offers no support for this hypothesis.

Thus, the results argue persuasively for the view that complexity of logical structure exerts a significant effect on the time needed to integrate information from the two premises.

Transitive Differences in Verification. The role of logical complexity in verification was to be a telling factor in the evaluation of the complete

Figure 14. Storage times for transitive and non-transitive problems in each block of Experiment 5



10
9
8
7
6
5
4
STORAGE TIME(SECONDS)
TRANSITIVE
NON-TRANSITIVE
1
2
3
4
5
6
BLOCK

FIGURE 14

and incomplete integration models, yet the data of Experiment 5 support neither position. The former predicts a main effect for transitivity along with an interaction with truth value, while the latter predicts no effects. The results indicated a main effect but no interaction.

The rationale behind the prediction of a main effect was that more information was stored for non-transitive orders, and thus more needed to be considered during the verification process. The interaction was expected, with the complete integration model, because the verification process should be especially difficult if the multiple information is inconsistent. Finding just one of these two effects is puzzling. It would be strange to assume that the main effect was due to, say, an increased number of set relations to scan and then discover that whether these relations are consistent or not makes not a bit of difference. This leads to some suspicion as to the validity of the transitive verification finding.

A simple explanation for the verification results can be considered. It is possible that subjects did not complete the integration of the premises prior to advancing to the conclusion. Subjects in Experiment 6 were instructed to fully process the the two premises before pressing the advance button.

If the transitive verification effect is simply due to a less than perfect correspondence between hypothesized stages and experimental operations, then the verification effect should disappear.

Truth Value and Verification Effects. The truth value findings for the Set group are intuitively reasonable: false problems, but not contradictory problems, take longer to verify than true problems. Thus it is not simply that false responses are less preferred, since when the false problem is easy, it is equivalent to the true item in errors as well as verification latencies.

For the Equivalence group, the verification differences can be explained with the help of a little imagination. It is the contradiction items that are the slowest. Since it is known that response probability affects RT (Smith, 1968), one possible explanation for this effect is that the false response is less frequent for the Equivalence group. Including fillers, "false" is correct 40% of the time. Experiment 6 controls response probability to examine this hypothesis. Note that this explanation cannot work for the Set data, in which the false response is the correct one 60% of the time, including the fillers.

A strange finding is that transitivity and truth value interact for the Equivalence group. This may be due to surface factors, in that conclusions that begin with A are easier than those that begin with C. It just so happens that in two of three instances, propositions beginning with C are non-transitive. This could account for the transitive verification difference as well as the interaction.

Experiment 6

Experiment 6 was designed to replicate Experiment 5. In addition, two specific purposes were (1) to determine the reliability of the transitive verification effects by strengthening the instructions and (2) to examine the role of response probability in the truth value effects found in Experiment 5. The nature of the inference rules being compared forces one to confound either stimulus or response probability with the factors of interest. In Experiment 5, all problems were presented equally often, but the "true" response was correct 40% for the Set group, and 60% for the Equivalence group. Experiment 6 thus presents the items an unequal number of times so that for each group the probability of true and false responses is equal.

Method

Subjects. Forty introductory psychology students received course credit for their participation in the experiment. Ss attended a single experimental session. Data from two subjects were discarded for failure to follow instructions.

Materials. Problems were identical to those of Experiment 5. Two lists were constructed, one for each group. For the Set group, true problems were represented a disproportionate number of times so that over all problems, including fillers, each response was correct 50% of the time. True problems were presented eight times each and the other problems appeared five times each. True fillers appeared seven times and contradictory fillers appeared five times. For the Equivalence group, contradictory problems rather than true problems appeared more often than the other problems.

For each list, six blocks of twelve problems were constructed, with each block subject to the same stipulations as in Experiment 5. All problems appeared at least once in each block. Problems that were more frequent appeared twice in three blocks and once in two blocks.

Procedure. Subjects were randomly assigned to experimental groups. Order of the blocks was randomized for each subject. One block served as warmup.

The instructions were identical to those in previous studies using the storage-and-verification procedure except that subjects were explicitly warned to fully process the two premises before pressing the advance button: "Prepare for any number of conclusions that might be based on the first two sentences. Don't press the advance button until you have reached this stage of preparation." Instructions were vague as to how to process the premises, but were specific in admonishing Ss to do so fully before attempting to verify the conclusions.

Results

The major results of Experiment 6 are tabulated in Table 12.

Errors. The overall error rate for the Set group was 13.8%. A one-way analysis of variance indicated that differences in error rates for the six problems were reliable, F (5, 95) = 10.88, $p < .001$. Newman-Keuls analyses showed that the False Non-Transitive (.37) was more difficult than any of the other problems ($p < .05$). The only other significant contrast was that the False Transitive (.20) was more difficult than the easiest problem, the Contradictory Non-Transitive (.03).

Error rates for the Equivalence group problems ranged from 0% to 8%, with an average of 3.4%. The one way analysis was marginally significant, F (5, 95) = 2.35, $p < .047$. The only significant contrast was between the True Transitive (.00) and the Contradictory Non-Transitive (.08).

Table 12

Mean Reaction Times[a] and Percentage of Errors in Experiment 6

Problems	Set: Errors	Set: Storage	Set: Verification
Transitive			
True	9	6.86	2.01
False	20	6.83	2.91
Contradictory	7	6.77	2.30
Total	12	6.82	2.41
Non-Transitive			
True	8	8.39	2.15
False	37	7.82	2.87
Contradictory	3	8.04	2.29
Total	16	8.08	2.44

Problems	Equivalence: Errors	Equivalence: Storage	Equivalence: Verification
Transitive			
True	0	3.88	1.07
False	2	3.89	1.18
Contradictory	8	3.56	1.26
Total	3	3.78	1.17
Non-Transitive			
True	2	3.83	1.25
False	3	3.83	1.05
Contradictory	6	3.79	1.26
Total	4	3.81	1.18

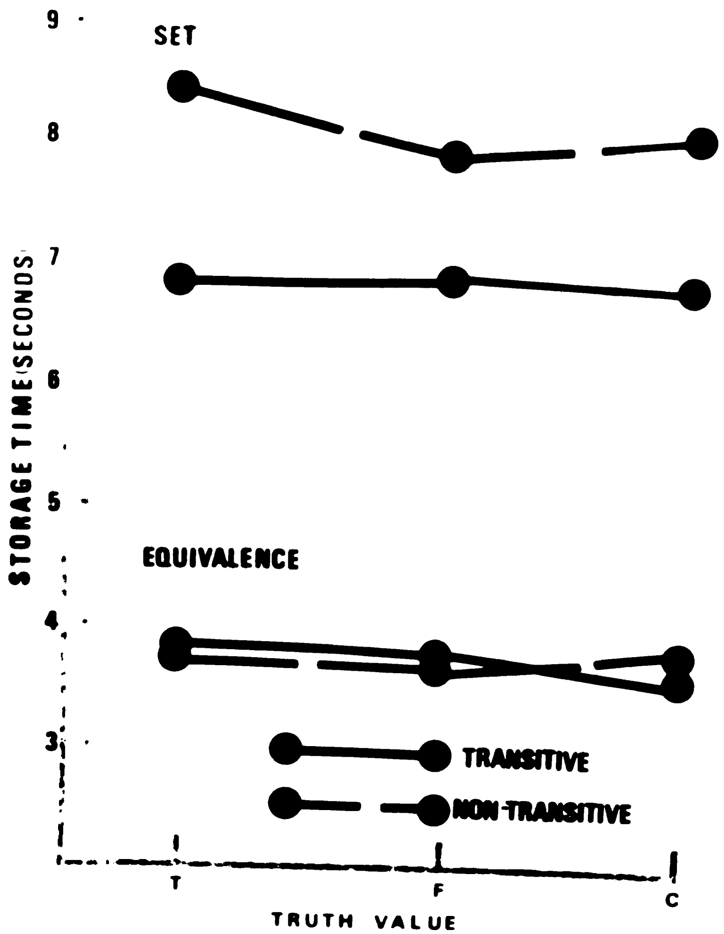
[a] In seconds

Storage Time. Since only two of 120 observations were missing, standard (.05) levels of significance were adopted as in Experiment 5. Latency data for correct responses are shown in Figure 15. The storage results are fully consistent with Experiment 5. Transitive problems were easier to store than non-transitive problems for the Set group (transitive = 6.82 ± 0.82 seconds, non-transitive = 8.08 ± 0.97 seconds; F (1, 19) = 5.94, p < .025) but not for the Equivalence group (transitive = 3.78 ± 0.31 seconds, non-transitive = 3.81 ± 0.35 seconds; F < 1). No other storage effects attained significance for either group.

Verification Time. Latency data for correct responses are shown in Figure 16. In contrast to Experiment 5, no transitivity effects were present for either the Set group (transitive = 2.41 ± 0.31 seconds, non-transitive = 2.44 ± 0.25 seconds; F < 1) or the Equivalence group (transitive = 1.17 ± 0.06 seconds, non-transitive = 1.18 ± 0.08 seconds; F < 1).

Other effects found in Experiment 5 were confirmed. For the Set group, a non-interactive truth value effect was found (true = 2.08 ± 0.20 seconds, false = 2.89 ± 0.41 seconds, contradictory = 2.30 ± 0.25 seconds; F (2, 38) = 7.42, p < .002). Newman-Keuls tests indicated that the false problems were

Figure 15. Storage times as a function of transitivity and truth value in set and equivalence tasks in Experiment 6



SET
EQUIVALENCE
STORAGE TIME (SECONDS)
9
8
7
6
5
4
3
TRANSITIVE
NON-TRANSITIVE
T
F
C
TRUTH VALUE

FIGURE 15

Figure 16. Verification times as a function of transitivity and truth value in set and equivalence tasks in Experiment 6

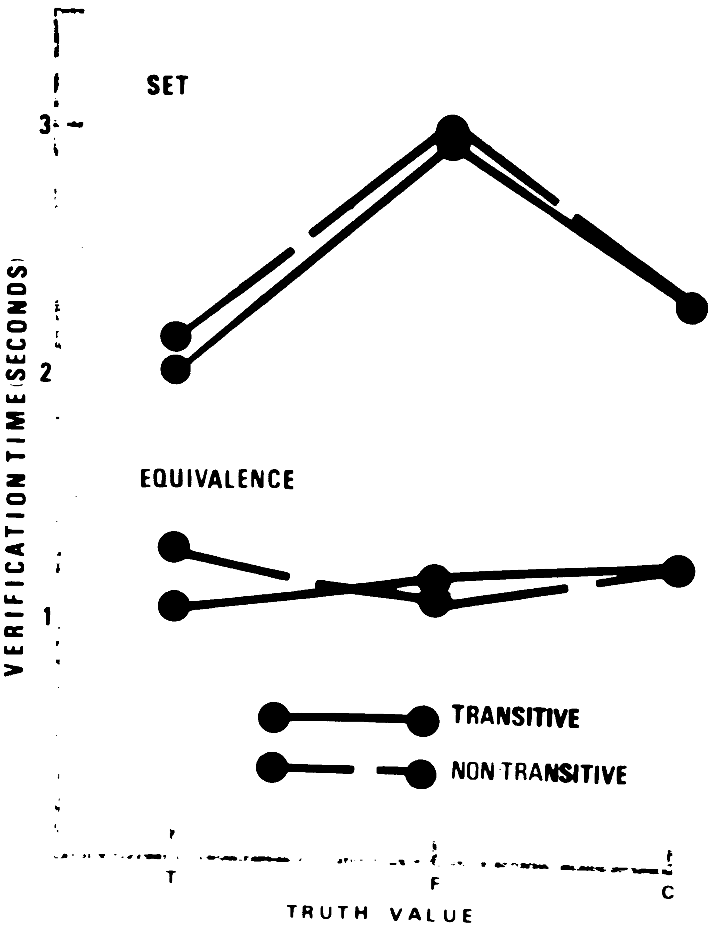

SET
EQUIVALENCE
VERIFICATION TIME (SECONDS)
3
2
1
TRANSITIVE
NON TRANSITIVE
T
F
C
TRUTH VALUE

FIGURE 16

slower than the contradictory or true problems which again did not differ.

For the Equivalence group, truth value effects favored true and false items (true = 1.16 ± 0.07 seconds, false = 1.11 ± 0.06 seconds, contradictory = 1.26 ± 0.09 seconds; F (2, 38) = 6.13, $p < .005$). The interaction between truth value and transitivity was also significant, F (2, 38) = 10.40, $p < .001$. Transitive verification latencies were shorter than non-transitive latencies for true problems, but not for false or contradictory items.

Discussion

Replications of Experiment 5. Since this experiment was in all respects but two a straightforward replication of Experiment 5, the points of correspondence between the two studies bear specific mentioning. For the Set task, the consistencies include: (a) storage times for transitive problems were shorter than those for non-transitive problems, (b) false problems took especially long to verify, (c) false problems were harder than the other problems, and the False Non-Transitive was the hardest, and (d) no interactions in either storage or verification times were found.

Replications in the Equivalence task include: (a) little or no error effects, (b) lack of any storage differences, (c) verification effects for the truth value factor, and (d) interaction effects for verification.

<u>Failures to Replicate</u>. Two findings were inconsistent between Experiments 5 and 6. In Experiment 5, transitive problems were easier to verify in both tasks than non-transitive problems. The latter study found neither effect. The failure to replicate the set-inclusion effect appears due to the instructions (see below), while the reason for either the presence or the absence of this effect in the Equivalence group is unclear.

<u>Interpretation of Storage Differences</u>. In the Set group, transitive problems were stored reliably faster than non-transitive problems, and several interpretations of this effect were considered in the discussion of Experiment 5. One alternative is to consider the finding merely due to the transitive order per se, but the results of each of the last two studies show that when the non-transitive order is unrelated to logical structure (in the Equivalence task), it does not produce longer storage times. Thus, the transitive effect is not simply an order effect.

The second alternative hypothesis was the "cautiousness" view. As in Experiment 5, there was no support for this position. Figure 17 shows the storage times for transitive and non-transitive problems as a function of practice, and it is clear that the storage

Figure 17. Storage times for transitive and non-transitive problems in each block of Experiment 6

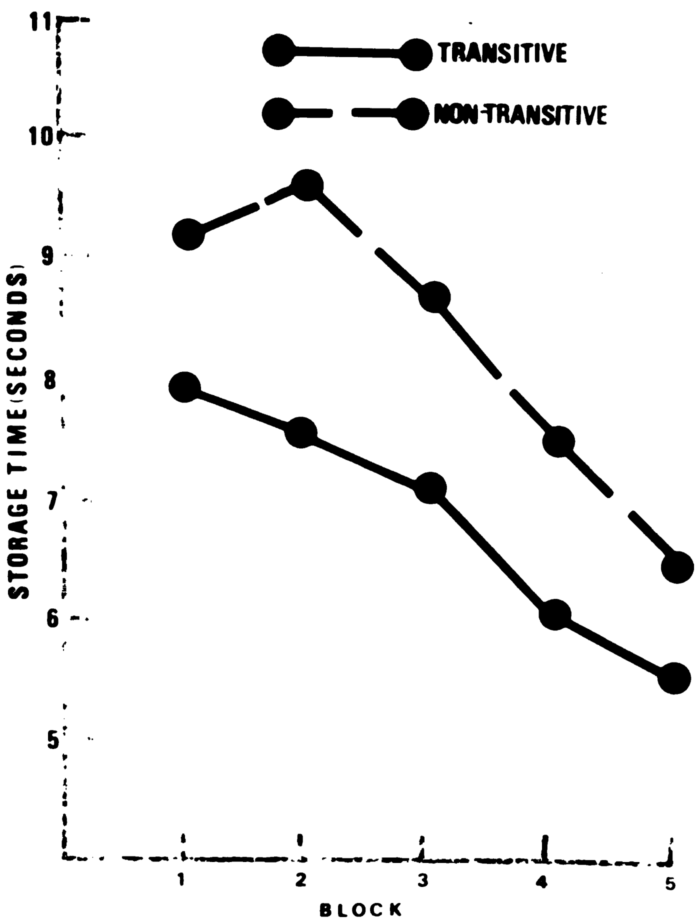

TRANSITIVE
NON-TRANSITIVE
STORAGE TIME (SECONDS)
11
10
9
8
7
6
5
1
2
3
4
5
BLOCK

FIGURE 17

effect is consistent across all levels of practice. In addition, the correlation between errors on the False Non-Transitive and the size of the storage effect was minimal and the direction is the opposite to that predicted ($r = -0.17$).

The conclusion to be drawn is that there is a direct relationship between logical complexity and storage time. The number and/or type of mental operations differs for logically simple and complex problems during the integration stage.

Effects of Instructions. By altering the instructions so that Ss are likely to have fully processed the two premises before proceeding to the conclusion, the verification difference between transitive and non-transitive problems found in Experiment 5 is eliminated. Further, in neither study was an interaction between truth value and transitivity evident.

This pattern of results would be difficult to explain if a number of set relations are stored during Stage 3 processing. However, a model that assumes that only a single relation is stored is not embarrassed by these negative findings. The conclusion is that Experiments 5 and 6, taken together, strongly support the incomplete integration model outlined earlier.

Effects of Response Probability. Relatively small effects were found that could be attributed to response probability. There should be a selective reduction in RTs for true set problems and contradictory equivalence problems. The results were in this direction, but were quite weak and did not alter the statistical results.

Summary and Conclusions

The general approach taken in this chapter seems to have been fruitful. By insuring the proper interpretation of the syllogistic propositions, considerable clarity has been achieved. The major finding was that logical complexity affects errors and storage times, but not verification times. A model that assumes that inference representations are unitary rather than complex is capable of providing an adequate account of these results.

The effects of logical complexity mirror the results for syntactic complexity presented earlier. The general principle appears to be that more complex problems increase processing time (either encoding time or integration time) but not the complexity of the inferential product (either the syntactic or set complexity of the inference).

Chapter 4

SUMMARY

This chapter presents the conclusions based on the studies described in previous chapters. A sketch of the relationship between the present work and earlier work is first outlined, followed by a summary of the hypotheses and findings, and their implications. A final section examines the utility of a chronometric approach to the experimental analysis of logical inference.

Relation to Previous Research

The purpose of using RT to analyze performance in a logical task is to provide information that is difficult to gain from inspecting error patterns.

Ceraso and Provitera (1971) were the first to demonstrate that the complexity of the set relationships can be a critical factor in syllogistic inference. Erickson (1974) incorporated this insight into his set-theoretic analysis and generated two inference models. The central point of difference between the complete and incomplete integration models was in the characterization of the way an inference is internally represented. For problems whose premises generate multiple set relations, the inference may be complex in that all such relations are stored; alternatively, the structure may be unitary in that only a single relation is stored.

As noted earlier, the error data marshalled by Erickson did not adequately discriminate between these two positions.

The present research both replicates and extends prior work. The general finding that more errors occur on problems with more complex logical structures is consistent with both Ceraso and Provitera (1971) and Erickson (1974). Both of the models offered by Erickson would predict this result. This consistency of error findings is important for the interpretation of RT results, since it gives reason to believe that the same process is being studied in the present as in past experiments.

The present work extends previous research in the sense that a different method is used to test some hypotheses generated by studies of error rates. Of central interest was the locus of logical complexity effects within the series of hypothetical processing stages. Both the complete and incomplete integration models predict that logical complexity will affect error patterns and the time needed to integrate the premises. However, only the complete integration model predicts that comparison time is also affected by set complexity. The crucial piece of evidence is whether verification time (of which one part is comparison time) is influenced by the complexity of the set relations.

Summary of Hypotheses and Findings

Chapter 1 provided a schematic outline of the logical inference process, and the purpose of the experiments that followed was to fill in some of the details of the process. This summary of the major findings is presented in the context of the original hypotheses.

Hypothesis 1

Negative premises take longer to encode than affirmative premises. Since there are many reports in the literature on comprehension that indicate that negatives take longer to comprehend than affirmatives (for review, see Clark, 1974), it is not surprising to find that this hypothesis is strongly supported in Experiments 2-4. In each study, large negation effects were found.

Another reliable finding was that first and second premise negation interacted in Experiments 3 and 4. The time taken to encode two negatives is less than what would be expected if the two premises were encoded separately. In some cases the time needed to encode two negatives is less than that needed to encode a single negative.

Hypothesis 2

The time needed to verify a conclusion is unrelated to the syntactic form of the original premises. This hypothesis is supported by the findings of Experiment 2. In that study, neither of the premise negation factors affected verification latencies. Results from Experiments 3 and 4 were consistent with this finding, but less clear-cut.

The time required to verify a premise (i.e., simple verification) has a more uncertain status. The data here came from the True Non-Transitive condition, in which either the first or second premise is simply repeated as the conclusion. In Experiment 2, in which premises were presented for ten seconds, premise verification was unrelated to the syntactic form of its original presentation. Congruence of form, however, produced large effects in Experiments 3 and 4. Temporal factors may be critical here. In Experiment 3, all information was shown simultaneously, and in Experiment 4 Ss controlled storage time and generally took less than ten seconds. Recoding of the premises evidently does not occur under these circumstances.

Hypothesis 3

Response time for true problems is less than for false problems. All four RT studies confirmed this relationship. Truth value also did not interact reliably with any of the other four variables (three negation factors, and transitivity). Truth value could either affect comparison processes, by assuming that one is "set" to assume a conclusion is true unless proven false, or response processes. The latter alternative may be more likely here, since errors are more frequent for false items, and this would seem to increase response uncertainty.

Hypothesis 4

Non-transitive problems take longer to integrate than transitive problems. This hypothesis was disconfirmed in Experiments 3-4 and supported in Experiments 5-6. The critical difference is probably that *Ss* in the latter two studies were constrained to encode the propositions as set relations rather than equivalences.

Hypothesis 5

(a) Non-transitive problems require more time during the comparison process than transitive problems. This prediction was given weak support in Experiment 2; the effect, however, is reliable in Experiment 5. The latter result was not confirmed under more stringent testing conditions (Experiment 6). It was concluded that the effect in Experiment 5 was simply due to an imperfect correspondence between experimental operations and processing stages. This hypothesis was rejected.

(b) The increment for non-transitive problems during the comparison process, relative to transitive problems, should be greater for false than true items. Truth value and transitivity should interact. This interaction was not found in any of the studies. The only experiment that observed a non-significant trend toward this interaction (Experiment 4) found a pattern opposite to that predicted. This hypothesis was rejected.

Hypothesis 6

Errors should be most frequent for the False Non-Transitive, least frequent for the True Transitive, and intermediate for the False Transitive. This compound hypothesis was supported by all studies except Experiment 4. The size of the transitive effect (i.e., the difference between false problems) varies somewhat; the effect is largest with practiced subjects and when the materials preclude erroneous encodings.

Acceptance of Hypothesis 6 does not imply that error patterns are fully predictable from a consideration of logical factors. Various non-logical factors (response bias, matching bias, premise misinterpretation) also affect performance. Nevertheless, the key prediction--that the False Transitive is easier than the False Non-Transitive--cannot be handled by non-logical factors. True-false differences are more ambiguous, but they persist even when these factors are controlled.

Theoretical Implications

What is the nature of the inference drawn in a syllogistic inference task? The results suggest that it is both abstract and unitary. That is, it is semantically- rather than syntactically-based and it is a single cognitive unit.

The evidence that the inference is abstract is that the syntactic form of the premises does not influence the time required to verify a conclusion. This finding indicates that problems that are expressed differently but contain the same information are stored in similar form. That different versions of the same problem are similar only at an abstract or semantic level implies that the representation is semantically-based.

The semantic unit most likely to be the basis for the representation is the set relation, since the complexity of these relations was found to affect both storage times and errors. Furthermore, the inference is best characterized as a single set relation, since the complexity of the logical structure does not influence verification times.

Some of the implications of this research extend beyond logical inference and fit quite well with what has become known as the "constructivist" approach to comprehension and memory.

Bransford and Franks (1971; Bransford, Barclay, & Franks, 1972) and their colleagues have been most responsible for the development of this view. They argue that one typically processes a series of events by constructing a wholistic semantic representation of the individual events. Initial support for this conclusion (Bransford & Franks, 1971) came from a recognition memory study in which Ss first were given a number of simple propositions such as "the rock rolled down the mountain," "the rock crushed the hut," and "the hut was tiny," and later were given a number of sentences in a recognition test. The major finding was that Ss often "recognized" complex sentences such as "the rock which rolled down the mountain crushed the tiny hut" that were not actually presented. The explanation was that the individual propositions were integrated into a single semantic unit used later in the recognition task.

Similar conclusions have been drawn from studies of linear orderings by Potts (1972, 1975) and Trabasso (1975; Trabasso & Riley, 1975; Trabasso, Riley, & Wilson, 1975). In Potts' studies, subjects were given paragraphs containing linear orderings such as "the bear was smarter than the hawk", "the hawk was smarter than the wolf," "the wolf was smarter than the deer," and so on. From these facts, one can derive that the bear was smarter than the deer, but this proposition was not actually presented. In a later verification task, Potts found that derived propositions were verified more rapidly than actually-presented sentences. Both Potts and Trabasso argue that inferences are generated during the acquisition phase and stored along with presented material in a single linear order. RT varies inversely with the distance between objects in the order. Thus, it is easier to verify an inference than a proposition, actually presented, about objects adjacent in the linear order.

If the constructivist thesis is a truly general cognitive theory, it must apply to other sorts of inferences, such as those based on set relations. Griggs (1974) has shown that similar underlying processes operate when set relations are substituted for linear orders.

The present work differs from previous studies in that instead of comparing inferred versus presented material, the focus has been on different inference conditions in a task in which <u>S</u>s are forced to draw inferences. Here the question is not whether an inference is drawn but rather what is the nature of the inference. Despite some procedural differences, the basic conclusions are similar. The present work supports the view that the representation formed in a syllogistic inference task resembles those formed in other tasks not generally thought to be similar to logical reasoning. This result can be construed as support for the generality of the "constructivist" approach.

Utility of a Chronometric Analysis

The value of the chronometric method of analysis is that it permits direct inferences about the processing components of the entire reaction process. In particular, it allows one (a) to determine the locus of experimental effects, and (b) to decide whether individual stages of processing are performed serially or in parallel.

Stage models of syllogistic inference have been formulated only recently. Before these developments, most research in syllogistic inference took a more global outlook toward major theoretical concepts. One example is the concept of conversion. Despite extensive discussion of the role of conversions in reasoning (Chapman & Chapman, 1959; Begg & Denny, 1969; Simpson & Johnson, 1966), it was only very recently that Revlis (1975a) made the first concrete proposal as to the temporal aspects of the conversion process (i.e., where it occurs in the series of information processing stages).

Both Revlis and Erickson generated multiple models, testing each by estimating the percentage of responses for which it could account. However, some of the models are quite complex, especially Revlis', and even in the rare case that one model is clearly superior to another, it is not evident which properties of the successful model are critical and which are incidental. While detailed models are ultimately of great interest, there is some question whether existing methods of error analysis permit discrimination between alternate proposals.

Use of RT as a dependent measure allows one to ask more general questions (e.g., does logical complexity affect comparison processes?). By establishing these general conclusions, certain classes of models are singled out as initially acceptable and deserving of further analysis. Just as important, other kinds of models can be ruled out even before all their details are worked out (e.g., the complete integration model). Thus, the chronometric method is well suited to the level of questions that remain unanswered.

Future Directions

Some of the current findings may serve as examples of the kinds of more detailed analyses that are possible. Two examples are the non-additive premise negation effect and the transitive storage effect.

Recall that the times for the condition with negatives in both the first and second premise were less than would be expected by assuming that encoding times for the two simple negation effects were purely additive. This underadditive effect was depicted in Figure 11 of Chapter 2. Several possible models of the encoding process were considered as explanations for this finding. Among the most serious candidates were: (a) a parallel processing model that assumed that the encoding of the two premises overlaps in time and (b) a complex effects model that assumed that while the second premise was not encoded until Stage 1 was completed, congruence of form with the first premise facilitated second premise encoding. In the discussion of Chapter 2, the second alternative was considered more likely, mainly because the double negative condition was generally easier than the single negative condition. While an overlap model can predict an attenuated negation effect, it would be hard pressed to predict a reversal of the effect.

One way to disentangle the two models is to present the premises sequentially. A parallel process model would predict that the non-additive effect would not appear if _S_s were forced to encode the two premises separately. Alternatively, the complex effects hypothesis would maintain that the effect should occur even if the premises do not appear together.

A second avenue for further research concerns the storage effect that was found. It was argued that logically more complex problems take longer to integrate than logically simpler problems, but the present study (purposely) confounded two types of complexity. One is the number of set relations, the other their compatibility. These two factors can be evaluated by examining problems with multiple yet consistent relations. One may speculate that the integration process consists, first, of a <u>generation</u> process that computes set relations followed by a <u>decision</u> process that chooses the one relation to be stored in memory. If it is the decision process that truly discriminates logically complex and simple problems, then a multiple-but-consistent problem should be no harder to integrate than a problem that generates just one set relation.

The general conclusion is that while the present series of studies still leaves us ignorant about many critical details of the stage model discussed, some of these more penetrating questions are within the grasp of the RT method. The schematic stage model has been of heuristic value both in the organization of the findings and in the identification of areas of experimental investigation. The present set of studies provides ample evidence that RT can be a valuable tool in the analysis of logical inference.

REFERENCES

Anderson, J. R. Verbatim and propositional representation of sentences in immediate and long-term memory. Journal of Verbal Learning and Verbal Behavior, 1974, 13, 149-162.

Begg, I., & Denny, J. P. Empirical reconciliation of atmosphere and conversion interpretations of syllogistic reasoning errors. Journal of Experimental Psychology, 1969, 81, 351-354.

Bransford, J. D., Barclay, J. R., & Franks, J. J. Sentence memory: A constructive versus interpretive approach. Cognitive Psychology, 1972, 3, 193-209.

Bransford, J. D., & Franks, J. J. The abstraction of linguistic ideas. Cognitive Psychology, 1971, 2, 331-350.

Ceraso, J., & Provitera, A. Sources of error in syllogistic reasoning. Cognitive Psychology, 1971, 1, 400-410.

Chapman, L. J., & Chapman, J. P. Atmosphere effect re-examined. Journal of Experimental Psychology, 1959, 58, 220-226.

Clark, H. H. Linguistic processes in deductive reasoning. Psychological Review, 1969, 76, 387-404.

Clark, H. H. Semantics and comprehension. In T. A. Sebeok (Ed.), Current trends in linguistics, Vol. 12: Linguistics and adjacent arts and sciences. The Hague: Mouton, 1974.

Clark, H. H., & Chase, W. G. On the process of comparing sentences against pictures. Cognitive Psychology, 1972, 3, 472-517.

Erickson, J. R. A set analysis theory of behavior in formal syllogistic reasoning tasks. In R. L. Solso (Ed.), Theories in cognitive psychology: The Loyola symposium. Potomac, Md.: Lawrence Erlbaum Associates, 1974.

Erickson, J. R. Memory search in formal syllogistic reasoning. Paper presented at Midwestern Psychological Association, May 1972.

Evans, J. St. B. T. Reasoning with negatives. British Journal of Psychology, 1972, 63, 213-219. (a)

Evans, J. St. B. T. Interpretation and matching bias in a reasoning task. Quarterly Journal of Experimental Psychology, 1972, 24, 193-199. (b)

Evans, J. St. B. T. On the problems of interpreting reasoning data: Logical and psychological approaches. Cognition, 1972, 1, 373-384. (c)

Evans, J. St. B. T., & Lynch, J. S. Matching bias in the selection task. British Journal of Psychology, 1973, 64, 391-397.

Frase, L. T. Validity judgments of syllogisms in relation to two sets of terms. Journal of Educational Psychology, 1966, 57, 239-245.

Frase, L. T. Associative factors in syllogistic reasoning. Journal of Experimental Psychology, 1968, 76, 407-412.

Griggs, R. A. Logical errors in comprehending set inclusion relations in meaningful text. Indiana Mathematical Psychology Report No. 74-7, 1974.

Huttenlocher, J. Constructing spatial images: A strategy in reasoning. Psychological Review, 1968, 75, 550-560.

Johnson, D. M., & Jennings, J. W. Serial analysis of three problem-solving processes. Journal of Psychology, 1963, 56, 43-52.

King, D. R. W., & Greeno, J. G. Invariance of inference time when information was presented in different linguistic formats. Memory & Cognition, 1974, 2, 233-235.

Kintsch, W. The representation of meaning in memory. Hillsdale, New Jersey: Lawrence Erlbaum Associates, 1974.

Kintsch, W., & Monk, D. Storage of complex information in memory: Some implications of the speed with which inferences can be made. Journal of Experimental Psychology, 1972, 94, 25-32.

Lippman, M. Z. The influence of grammatical transform in a syllogistic reasoning task. Journal of

Verbal Learning and Verbal Behavior, 1972, 11, 424-430.

Meyer, D. E. Verifying affirmative and negative propositions: Effects of negation on memory retrieval. In S. Kornblum (Ed.), Attention and Performance, IV. New York: Academic Press, 1973.

Meyer, D. E. Long-term memory retrieval during the comprehension of affirmative and negative sentences. In A. Kennedy & A. Wilkes (Eds.), Studies in long term memory. London: Wiley, 1975.

Neimark, E. D., & Chapman, R. H. Development of the comprehension of logical quantifiers. In R. Falmagne, (Ed.), Reasoning: Representation and process. Hillsdale, New Jersey: Lawrence Erlbaum Associates, 1975.

Pachella, R. G. The interpretation of reaction time in information-processing research. In B. Kantowitz (Ed.), Human information processing: Tutorials in performance and cognition. Hillsdale, New Jersey: Lawrence Erlbaum Associates, 1974.

Pachella, R. G. Additive factors and underadditivity within stimulus encoding processes. Paper presented at the Midwestern Psychological Association, 1976.

Parkman, J. M., & Groen, G. J. Temporal aspects of simple addition and comparison. Journal of Experimental Psychology, 1971, 89, 335-342.

Parrott, G. L. The effects of premise content on accuracy and solution time in syllogistic reasoning. Unpublished Master's thesis. Michigan State University, 1967.

Parrott, G. L. The effects of instructions, transfer, and content on reasoning time. Unpublished doctoral dissertation. Michigan State University, 1969.

Pezzoli, J. A., & Frase, L. T. Mediated facilitation of syllogistic reasoning. Journal of Experimental Psychology, 1968, 78, 228-232.

Potts, G. R. Information processing strategies used in the encoding of linear orderings. Journal of

Verbal Learning and Verbal Behavior, 1972, 11, 727-740.

Potts, G. R. Bringing order to cognitive structures. In F. Restle, R. Shiffrin, J. Castellan, H. Lindman, & D. Pisoni (Eds.), Cognitive theory, Vol. 1. Hillsdale, New Jersey: Lawrence Erlbaum Associates, 1975.

Potts, G. R., & Scholtz, K. W. The internal representation of a three-term series problem. Journal of Verbal Learning and Verbal Behavior, 1975, 14, 439-452.

Revlis, R. Two models of syllogistic reasoning: Feature selection and conversion. Journal of Verbal Learning and Verbal Behavior, 1975, 14, 180-195. (a)

Revlis, R. Syllogistic reasoning: Logical decisions from a complex data base. In R. Falmagne (Ed.), Reasoning: Representation and process. Hillsdale, New Jersey: Lawrence Erlbaum Associates, 1975. (b)

Roberge, J. J. A re-examination of the interpretations of errors in formal syllogistic reasoning. Psychonomic Science, 1970, 19, 331-333.

Roberge, J. J. Further examination of mediated associations in deductive reasoning. Journal of Experimental Psychology, 1971, 87, 127-129.

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

Shanteau, J., & McClelland, G. H. Mental search processes in problem solving. Memory & Cognition, 1975, 3, 627-634.

Simpson, M. E., & Johnson, D. M. Atmosphere and conversion errors in syllogistic reasoning. Journal of Experimental Psychology, 1966, 72, 197-200.

Smith, E. E. Choice reaction time: An analysis of the major theoretical positions. Psychological Bulletin, 1968, 69, 77-110.

Staudenmayer, H. Understanding conditional reasoning with meaningful propositions. In R. Falmagne (Ed.), Reasoning: Representation and process. Hillsdale, New Jersey: Lawrence Erlbaum Associates, 1975.

Taplin, J. E. Reasoning with conditional sentences. Journal of Verbal Learning and Verbal Behavior, 1971, 10, 218-225.

Taplin, J. E., & Staudenmayer, H. Interpretation of abstract conditional sentences in deductive reasoning. Journal of Verbal Learning and Verbal Behavior, 1973, 12, 530-542.

Trabasso, T. Representation, memory, and reasoning: How do we make transitive inferences? In A. D. Pick (Ed.), Minnesota symposia on child psychology, Vol. 9. Minneapolis: University of Minnesota Press, 1975.

Trabasso, T., & Riley, C. A. On the construction and use of representations involving linear order. In R. L. Solso (Ed.), Information processing and cognition: The Loyola symposium. Potomac, Md.: Lawrence Erlbaum Associates, 1975.

Trabasso, T., Riley, C. A., & Wilson, E. G. The representation of linear order and spatial strategies in reasoning: A developmental study. In R. Falmagne (Ed.), Reasoning: Representation and process. Hillsdale, New Jersey: Lawrence Erlbaum Associates, 1975.

Trabasso, T., Rollins, H., & Shaughnessy, E. Storage and verification stages in processing concepts. Cognitive Psychology, 1971, 2, 239-289.

Traub, B., & Erickson, J. R. Determinants of difficulty in judging the validity of syllogistic arguments. Paper presented at the Midwestern Psychological Association, 1975.

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

Wilkins, M. C. The effect of changed material on ability to do formal syllogistic reasoning. Archives of Psychology, 1928, 102, 1-83.

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