December 8, 1952

To the Graduate Council:

I am submitting herewith a thesis written by William E. Morris entitled "A Theoretical-Experimental Analysis of Language: A Study of the Charles Morris Types of Discourse." I recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Psychology.

William O. Jenkins
Major Professor

We have read this thesis and recommend its acceptance:

Gerald A. Porter
Edward L. Cureton
Virgil E. Long
Wilis Moore

Accepted for the Council:

E. G. Wates
Dean of the Graduate School

A THEORETICAL-EXPERIMENTAL ANALYSIS OF LANGUAGE: A STUDY OF THE CHARLES MORRIS TYPES OF DISCOURSE

33

A THESIS

Submitted to
The Graduate Council
of
The University of Tennessee
in
Partial Fulfillment of the Requirements
for the degree of
Doctor of Philosophy

---

by

William E. Morris

December 1952

LIBRARY
UNIVERSITY OF TENNESSEE
KNOXVILLE

ACKNOWLEDGMENT

A dissertation stands as a demonstration of man's individual helplessness--his need for others. So many have assisted the writer in his research that it is impossible to present the entire roster. However, those whose names do not appear should know that their graciousness is neither unappreciated nor forgotten.

The writer is very grateful for the generous cooperation given him by Charles Morris of the University of Chicago. Professor Morris' book, Signs, Language and Behavior, provided the impetus and the framework of this writer's study. Professor Morris offered many valuable suggestions, but he is in no way responsible for any errors of measurement or interpretation.

Special credit is due Dr. William O. Jenkins, who was mainly responsible for the design of the experiments and who patiently suffered all the writer's methodological inadequacies. Dr. Jenkins' assistance has surpassed anything that could be summarized as "guidance"--though this term would adequately define that which any graduate student should reasonably expect.

The writer wishes to thank Dr. Edward E. Cureton, who was the first to conceive of the group-test form for the experiments; he also suggested many of the statistical procedures. The writer has greatly benefited from his objectivity and patience.

Each of the other members of the committee gave special aid. Dr. Virgil Long outlined many sociological uses for the writer's experimental procedure. Dr. Willis Moore's particular competence is evident

305517

throughout, but especially in the sections which deal with the logical and philosophical analysis of language. To Dr. Gerald R. Pascal the writer is deeply grateful for many of the suggested research implications, particularly those in clinical psychology.

Thanks should be expressed to Dr. James H. Elder of The State College of Washington, who supervised the writer's early graduate training and encouraged his linguistic research, and to Dr. L. S. McGaughran, now of the University of Houston, under whose direction the pilot study was conducted.

The writer is most deeply indebted to his wife, who did much more than prepare all the experimental materials, type the manuscript, and patiently correct arithmetical and grammatical errors. Without her encouragement, faith, and help, the work would not have been done.

To all of you--thanks; had you allowed me more time to master rhetoric, this would be better expressed, but not more sincerely intended.

TABLE OF CONTENTS

CHAPTER		PAGE
I.	HISTORICAL BACKGROUND	1
	Preface	1
	Introduction	2
	Language Analysis	3
	Recapitulation	23
II.	TYPES OF DISCOURSE IN THE SEMIOTIC OF CHARLES MORRIS	25
	The Modes of Signifying	28
	The Uses of Signs	31
	The Types of Discourse	35
III.	RATIONALE AND STATEMENT OF THE PROBLEM	40
	Why Another Language Study	40
	Is a Science of Signs Necessary	44
	Why Charles Morris	49
IV.	METHOD	51
	Experiment I	61
	Experiment II	61
	Experiment III	63
V.	RESULTS OF EXPERIMENTS I, II, AND III	65
	Experiment I	65
	Experiment II	72
	Experiment III	76
VI.	EXPERIMENT IV: RESULTS AND DISCUSSION (THE C. M. SCORING KEY)	98

CHAPTER PAGE

VII. EXPERIMENT V: RESULTS AND DISCUSSION OF THE EFFECT OF THE PRETEST 121

VIII. DISCUSSION 134

IX. SUMMARY AND CONCLUSIONS 150

BIBLIOGRAPHY 159

APPENDICES

APPENDIX A 167

APPENDIX B 176

APPENDIX C 210

APPENDIX D 223

LIST OF TABLES

TABLE PAGE

I. Results of the 1-1-4 and 1-4-16 Scoring for Experiment I 68

II. Experiment I: Summary of the Raw Data and Results of the Multiple Regression Analysis 71

III. Experiment II: Summary of the Raw Data and Results of the Multiple Regression Analysis 74

IV. Experiment III: Summary of the Raw Data and Results of the Multiple Regression Analysis 77

V. The Intragroup Coefficients of Correlation 79

VI. Comparison of Workers from Experiments I, II, and III . . 86

VII. Raw Data Summary and Tests of Significance for all Ss Not Used in Worker Analysis 90

VIII. Experiment I: Labels Given by all Ss on Pretest 92

IX. C. M.'s Paragraph Evaluations 100

X. C. M. versus E: Paragraph Scoring 101

XI. Experiment I: Multiple Regression Analysis With the C. M. Scoring Key 109

XII. C. M. Scoring: Experiments I, II, and III 112

XIII. The C. M. Scoring Key Applied to E's "Worker" Group . . . 115

XIV. C. M. Scoring: "Workers" plus and minus One Standard Deviation from the Mean of the Pretest for the Combined Groups 116

TABLES PAGE

XV. Comparisons between Item No. 3 and All Items for Experiments I, II, and III Employing E's Scoring Key . . 118

XVI. Experiment V: Summary of the Results for Both Groups . . 125

XVII. Item No. 3 Only: Comparison of the Workers, Experiments II and III with Pretested Group of Experiment V 131

LIST OF FIGURES

FIGURE PAGE

1. Examples of the Major Types of Discourse Taken from Charles Morris 27
2. Diagram for the Experimental Design of Experiments I and II 57
3. Projected Mode-Use Matrix 146

CHAPTER I

HISTORICAL BACKGROUND

Preface

To add point to the remainder of this chapter, it was felt that a preliminary statement as to the nature of the problem with which we are concerned should be given. This study is NOT directed toward the problem of linguistic communication if a rigorous definition of communication is applied. "Communication means that information is passed from one place to another" (52, p. 6). ("Information" and "content" are defined on page 5 of this thesis.) Instead the experiments later to be described deal with the signification (nature, use, and interpretation) of written language messages. (It is assumed that the responses to written and spoken signs are in some way similar, but this must await further experimentation.) Take this message for example: "Man's days are fruitless and few." You will see later how psycholinguists, phoneticians, ethnologists, logicians, and others would set about analyzing such words. One might ask: (1) What sort of message is it? Is it a scientific statement, fantasy, or what? and (2) What type message does the reader (or listener) think it is? That is, how does he behave? If he interprets it as a scientific utterance, he may prepare for the worst and/or establish an hedonic program for himself. If he interprets the statement as a moralist's admonition, his behavior will be different from that which an empirical evaluation would dictate. Something of this philosophic approach will be discussed.

However, very little if any EXPERIMENTATION has been aimed at answering the above two questions; this thesis is an invitation to "meaning" to enter the laboratory, a small excursion to a land that has been mapped, but the maps have been collectors' items and not mariners' charts!

Introduction

And words--

> They don't seem dangerous certainly, just little puffs of air, little clicks in the mouth, neither one thing nor the other. . . . One's unsuspicious about words . . . and some misfortune ensues.
>
> Among them, there are some hidden away under all the others, like pebbles. You don't particularly notice them and suddenly they've made all the life there is in you tremble. . . . And you're terrified . . . The thing's an avalanche . . . You swing in the air above a torrent of emotion like a hanged man . . . A hurricane has come up and passed on, and it's been much too strong for you, so violent that you'd never have believed it could be as violent as that and yet be made up of nothing but just feeling. . . . . So one's never distrustful enough of words (91, p. 150).

This much is evident: Experimental psychologists have been unsuspicious about words. Boring (8) refers to Broca's speech center and memory studies of nonsense and meaningful material, but there is no reference to a study of language per se. Gardner Murphy (57) mentions the work of eighteenth century German scholars with language and phonetics: Their studies were primarily philological, i.e., concerned with the authenticity and accuracy of historical documents, etc.; Wundt's belief that a given culture could best be described by an analysis of the language employed by the group, Mead's thought as to the importance of language and group membership, and Watson and Holt on word-meanings as conditioned responses are given some space. Language has played a small role at best

in the theories of the various schools of psychology. Heidbreder (36) writes of the importance of talking and psychoanalytic therapy and William James's linguistic sophistication, but these are the only language references. However, many students--for the most part philosophers, logicians, linguists, and anthropologists--have been concerned both with the nature of language itself and with language behavior; this is particularly true of those who followed in the positivistic trend of Auguste Comte. Recently psychologists have moved into the field, particularly in the area of communication. Let us turn now to a review of the several approaches to the study of language.

Language Analysis

The investigation of signs and languages may be undertaken from several vantage points. Sign behavior may be classified in several ways: One might use encoding, organismic, and decoding processes, the traditional framework of grammar, rhetoric, and syntax, and so on. The writer has chosen to codify language studies as to STRUCTURE, CONTENT, AND MEANING. These three headings are further divided as to disciplines and/or methodologies. No plea is made for this particular systematization, no attempt to develop generally useful definitions of these three key terms is made, and--above all--it is not implied that the men used for illustrative purposes confine their work to any ONE of these heuristic areas. The reader is asked--if the need is felt--to derive his own definitions from his reading of a given section; at best multiple definitions would be required. From the standpoint of psychology, the studies to be

reviewed could be separated as to whether stimulus, organismic, or response variables are being investigated. However, the borderlines might be even hazier if this division were employed. For example, communications engineers are concerned with both the morphology and phonology of words (their production and reproduction) and the organism's auditory and verbal apparatuses, as well as the mechanical linkages in the communication chain. How would we classify the work of these engineers in such a behavioral scheme?

Linguistic Structure

Grammar, rhetoric, and syntax. All of us are familiar with this approach, and no elaboration by the writer is necessary. However, it is interesting to note that rhetoric has virtually dropped out of the curricula of United States universities. Kenneth Burke (12) has recently made a plea for the reinstatement of rhetorical analysis favoring an extension of the classical definition of rhetoric, i.e., the persuasive use of language.

Structural linguistics. This method, sometimes called descriptive linguistics, divides the study of language signs into morphology and phonology which in turn are subdivided into morphemic and phonemic elements. The procedures involve the isolation of these elements; their interrelationships and distributions are then analyzed. This methodology is invaluable in the description of primitive languages and second-language teaching, but it is not germane to the problem later to be discussed. The work of Zellig Harris (31) is typical. A related discipline is psycholinguistics, which ". . . is concerned with the processes of

encoding and decoding, i.e., the relations which the message bears to the intentive behavior of the speaker, on the one hand, and to the interpretative behavior of the hearer, on the other" (15, p. 8). Psycholinguistics, as the name implies, is a cooperative venture of psychologists and linguists. They study the communication process under the rubric of information theory. At this early stage psychology is playing a very small part indeed. The emphasis is on information as conveyed by speech. No definition of information or message (content) is stated in the booklet, but Miller writes that

> . . . 'information' is used to refer to the occurrence of one out of a set of alternative discriminative stimuli.
> . . . . . .
> The content of the information concerns the particular discriminative stimulus that does occur (52, p. 41).

With the above definitions in mind, one could say that psycholinguistics, AT THIS TIME, is dealing with the "how" of information; i.e., how information is encoded, transmitted, received, and decoded. Questions as to what type of information is given and what differences do these various types make to the organism are not now being asked. The report states:

> The speech of an individual is bound by the linguistic patterns imposed upon him by his culture and yet is free enough to allow for the expression of his particular personality and current emotional state. While the study of this no-man's-land is obviously the psycholinguist's task, it was also the sense of the Seminar that the unique phenomena nevertheless impinge upon the technical linguist's domain as well, since modifications of the formal structure undoubtedly have their starting point in just this area. Since personality and emotional factors are probably not codified digitally but rather continuously, current linguistic methods may be incapable of describing them (15, p. 19).

(Re digital computing (codification), it should be stated that the work of Miller and the psycholinguists has been profoundly influenced by Shannon's (71) theoretical model.) It is with certain aspects of this

no-man's-land that the present study is concerned.

Phonetics. This study is concerned with the formation and apprehension of language sounds, especially as applied to language teaching and speech correction. Thus it is related both to linguistics and to the neurology of speech and audition. Phonetics is an important tool for speech and reading correctionists.

If one used a different connotation of the word "structure," then logical analyses, Aristotelian and non-Aristotelian, could be subsumed under this classification; certain ethnological, mathematical, and philosophical analyses could be added. It has, however, seemed best to the writer to consider all of these in the section on linguistic meaning.

Communications engineering. Shannon (71) and others emphasize the efficiency of communication; only indirectly do they deal with "meaning." Mathematical models are developed and experimentation is directed toward the process of encoding and decoding spoken or electronic (94) messages, i.e., the nature and perception of speech sounds. This work is of tremendous significance to researchers in psycholinguistics, but more particularly to those who are NOT concerned with connotation (signification) or context (the behavioral implications and settings of messages). If Joe said to Moe, "Hello! How are you?" and Moe replied, "I'm five feet eight," the engineer might first investigate the functioning of the electronic components of the chain and after satisfying himself on this score, announce that his instruments and concepts assumed the normal functioning of the animal factors involved even though some people are funny!

Statistical analysis. Zipf (98) probably has carried this approach furthest; his efforts, though related to descriptive linguistics, are more closely connected to those of the philologists. Elaborate statistical counts and derivations are used to infer the nature of work, culture, personality development and deterioration, etc. His units of measurement include words as well as phonemes, etc. He is dealing with language and meaning in general and thus only illustratively with the meaning of any given unit of discourse and its interpretation by the individual organism. As far as he is dealing with language per se, it may be said that he builds on the semantic counts, etc., of Thorndike and others (86, 17). His findings have considerable significance, but since they do not contribute in any DIRECT way to this study, they will not be discussed more fully.

Linguistic Content and Behavioral Effects

Under this heading a vast literature could be assembled. The range would extend from the Epicureans, who anticipated the nominalists and positivists, with their emphasis on the process nature of reality and the inadequacy of words to express this process, to Osgood (62), who has experimented on the connotative meaning of words. Let us try to summarize some of this work.

Applied psychology. Psychologists have a definite interest in speech and language. Considerable work has been done on the "language" of advertising. Clinicians have been concerned with the aphasias, concept-formation, and symbolism; psychoanalysts in particular devote much thought to symbolism--its ethnological and individual interpretation. It is

interesting to note that the role of the individual organism (his interpretations, etc.) is slighted by many of these students. Testimonials are as effective with John as with Jim (32). Visual agnosia results from injuries to area 18 of the cortex (16); wide individual differences in cortical structure and organization are of little consequence. To dream of a snake is the same for both Minnie and Myrtle (26).

Wendell Johnson (41) exemplifies the approach of therapists (speech correction and personality adjustment) to the language process. Johnson, building on the foundation of Korzybski (43), theorizes about the semantic origin of psychological maladjustment. Unconscious identification and projection, two-valued orientations, the belief in absolutes, elementalism, etc., are fostered by the very Aristotelian structure of Indo-European languages; the logical structure of these languages enhances our faith in words and leads us to search for verbal chimeras and to despair over linguistic phantoms. Prophylaxis is the natural outcome of a semantic sophistication which is founded on consciousness of abstracting, awareness of the process nature of reality, and recognition of the fact that the "word is not the thing." This and similar points of view will be considered further in the section on linguistic meaning. (See Cameron (13) for the application of linguistic analysis to schizophrenic processes.)

Propaganda and content analyses have been undertaken by applied psychologists (these studies will be discussed later in this chapter). The findings of Flesch (21, 22) and Thorndike (85) contribute to our knowledge of vocabulary difficulty and readability. Their work helps one get

a message across to a given audience and thus is related (on the content side) to psycholinguistics.

Experimental psychology and learning theory. Workers in these fields have given much to students of language, too much for any short-space summary. Miller (52) gives the most complete presentation of objective data re language. It is no comment on Miller but rather on the nature of the present writer's problem that from the very extensive bibliography given in Language and Communication, only eighteen references were considered as being of CENTRAL importance to the present study. In fact, Miller has written, "Philosophical discussions of language are somewhat outside the scope of experimental psychology . . ." (53, p. 789). He means by this that, as of this date, scant empirical data are available in the field of "meaning," i.e., these discussions have not led to experimentation, and that he (like Shannon and the psycholinguists) is primarily dealing with verbal content. This is NOT to say that Professor Miller is unaware of these problems (he has done the most extensive work with language among psychologists) but only that he has not turned his attention to them. C. E. Osgood, in reviewing Language and Communication, has this to say:

> Another outgrowth of the limited learning model (B. F. Skinner's) is Miller's treatment of meaning--or better, lack of treatment. The entire problem of linguistic forms as signs having representational properties, of the development and measurement of significance, is missing. This also presumably reflects Miller's belief that we cannot be scientific and objective about meaning (63, p. 362).

Of the learning theorists Skinner (74, 75, 76) is perhaps the one most concerned with language processes. Since his book is not available

to the writer, no extensive comment should be made on his position. On the basis of his learning theory, he has discussed the variables which control verbal behavior. It is just possible that Charles Morris' (54) semiotic, which provides the basis for the experiments later to be discussed, could be partially translated into Skinner's terminology and vice versa. Skinner is an associationist, and Morris has been greatly influenced by Hull and Tolman. The possibility of this translation is intriguing, but it must be postponed.

To revert to Osgood's evaluation of Miller's book, it would seem that his conclusions may be too strong. It is not that one cannot be scientific about meaning or that Skinner's model is inadequate, but rather that experimental verification is lacking; this does not mean that it won't be forthcoming. As a matter of fact, the group with whom Osgood is associated has doubts similar to Miller's (see page 5).

In important ways the work of Watson (90), who virtually equated thought and subvocal speech, and many studies dealing with the relationship between language and perception (46) should be included even in a brief survey; they <u>are</u> contributory but not necessarily propaedeutic, i.e., if signs and sign processes are defined and investigated in terms of stimulus objects and response sequences. Watson was theorizing as to the nature of <u>thought</u>, and only incidentally was he concerned with the nature of <u>language</u> per se. The student of communication will be influenced by the findings of other investigators, but <u>they</u> will have to tell <u>him</u> how it is with perception, thinking, and learning; he will tell them of language and behavior. It just might be that a <u>science of sign-behavior</u>

is, in fact, the prior desideratum.

Developmental psychology. Investigators of child behavior ask questions regarding the origin and growth of written and spoken language behavior in the child, the representational character of words, and the relation of language to thought processes (29). Changes in concept formation are stressed by Piaget (65) and others. Child psychologists have emphasized the acquisition of language and have not delved deeply into the emotional and referential character of words or into individual differences re connotation.

Propaganda and content analyses. Numerous scholars representing many disciplines have explored the nature, creation, dissemination, and effects of propaganda; an adequate review is impossible here, and we shall offer only some comments on Harold Lasswell and his associates, who typify this orientation (44, 45). (Content analysis involves "meaning," but for titular reasons--if nothing else--it is taken up here rather than in the section on linguistic meaning.) The data employed by these analysts are words, or better, "words about words." In The Analysis of Political Behaviour (44, pp. 158-172), Lasswell and Renzo Sereno describe the changing structure of the Italian ruling group before and during the Fascist regime. Employing an Italian "Who's Who," newspaper editorials, etc., and using a well-defined frame of reference (skill analysis), they were able to show the social strata from which new political leaders were recruited. With this information (the writer is oversimplifying this), they were able to make predictions as to the trend of governmental policies, etc. Propaganda is evaluated as an important social tool, one that

should be respected, sharpened, and used. Workers in this area are prone to talk to one another; not that they are against the wide dissemination of this information (Lasswell himself has extended the implications of his research to psychology, the law, sociology, etc.), but rather, in these early stages, they are not introducing John Doe to the mysteries of propaganda; instead they seem bent on protecting and guiding him. These investigations are thrilling and important, but it could not be determined how these methodologies could be used to investigate the problem with which this study is concerned.

Linguistic Meaning

The literature of "meaning" is vast indeed. As a philosophical problem its history is extensive. In a survey such as this, one must be a bit arbitrary in selecting a starting point. The writer chose Ogden and Richards' The Meaning of Meaning (61) for several reasons: It provided the impetus for much of the recent work--i.e., it reopened this area of scholarship; the book is the product of a psychologist and a literary critic-esthetician and thus presages the efforts of the "Unity of Science" group; and it was the first book to truly reveal these semantic problems to the writer. Many men and disciplines have contributed, and one must select from among them. Even the units of language measurement are varied, ranging from connotational definitions of words (62), through the reading of a page (67), to reading a book (1). (One could include phonemes and morphemes as "meaningful" units, and certain systems probably function best if the complete works of a given writer are evaluated. So far as the writer knows, the present study is the first to use the

paragraph as a unit of measurement. Tests of paragraph-comprehension and vocabulary cannot be counted, since the question asked is: Do you know WHAT you have read? and not: Do you know the MEANING of what you have read?) Ogden and Richards and others have clearly shown what a slippery word "meaning" really is; countless definitions are possible and acceptable. Charles Morris writes:

> The term 'meaning' is not . . . among the basic terms of semiotic. This term, useful enough at the level of everyday analysis, does not have the precision necessary for scientific analysis.
> . . . . . .
> 'Meaning' signifies any and all phases of sign-processes (the status of being a sign, the interpretant, the fact of denoting, the significatum), and frequently suggests mental and valuational processes as well; hence it is desirable for semiotic to dispense with the term and to introduce special terms for the various factors which 'meaning' fails to discriminate (54, p. 19).

(Consult the Glossary in Appendix A for definitions of Morris' terms.) Since this chapter is in fact an everyday analysis, the term has been retained. Another important consideration should be mentioned: Quine (66) pleads for the separation of meaning and reference and warns against the unwarranted extension of singular terms (naming).

> The main concepts in the theory of meaning . . . are synonymy (or sameness of meaning), significance (or possession of meaning), and analyticity (or truth by virtue of meaning). Another is entailment, or analyticity of the conditional. The main concepts in the theory of reference are naming, truth and denotation (or 'truth-of', in the sense in which the term 'green' denotes or is true of each green thing). Another is extension, or the class of denoted objects; also the notion of values of variables (66, p. 91).

In anticipation and with these statements by Morris and Quine in mind, it may be said that this writer's study deals primarily with the SIGNIFICATION aspect of meaning and in no way with denotation or reference (at least not directly). This subject will be brought up again and discussed

more fully in Chapter II.

It will be obvious, in that which follows, that different aspects of meaning are being discussed. The men selected represent the fields of psychology, sociology, ethnology, logic, esthetics, etc., but the discussion has been divided into four areas: social science, linguistic science, criticism, and philosophy. It is interesting that many of these scholars resist classification; the writer knows not how Harold Lasswell labels himself, but he hesitates to label him. This may indicate how far the departure from departmentalization has progressed in this area and the genuine possibility that the linguistic unification of science is at hand. WHEN THESE MEN INVESTIGATE LANGUAGE BEHAVIOR, the major difference between them is in choice of methodology. Classificatory arbitrariness should be admitted in relation to the men we have discussed previously. Virtually all of them work with structure, content, and meaning; it is only what the writer considers to be their primary emphasis which has caused him to include them in the sections on structure and content. Much of the remainder of this chapter has been freely adapted from Charles Morris (54) and George A. Miller (52).

Social science. Several psychological workers who have made contributions to language-meaning have already been mentioned. However, other approaches should be noted. For example, Baldwin (3), using a variation of content analysis, examined the letters of a mother to check his prior clinical interpretation of her personality structure. He isolated fifteen discussion topics, which in turn were subdivided into attitudes. Consistencies were then sought; e.g., negative remarks about her

son were associated with negative remarks about women. This is an interesting example of a "second-order" word analysis. Certain experiments with syllogisms are informative. Woodworth and Sells (97) have shown that the choice of words used in a question have much to do with the answer to it. Words like "all" and "no," when used in syllogistic premisses, call out a corresponding response in the conclusions ("atmosphere effect"). This type of error shows a slight positive correlation with intelligence. Maier (47) demonstrated the importance of preliminary verbal instructions for problem-solving behavior. Social psychologists are most concerned with the circulation of information, propaganda, and rumor. Certain studies examine the dynamics of communication nets: Bavelas' (4) experiment illustrates how a breakdown in the channels of communication can prevent the solution of a group task; questions as to individual interpretations of the messages were not asked. Statistical evaluations of the effects of mass media of communication are presented by other social investigators. Boder's (6) findings are somewhat related to our own experiments. By means of the adjective-verb quotient, Boder was able to show reliable differences among authors, different samples from the same author, and various types of written material (dramatic dialogues, Ph. D. theses, legal tracts); e.g., fiction had about three verbs per adjective, while legal statutes had five. This very promising study would seem to have many ramifications, but--so far as can be ascertained--additional investigations have not been carried out.

Psychopathologists have tackled the "psycho-semantic" problem. Johnson (41) and the American Psychiatric Symposium (13) have been

mentioned. Freud (25) and Goldstein (30) have made contributions, but their conceptions will not be considered here because they deal with the more nominal aspects of symbolic processes (i.e., dream interpretation, aphasias, etc.), and the individual organism's interpretation ("meaning") is--or is assumed to be--straightforward; for the problems with which they were dealing, questions about the nature of verbal stimuli were not asked. Jules Masserman, on the important connection between signs and neurotic behavior, writes: "The motivations of behavior . . . may become conflictful, either because of the simultaneous evocation of antagonistic inner needs or through the necessity of difficult adjustments to complex and contradictory external symbolisms" (50, p. 6). For other psychiatric evaluations as to the significance of symbolic behavior in human adjustment, see Sullivan (84). Henry Winthrop (96) offers a study which has more direct bearing on the present writer's problem. Using the Aristotelian Diagram of Opposition (all S is P; no S is P; some S is P; and some S is not P), he constructed two forms of an attitude test. The sentences of Test II were related to those of Test I as logical contradictions, contrarieties, or equivalences. "All Europeans want to depend upon Uncle Sam in order to stabilize themselves economically." This might be a sentence in Test I. The subject was asked to check this item "Yes," "No," or "?." Two weeks later in Test II this version would appear: "Some Europeans want to get on their own feet economically without the help of Uncle Sam." What of the subject who checks "Yes" to this proposition in Test I and "No" to its counterpart in Test II? Winthrop found that the attitude-consistency score for an "elite group" was 47 where 100 was the

maximum possible score. The author gives a challenging interpretation of his findings in terms of conflict and rigidity. Obviously, DESPITE individual consistency or lack thereof, certain sentences, although logically similar, hold different meanings for the subjects; this task was not taken by Winthrop. Since he was not investigating similarities per se, the subjects were not asked whether or not they considered two propositions to be equivalent. The writer once overheard two coeds, and their conversation went like this: A. "I don't like to shop up North; those Yankees shove you around so." B. "I'm a Yankee, and I don't shove anyone around." A. "Oh, you're from Ohio, and I don't consider people from Ohio Yankees." This story brings out one question that the present study seeks to answer: What sort of writing (type of discourse) do you consider this paragraph to be? In terms of the analogy: What is your definition of Yankee?

Linguistic science. Anthropologists and ethnologists are included under this heading; often the man up for discussion is an expert in both cultural and linguistic sciences. From among the anthropologists one immediately thinks of Malinowski (48), Warner (89), Trager (77, 88), Whorf (92, 93), and many others; and of the linguists, Bloomfield (5), Gardiner (28), and Sapir (68, 69, 70). This discussion is limited to the work of B. L. Whorf and Edward Sapir.

Whorf's most original contributions have been in the area of metalinguistics.

> Microlinguistics concerns itself only with differential meaning, that is, whether utterance fractions are the same or different. Metalinguistics, however, which may be described as the study of what people talk (or write) about and why, and how they

> react to it, is concerned with meaning on all levels. Where microlinguistics stops with the consideration of the structure of the sentence, metalinguistics deals also with the organization of sentences into discourse and the relation of the discourse to the rest of the culture (77, p. 155).

(Exolinguistics has been offered as a less confusing title than metalinguistics (15, p. 4), since it serves to distinguish this term from metalanguage, i.e., language about language, as used by semioticians.) More particularly, Whorf investigated the relationship between language structure and the ordering of experience; e.g., what differences in "world-view" result because English is a "temporal" language while Hopi is not? "The timeless Hopi verb does not distinguish between the present, past, and future of the event itself but must always indicate what type of validity the speaker intends the statement to have: . . ." (92, p. 7). Re thought, language, and behavior Whorf (92, p. 27), in analyzing reports which dealt with the starting of fires, found that workers exercised great caution when around stored "gasoline drums" but became careless in the storage of "empty gasoline drums," which are in fact (if not in word) the more dangerous. Scientists themselves are not immune:

> Science . . . has not yet freed itself from the illusory necessities of common logic which are only at bottom necessities of grammatical pattern in western Aryan grammar; necessities for substances which are only necessities for substantives in certain sentence positions, necessities for forces, attractions, etc. which are only necessities for verbs in certain other positions, and so on (93, p. 188).

For example, English grammar requires and/or condones the use of a substantive ("force") and may lead us to search for the substance instead of defining it. The writer hopes that this cursory review will give you something of Whorf's conception of the linguistic patterning of thought.

Whorf was a student of the linguist and cultural anthropologist, Edward Sapir, who stressed the hold of language upon thought processes. Sapir's efforts extended from investigations of the relationship between speech melody, etc., and personality (68) to the structural analysis of American Indian languages (70). In a sense he personalized anthropology and related it to psychology and psychiatry. Perhaps his greatest contributions were theoretical orientations which greatly influenced social scientists in their question-framing and methodology.

Criticism. Anyone who has been through sophomore English has been exposed to the "what-was-the-author-trying-to-convey" type of criticism; this requires no elaboration. The semantic approach, however, should be briefly discussed. Hayakawa (34) has popularized general semantics and turned it to good account in interpreting ethics, art, poetry, etc. (Further discussion of general semantics will be reserved for the following section.) I. A. Richards (61, 67) has been introduced. Kenneth Burke is now developing a metalanguage, not in semiotic terms, but rather in terms of a grammar (11), rhetoric (12), and symbolic (not yet published) of motives. In the grammar he codifies motives by five key terms: ". . . any complete statement about motives will offer some kind of answers to these five questions: what was done (act), when or where it was done (scene), who did it (agent), how he did it (agency), and why (purpose)" (11, p. xv). The rest of the book is an elaboration of these analytic terms using literary and philosophical examples. Burke broadens the classic concept of rhetoric as persuasive discourse into IDENTIFICATION; for him it is the mechanism of identification that leads--via

language—to human strife or cooperation.

> Thereafter, with this term as instrument, we seek to mark off the areas of rhetoric, by showing how a rhetorical motive is often present where it is not usually recognized, or thought to belong. In part, we would but rediscover rhetorical elements that had become obscured when rhetoric as a term fell into disuse, and other specialized disciplines such as esthetics, anthropology, psychoanalysis, and sociology came to the fore. . . .
>
> But besides this job of reclamation, we also seek to develop our subject beyond the traditional bounds of rhetoric. There is an intermediate area of expression that is not wholly deliberate, yet not wholly unconscious. It lies midway between aimless utterance and speech directly purposive.
>
> . . . . . .
>
> Here is a rhetorical area not analyzable as sheer design or as sheer simplicity. And we would treat of it here (12, pp. xiii-xiv).

(It is not certain that many of Burke's hypotheses would not lend themselves to experimental test; however, the writer has given but little thought to the task.)

Philosophy. Under this heading are included logicians and semioticians. Analysis of language "meanings," or the science of signs and sign-behavior, is called semiotic, a term used by Locke, Peirce, and—more recently—Charles Morris. (Semantic is often a synonym for semiotic; but because of its more specialized significance in the usage of Charles Morris and certain others, semantics is considered as one aspect of semiotic.) Semiotic is divided into PRAGMATICS, which investigates the psychosocial functions of words; SEMANTICS, the study of the meaning of signs, expressions, or terms, i.e., the signification and thus the designation of words; and SYNTAX, or SYNTACTICS, which ignores meaning in any referential sense and investigates words in relation (connection) WITH ONE ANOTHER. These three areas are capable of being analyzed from at least three points of view: descriptive (metalinguistics), applied

(propaganda-analysis), and pure (metalanguage-construction). An adequate coverage of semiotic will not be attempted; instead certain high spots will be etched merely to provide a setting for Charles Morris' behavioristic science of signs. (Feigl and Sellars (20) have provided a good compilation of the modern literature of philosophical analysis. Many of the essays deal directly with semiotic; all are influenced by semantic concepts. The following discussion draws heavily from this source.)

Charles Peirce was, very possibly, the American most aware of linguistic pitfalls. For Peirce pragmatism was an axiom of logic; he stated it thus: "In order to ascertain the meaning of an intellectual conception one should consider what practical consequences might conceivably result by necessity from the truth of that conception; and the sum of these consequences will constitute the entire meaning of the conception" (64, p. 6). This maxim may be called the rallying cry for many present-day philosophers of meaning. An extension of this maxim which found a wide and responsive audience was Bridgman's (10) statement of operationism, i.e., a concept means nothing more than a set of operations. (For a concise statement on operationism, consult the "Feigl Treatment" in Appendix B.) Ayer (2) formulates it differently but to the same effect. The tool of symbolic logic aided the so-called "Vienna Circle" (logical positivists or empiricists) and others in their attack on the logical structure of language: Schlick, Wittgenstein, and others maintained the experiential (Socratic) approach, while Carnap, Hempel, and Neurath advanced a physicalistic or behavioristic theory. Carnap (14), who may be taken as representative of at least his group and its methodology, has derived a

complete linguistic system which deals primarily with the logico-mathematical aspects of language, i.e., semantics and syntactics. Since the present study may be said to deal with applied pragmatics, the work of Carnap (and others) will not be explored further at this time. (For a good review of logical empiricism, see Jørgensen (40).)

If, as Peirce would have it, signs (or words) modify the organism's response (action) tendencies, how shall we go about classifying words? Ogden and Richards (61) have classified words as "emotive" and "referential," and these have four main expressive functions: symbolization of referent, attitude towards listener, attitude towards referent, and advancement of intended effects.

> The contrast occurs in many other forms: various writers distinguish cognitive and non-cognitive signs, cognitive and instrumental signs, referential and expressive signs, referential and evocative signs, and then refine these basic dichotomies by various and varying subdivisions of each of the contrasted classes (54, p. 60).

Such classifications have been called by Charles Morris modes of signifying. A mode of signifying is defined as "a differentiation of signs in terms of the most general kinds of significata" (54, p. 351), a significatum being "the conditions such that whatever meets these conditions is a denotatum of a given sign" (54, p. 354). Signs may also be classified in terms of the <u>use</u> to which they are put. Use and modes of signifying will be discussed more fully in Chapter II. So much for classification.

The general semantics of the late Alfred Korzybski (34, 41, 43) represents an attempt to apply the tenets of a non-Aristotelian logic to daily life. Hayakawa writes:

> Underlying our 'evaluations,' or 'semantic reactions,' says

> Korzybski, are neurologically channelized epistemological and linguistic assumptions. These assumptions, when infantile, primitive, or unscientific, lead not only to confusions and perplexities in discourse and discussion, but also to misevaluations in everyday life: misevaluations which, when serious enough, require the attention of psychiatrists (33, pp. 118-119).

Korzybski offers many devices which are designed to remind us of the nature (metaphysics and logic) of language and to help prevent misevaluations. Since 1933 many interesting applications of general semantics have appeared in the literature. General semantics is aimed mostly at reference, i.e., naming, denotation, etc.; while the writer's studies, it will be recalled, deal with signification.

Recapitulation

We have seen how investigators, using a variety of techniques, have advanced on the language problem from all sides. Linguistic structure has been described from the standpoint of logic and by statistical and phonetic analysis. Content has been assayed in terms of propaganda, nonsense syllables, etc. Meaning has been discussed philosophically, philologically, and experimentally. Many disciplines have contributed to these analyses, and different scholars have accented stimulus, response, and/or organismic variables. The units of measurement have ranged from the phoneme to the complete works of an author.

In the area of meaning, however, there has been but little experimentation, and that has been directed at restricted units of language and/or vocabulary or dictionary meaning. Indeed, experimental psychologists have avoided the issue or despaired of developing methods for

handling it. Those who have written most provocatively AND often scientifically about the theory of meaning have not actively sought interpersonal, empirical verification or else have not framed their hypotheses in terminologies which lend themselves to experimental test. Doubtless the approach of the metalinguists is capable of extension. It is hard to tell how the INDIVIDUAL reader (or listener) could use the methods of content analysis in any practicable way, because of lack of time, for example. Ability to use Korzybski's "structural differential" could be measured, but resulting behavioral changes due to training in general semantics would involve the student in questions of criterion reliability and validity that would be difficult, although not impossible, to solve. Osgood has not developed his method to the point where it can handle units larger than a single word or concept. Like general semantics the metalanguage which Kenneth Burke has developed (and is still developing) would lend itself to a more empirical handling, but as yet this has not been accomplished.

Of those semioticians who have tried to develop a comprehensive science of signs (at least a metalanguage of semiotic), Charles Morris seemed best suited for our purpose; best because the CENTRAL concepts can be readily grasped and because the systematic foundation is behavioral and consequently testable. We turn now to examine CERTAIN ASPECTS of this system.

CHAPTER II

TYPES OF DISCOURSE IN THE SEMIOTIC OF CHARLES MORRIS

Before developing the semiotic of Professor Morris in some detail, we should say a word perhaps about George H. Mead. Neurath (58, 59) set the framework for Morris' metalanguage in "behavioristics," Hull and Tolman influenced its terminology and methodology, while Mead provided its direction and goal. Mead (51) gave much thought to the genesis of language which was for him the gesture or significant symbol; if a gesture implies the idea behind it and brings out that idea in another, it becomes a significant symbol, a true social communality. (Where Mead uses "idea," the substitution of "behavior" or "response sequences" is warranted.) To have a self one must live in a society that has a language composed of symbols, and the symbolic content must be to a large extent interpersonal, i.e., shared by all members of that society or culture; a symbol is significant only if the user of it can take the role of one who hears it. Mead's significant symbol includes not only the social but the individual sign-substitutes or symbols as well. In this realm interpersonality does not exist; here the individual is "talking to himself." Mead's work is more suggestive than systematic; i.e., he did not develop a complete theory of language, but he was influenced by Pavlov's conditioning experiments, which helped him point the way to a less mentalistic conception of language and meaning.

While it is not possible to properly evaluate the background of Professor Morris' science of signs at this time, if the reader will keep

in mind the three influential men mentioned above and will consult the Glossary (prepared by Dr. Morris) in Appendix A, he should be able to realize something of the psychological tenor of the system.

Professor Morris tried to develop a science of signs capable of handling ALL signs, language and non-language, as employed by all organisms; we shall be actively dealing with written words only, and these will be discussed from but one of the many possible viewpoints. The present writer's studies do NOT cover all of Morris' metalanguage, and will, in no way, take the place of his book and articles for those who want to understand his conceptions fully. These studies are designed to test the workability of the Charles Morris system (mode-use) of classifying types of discourse. A type of discourse is defined by Morris as: "A specialization of language for the accomplishment of specific purposes. The types of discourse are here classified on a mode-use basis, that is, in terms of the combination of some dominant mode of signifying with some dominant use" (54, p. 356). (There are four modes and four uses which yield sixteen major types of discourse. Types of discourse and other technical terms are defined in Appendix A.) For a simplified discussion of types of discourse as used in the experiments--employing a somewhat different approach and with certain titles changed--see the "Morris Treatment" in Appendix B.

The Charles Morris mode-use classification can be put in the form of a matrix; the intersection (common cell) of mode and use names the type of discourse, or better, an example of that type of discourse. This matrix is reproduced on the following page.

Use / Mode	Informative	Valuative	Incitive	Systemic
Designative	Scientific	Fictive	Legal	Cosmological
Appraisive	Mythical	Poetic	Moral	Critical
Prescriptive	Technological	Political	Religious	Propagandistic
Formative	Logico-mathematical	Rhetorical	Grammatical	Metaphysical

Figure 1

Examples of the Major Types of Discourse
Taken from Charles Morris (54, p. 125)

The Modes of Signifying

From Signs, Language and Behavior, we learn that a mode of signifying is:

> A differentiation of signs in terms of the most general kinds of significata. Five modes of signifying are distinguished (identificative, designative, appraisive, prescriptive, and forma ve, an signs signifying in these modes are called respectively identifiors, designators, appraisors, prescriptors, and formators (54, p. 351).

(We will not be concerned with the identificative mode of signifying.) Sentences or ascriptors in the DESIGNATIVE MODE indicate characteristics of the environment; those in the APPRAISIVE MODE assay the capability of an environment to satisfy a need; and those in the PRESCRIPTIVE MODE specify ways of behavior which will lead to the satisfaction of a need. The FORMATIVE MODE is difficult for some to comprehend; signs in this mode are often called formal or logical signs--"not," "or," "&," "( )"--as opposed to lexicative signs. Morris writes:

> If a dog were trained to respond differentially to five food pans containing food when each of five sounds was produced, and now if a sixth sound were introduced always with some two of the other sounds, and if the dog then always went first to one of the two designated pans and to the other if and only if food was not obtained in that pan which it first approached, it might be said that the sixth sound signified to the dog what 'or' in one of its English usages signifies (54, p. 87).

What does the addition of this "or" do to the response sequence? What are the characteristics of this sign "or"? Charles Morris lists four:

> 1. Some stimulus is added to signs which already have a pluri-situational signification . . . ;
> 2. When the new factor is added the signification of the particular sign combination in which it appears is different from when it is absent, as is evidenced by the difference in behavior correlated with its appearance;
> 3. The new stimulus does not itself signify additional

> stimulus features of the otherwise designated situation . . . , nor does it add to the appraisal of what is otherwise designated nor add to the prescription of how to act to what is otherwise designated;
>
> 4. The new stimulus influences the response of the person stimulated to the signs with which it appears in a particular sign combination by affecting the interpretants aroused by the other signs in the sign combination; . . . (54, p. 87).

Here we are dealing with relations of implication and not with designation, appraisal, or prescription. Formators do influence (alter) behavior, i.e., tendencies to respond. The signification of formators can be determined but they denote if and only if their antecedents denote. (Professor Morris utilizes additional break-downs of the four modes, but we have not employed them.)

To put it baldly--only ONE question is asked in evaluating a sentence or phrase: What sort of WORDS are these? For example: "There is your "Flyforlet" convertible coupe" (designative mode). "It is a powerful and beautiful car" (appraisive mode). "If you want it to function properly, you should change the oil in the crankcase every 2,000 miles" (prescriptive mode). "But--there are some owners who change the oil and some who never change it" (formative mode). In a lengthy and/or compound sentence, the mode of signifying may not be readily apparent. In such cases (and all others for that matter), the predicate (NOT JUST THE VERB) most frequently determines the mode of signifying. In dealing with connected discourse instead of separate sentences, it should be said that "in general prescription rests on appraisal, and appraisal on designation" (56). One should not expect to find discourse that employs only one of the modes; to determine the primary mode of signifying for a unit of discourse, one would have to ascertain which mode of signifying is

most frequent (number of sentences, etc.), which is most prominent (by reason of placement or emphasis within the given unit), and which is implied. Let's develop these quasi-measures more fully. For a given unit of written discourse (the same guides would serve for speech)—a paragraph for example—one could (1) examine the sentence predicates, determine the mode of signifying for each sentence, and choose the one with the greatest frequency; (2) pick out the mode of signifying which was present in the first sentence, the next to the last sentence, or the last sentence, and accept the one with the most prominence; (3) determine the primary mode of signifying on the basis of implication (e.g., questions are most frequently prescriptive or appraisive—"Don't we all KNOW that excess speed is the most frequent cause of all accidents?" and "Isn't she a pretty baby?"—the first question "means" we must slow down, and all of us know the answer to the second question); and (4) detect that mode of signifying which by reasons of rhetoric, punctuation, etc., received the most emphasis. Charles Morris suggests the following order of preference: "I (asked) myself in each instance how high in the ladder did the example go, i.e., was prescription present and prominent; if so the label 'prescriptive' was used; if not, was appraisal present and prominent; . . . ; etc." (56). He is suggesting that a reader should first alert himself for the observation of prescriptors, next appraisors, then designators, and finally (by implication) formators. The scheme which we have just outlined is meant to be procedurally suggestive—NOTHING MORE!

We have seen how something may be signified in at least four ways.

Each of these modes of signifying may stand independently in a given unit of discourse or may show various interrelationships. For example, to state that something is "good," without specifying "which" something, would give an organism no behavioral orientation. One distinction between this semantic system and others lies in the fact that different modes of signifying are not only recognized but no one of them is given preferential status. Burke stresses the prescriptive nature of *all* signs, and--although the *terms* are not used by him--he would reduce all signs to the prescriptive mode of signifying. Also--since all signs (potentially) *can* be reduced to the designative mode--many extreme positivists would regard the designative mode of signifying as the only necessary one. We have seen how some students classify signs as referential and emotive, cognitive and noncognitive, etc. The concept of mode of signifying will describe only *one* feature of such dichotomies. The additional concept of sign-usage is required to round out the picture of signs in the Charles Morris scheme.

The Uses of Signs

The modes of signifying were discussed from the standpoint of the interpreter. (From the Glossary in Appendix A, we learn that an interpreter is an organism for which something is a sign. Thus an interpreter *may* be the producer of the sign as well as the respondent.)

> The difference between designators, appraisors, prescriptors, and formators was put in terms of differences in the behavior they prepared. To say that a sign is in one or the other modes of signifying is to say that the sign operates for some interpreter in one or another of the ways described, or *generally* so

operates for some set of interpreters.

. . . . . .

Attention must now be turned to one phase of the production of signs: to the question of the purpose for which an organism produces the signs which it or other organisms interpret. We now look at signs not from the point of view of their interpretation but in terms of their relation to the purposive behavior in which they are produced and which they serve.

A sign S will be said to be used with respect to purpose y of an organism x if y is some goal of x and if x produces a sign which serves as means to the attainment of y. If a person is seeking money and writes a short story to achieve that end, the sign complex which is the short story is "used" for the purpose of getting money. The usage may be voluntary or involuntary, however these terms are defined (54, pp. 92-93).

(The economic use, practically important though it be, is not one of the four uses of signs, but this does not mean that an interpreter should ignore it.)

There are four primary sign usages:

Signs . . . may be used to inform the organism about something, to aid it in its preferential selection of objects, to incite response-sequences of some behavior family, and to organize sign-produced behavior (interpretants) into a determinate whole. These usages may be called in order the informative, the valuative, the incitive, and the systemic uses of signs. These are the most general sign usages; other usages are subdivisions and specializations of these four. They are the purposes for which an individual produces signs as means-objects in the guidance of his own behavior or in the guidance of the behavior of others. They may be employed with respect to things other than signs or to signs themselves (54, p. 96).

The Informative Use

"Insofar as signs are used to cause an interpreter to act as if something has certain characteristics they are used informatively" (54, p. 98). Any of the modes of signifying may be used in an informative way; but the designative mode is always apparent, and--conversely--designators always inform although they may be used for other purposes. (It

should be remarked that a naive conception of a designator, e.g., concrete noun, is incomplete; if a sign specifies stimulus characteristics of an object, it is a designator. Thus signs like "neutron" or "heavier" may be designators.) The information given may be of various sorts and not necessarily concerned with the physical environment.

The Valuative Use

All of the modes lend themselves to the valuative use, but the appraisive mode is best suited to the task of producing preferential behavior on the part of an interpreter. The interpreter may or may not be the producer of the signs. (In our experiments the interpreter is never the producer of the signs; the subjects were asked to judge written paragraphs, one paragraph for each of six authors.) IN THE WRITER'S INTERPRETATION the preferential behavior may be indirectly activated; i.e., object A may be negatively evaluated so that object B will be preferred. The salesman's spiel, when pointing out the bad features of a competitor's product, would exemplify this, as would the trick of negative selling; e.g., Salesman: "I feel that this item is expensive, and. . . ."

The Incitive Use

Prescriptors are best suited to the purpose of eliciting specific responses, but signs in all the modes of signifying are used also. Signs are incitively used to cause the interpreter to act towards something in specified ways, e.g., "Take the right fork of the road." Many degrees of subtlety are detectable in the prescriptive use of signs; i.e., the actual prescription may rest on prior appraisors and the appraisors on

LIBRARY
UNIVERSITY OF TENNESSEE
KNOXVILLE

designators. No matter how the several modes and uses are intermingled in a political speech, the purpose of the address is the prescription, "Vote for me!"

The Systemic Use

"In the systemic use of signs the aim is simply to organize sign-produced behavior, that is, to organize the interpretants of other signs. This may be done with respect to all kinds and combinations of signs, and by the use of signs in the various modes of signifying" (54, p. 104). Formators require signs in the other modes of signifying, and the systemic use contributes to the other three uses. Signs used systemically do NOT motivate behavior but rather prepare the interpreter to receive the message which other signs carry.

It is difficult to conceive of the systemic use of signs; this is the realm of the analogy, slanted teaching, propaganda, and normative logic ("reasoning"). Designators, appraisors, prescriptors, and formators are used in a contributory fashion to more firmly establish the designation, appraisal, or prescription. A lecture on the importance of "learning-by-doing," so that a proposed change in curriculum will be more readily accepted, would be one example of appraisors used systemically.

As a guide to determine the use of a given language unit, one may follow the plan we outlined for modes of signifying; namely, frequency, prominence, implication, and emphasis may be estimated, and the corresponding order of preference, i.e., incitive, valuative, informative, and systemic, may be followed (see page 30). In determining use, one can perhaps do no better than admit that the mechanism of projection will be

at work and guard against it. However, it would seem that a fifth factor, innuendo, should be added to the list above. We are all familiar with the pseudo-medical lingo of the carnival medicine man. It would be a grievous error to assume that he purposes to give us information; no matter how many designators he employs, he is prescribing his cure-all--if not to consume at least to buy! It will be recalled that the use of signs pertains to the purpose for which they were produced, while the modes of signifying are gauged from the standpoint of the interpreter; the signs stand by themselves, and thus there is no place for innuendo in making judgments about the modes of signifying. If innuendo is not obviously a factor, the sign-use is most apt to be organizational, i.e., systemic; this is a difficult distinction to make.

The Types of Discourse

A type of discourse has already been defined as "a specialization of language for the accomplishment of specific purposes." We are all familiar with certain rough-and-ready classifications of discourse. For example, publishers may list their offerings under "religious," "scientific," "literary criticism," and many other headings. If the reader examines the matrix on page 27, he can note the illustrative examples of the sixteen major types of discourse; for example, "scientific discourse" is a sign complex which favors the designative mode of signifying and the informative use. It will be obvious that much writing that passes as scientific will not fit these criteria. Other types of discourse may employ designators for informative purposes; thus we should remember that

Professor Morris is trying neither to define scientific discourse (in any limitational sense) nor is he saying that all designative-informative discourse is scientific; instead he is illustrating.

> It must be realised that the illustrations are tentative and indicate merely the problem and a direction of solution; we are not proposing a definition of, say, 'religion' or 'religious discourse,' but illustrating a type of discourse--in this case prescriptive-incitive discourse--by religious literature (54, p. 126).

We shall not discuss the various types of discourse in any detail. It is readily apparent that once the primary mode of signifying and the most prominent use have been determined, one only needs to consult the matrix to find an example of this type of discourse. The thought that the mode-use classification comes through most clearly when one tries to employ it has led the writer to include the suggestive scheme of procedure (pages 29-30, 34-35) should the reader be inclined to try evaluating written material.

Not the least of the merits of a mode-use classification is the fact that no value judgments are implied; one type of discourse is as worthwhile as another, even though each is not equally suitable for a given linguistic production, and no one type of discourse is adequate for all purposes. No matter how great our semantic sophistication, there is a certain sense in which the signification of a word or words will remain private or subjective; doubtless this is as it should be, and this does not mean that language cannot be categorized with accuracy in mode-use terms.

> The difficulties arise in the fact that interpersonality of language signs is seldom completely achieved, and that even where a common core of signification is obtained, the signs may have

> to different individuals of the community different additional significations. The term 'good' is at times an appraisor and at times a designator, so that a person intending to communicate an appraisive signification may by the use of this sign actually produce in another individual a designative sign-process, and thus fail in communication. Or a term which has a common designative core may differ appraisively to two individuals, so that the total communication intended by a person may fail even though a partial communication is effected (54, p. 120).

However, these pitfalls can be minimized if not completely eliminated, especially for scientific and logico-mathematical discourse. Physical scientists and mathematicians have already achieved a fairly large body of reliable signs. But the mode-use classification is not designed to limit language behavior.

> Since scientific language is pre-eminently designative in mode and informative in use, the tendency is strong to place reliance upon scientific discourse and to avoid, or even derogate, the other types of discourse. In so far as such "scientism" serves the purpose of obtaining more adequate appraisers, prescripters, and formators it is inevitable and salutary; in so far as it discourages the use of non-designative signs it is theoretically questionable and culturally dangerous. For organic processes necessitate giving preferential statuses to some objects rather than others, the selection of some courses of action rather than others, and the choice of certain linguistic structures among various possible ones. And since all phases of behavior reflect themselves in sign-behavior and need signs for their effective operation, behavior is stultified when it attempts to dispense with signs other than designators (and hence with types of discourse other than scientific) (54, pp. 116-117).

No matter what the subject matter, be it literary criticism or legal culpability, the possibility of creating a body of reliable signs exists; a rigorous metalanguage with a minimum of primitive signs and behavioral definitions of these and derivative signs can be used for designative-informative (scientific) discourse. The actual development of the metalanguage would be an example of formative discourse. We have

seen that formators may serve any of the four uses though being best adapted to the systemic use. It may be worthwhile to consider a bit further the differences between formative discourse (in general) and lexicative discourse. Formative signs may be defined in terms of their signification or not. (Charles Morris favors the first method.) What IS the signification of formators? For the writer the most concise, practicable (albeit difficult) way to analyse a given unit of discourse is to ask this question: Are the signs we are considering "about" language or "about" the world? If they deal with language, they are formative in mode of signifying. Thus Euclidean geometry is formative discourse, discourse about language. How much trouble could be avoided if we realised this! The fact that you can survey a building lot very adequately by the use of Euclidean geometry does NOT mean that the geometry says anything about the world! In Clark Hull's theory of behavior, his postulates are rules of language and terminology, and no amount of experimental testing, or verification, of theorems (derived from the postulates) can make the postulates themselves lexicative. Since many of these empirical tests DO square with the world, we are prone to consider the postulational formators as designators, at least until we find instances where the "fit" is not too good. The development of non-Euclidean geometries demonstrates not the incorrectness of Euclid's deductions but rather the fact that he was not writing about the world. Thus, the writer offers another rule-of-thumb; namely, if a unit of discourse is concerned with language, it is in the formative mode of signifying, and we should <u>first</u> determine if the use is systemic. It should be

clearly understood that we are looking for a simple, usable system for analyzing types of discourse, and that the writer's use of Charles Morris' semiotic does not do full credit to his system and often oversimplifies and broadens it. Professor Morris does not agree completely with this interpretation of formative discourse. This point will be discussed more fully in Chapter VI. However, in anticipation of the experiments to be discussed in Chapter VI, it should be stated that at various times prior to these experiments, the writer considered (1) omitting the formative mode and systemic use, or (2) employing instead a "neutral" mode of signifying and a "general purpose" use. This last point will be taken up again in Chapter VIII.

CHAPTER III

RATIONALE AND STATEMENT OF THE PROBLEM

Why Another Language Study

Confucius has equated one picture with a thousand words. It may be that one parable can equal one picture and thus save us a few words:

"The Parable of Objectivity"

> It is in the paramount interest of the State and the nation to prevent (the common) people from falling into the hands of evil, ignorant or even malevolent educators. "How many readers have enough personal knowledge to distinguish fact from fiction, ignorance from knowledge, interest from impartiality" without their paper's help? The State, therefore, has the duty to supervise their education. Even the most competent newspaper editor is often so convinced of the need to be objective that when he spots a "downright misstatement of facts" in a speech, he never follows it with a bracketed insert to the effect that "This simply is not so." (Thus the nation) has to watch especially the press for its influence is by far the strongest and the most penetrating on (the) people, as it is applied not temporarily but permanently. If the Honorable John P. Hoosis is an important person, "you (may still) see him quoted at length in newspapers on almost any subject, with no indication that he knew nothing at all about it . . . (for) to do that would be editorializing, interpreting the news, failing in objectivity." In the persistent and eternal repetition of this instruction lies (the newspaper's) entire unheard-of importance. "You can do it (editorialize) to Stalin; you could do it to Hitler in his day; but tradition forbids doing it to one of our fellow citizens." Therefore, if in any place at all, the State must not forget that just in here all means must serve an end. "Failure to make such a correction may salve a man's conscience about his loyalty to objectivity, but how about his loyalty to the reader?" (The government) must not be persuaded to fail in its duty and to put before the nation the food that it needs and that is good for it. "The good newspaper, the good news broadcaster, must walk a tightrope between two great gulfs--on one side the false objectivity that takes everything at face value and lets the public be imposed on by a charlatan with the most brazen front, on the other the 'interpretive' reporting which fails to draw the line between objective and subjective, between responsible and well-established fact and

> what the reporter or editor wishes were the fact." (Thus the State) must assure itself with ruthless determination of this means for educating the people and to put (it) into the service of the State and the Nation. "No wonder that too many fall back on the incontrovertible objective fact that the Honorable John P. Hoosis said, colon, quote--and never mind whether he was lying or not."

We hope the reader has tried to determine the mode-use classification of the parable. Many questions could be asked about our parable; the methodologies mentioned in Chapter I could provide answers to most of these queries. If we ask, "What is the meaning of the parable?", many answers would be forthcoming: Some would tell us that it is "emotive" language and would deduce the implicit evaluations; others might say that the signs of the parable are "referential"; the general semanticists would apply a level-of-abstraction analysis; logicians would manipulate the propositions contained in the message; and so on. Some over-all agreement MIGHT be achieved, but the judgmental frames of reference and the amount of agreement could not be adequately specified by any of the analytic schemes, at least not if the signification of the words is being investigated; moreover, experimental data would not be offered. The semiotic of Charles Morris was designed to answer these questions, AND it seemed to lend itself to experimental test. That such a test should be made will be obvious if we tell you that Adolf Hitler, Elmer Davis, and Time magazine were the coauthors of the parable (7, p. 128; 100, p. 51). Sentences 1, 3, 5, 7, 9, 11, and 13 are Hitler's, with the exception of the words in parentheses which are the present writer's. Sentences 2, 4, 6, 8, 10, 12, and 14 are from Time; the present writer added "editorialize"; the statements in quotations are Davis'.

You will find that the sentences make sense if read separately, i.e., as TWO units of discourse.

The writer would like to give you a summary of his introspections about the parable; they may help establish the background or "feel" of the experiments, which required the subjects to evaluate paragraphs in mode-use terms. This has not been done before on an experimental basis. As a matter of fact, it is not easy to judge whether or not an experimental study of meaning has ever been conducted. Reading the parable as a unit, it seemed to be prescribing that we watch the "press" and make certain that the people get an interpretation and not straight reporting because they are not too "bright" and/or not in a position to ascertain the facts. The admixture of designaters and appraisors makes the prescription less turgid and thus perhaps more palatable (prescriptive mode of signifying). The use is incitive and not systemic, because the newspapers are discussed, for the most part directly. "We are going to do this to the press, and we want you to approve it and aid us" (incitive use). Looking at the odd-numbered sentences (Hitler's), they were judged to be--in terms of frequency, prominence, implication, emphasis, and order-of-preference--overwhelmingly prescriptive in mode of signifying; for the same reasons the use is patently incitive. This might be considered further evidence for labeling Fascism a religion. The writer is certain that each of us has participated in at least one discussion in which the question dealt with the make-up of Hitlerism; is it a political philosophy, a religion, or both? Aren't we inquiring as to the type of discourse favored by its chief spokesmen? If our questions were based on this

frame of reference, we might be in a better position to answer the question. Consciously or not we are actually judging types of discourse quite often; e.g., "Dr. X is a poet though he thinks he's a scientist."

When the writer turned to the even-numbered sentences of Time and Elmer Davis, he had more trouble. In evaluating the mode of signifying, frequency is not decisive; the first and next to the last and the longest sentences are prescriptive (prominence); prescription is implied by the questions; emphasis is probably on the appraisive mode of signifying; prescription is first in the order-of-preference; therefore, it was concluded that the passage is predominantly in the prescriptive mode of signifying. What of the usage? The article is systemic, insitive, valuative, and informative. However, the systemic and informative uses are not prominent. Order-of-preference, etc., slightly favor the insitive use, but innuendo points to valuative. In weighing these two uses, the writer finally voted for valuative; it seemed to him that "Davis-Time" was a bit fearful about having its prescriptions established as a general policy. What is really intended is that we should approve of the policy of news interpretation (editorializing and not mere reporting) if and only if qualified, literate, and well-meaning writers are doing the interpreting. Davis feels that he has the qualifications, but beware of the "objectivists," though they should be tolerated in a democracy. Thus, the even-numbered sentences yield a type of discourse exemplified as political. The reader is certainly free to make his own evaluations, and the writer doubts that in all instances they will agree with his; I am confident that no one will consider the entire parable or the separate

"paragraphs" to be logico-mathematical, cosmological, or scientific. We are not commenting on Time magazine or Elmer Davis but only trying to classify this particular article; one does not generalize from a sample of this size. I do think that Hitler's contribution to the parable is unequivocal. The fact that the remainder of the piece is not so obvious only points up the need for a method which will categorize it; then if we agree that we favor this type of reporting, we will know on what our agreement is based. However, it is suggested that none of the methods discussed in Chapter I lend themselves to this sort of analysis. The writer is convinced that some such analysis is important; and that whether or not you agree with his evaluations, we could have a profitable and unemotional discussion of them if we followed the types-of-discourse schema. Moreover, the writer is convinced that a number of readers would interpret the parable as "true" and thus by implication scientific discourse, since truth and science are--for many people--synonymous in 1952. Would the mode-use classification help us achieve a certain measure of freedom from speech? And note that we have not said that the parable should never have been written. Perhaps we should accept its prescriptions, but we should receive them as selective--not hypnotized--organisms; i.e., we should give symbol and not signal reactions, and this is dependent on knowledge of the stimulus--type of discourse.

Is A Science of Signs Necessary

Even if the possibility exists that the significance of written material can be fairly objectively estimated by the mode-use system, are

such analyses important to psychologists--to the language of psychology and the psychology of language? Psychologists deal with behavior, not words. Numerous authors, in one way or another, tell us that the language of a science is the better part of that science. The "Unity of Science" group is chiefly concerned with the development of a language, or type of discourse, that will be common to all scientific disciplines. The writer feels that at least three types of scientific discourse could be tabulated, but most of us would agree that a science should contain, indeed should be founded upon, a core of reliable signs, signs that are for the most part designators and which are used for no other purpose than the giving of information. Logical empiricists, operationists, and theorists are striving constantly for the achievement of a body of reliable signs; they are at some pains to convince investigators of psychosocial phenomena of the need for referential language and to instruct them as to how such a language can be developed. The semiotic of Charles Morris specifies one set of conditions under which language may be considered as scientific and/or logico-mathematical. Once we can agree as to what is signified by a given sign-complex, certain methods (mentioned in Chapter I) are available to determine whether or not these signs denote, the most obvious being the experimental test. We do not suggest that the methods previously mentioned have not borne fruit; the logico-mathematical discourse in Hull's Principles of Behavior and Skinner's operational treatment of The Behavior of Organisms--to mention only two investigators--will prove that they have. However, it would seem that logical empiricism has not permeated the work of any large number

of social scientists and that many students have only a nodding acquaintance with these concepts. Koch (42), in reviewing Helson's collection of articles, Theoretical Foundations of Psychology, writes:

> . . . the one uniformity common to most of the seventeen contributors seems to be a predilection for some kind of theoretical framework of the sort variously called "field," or "neo-gestalt," or "dynamic," or "holistic" (p. 349).

> Terms like "molar," "molecular," "field," "intervening variable," "operational definition," "theory," etc., become chameleons which assume new semantic and emotive shadings in each chapter (p. 350).

> A third feature of the book, not without symptomatic significance for psychology in general, has to do with a complex of factors epitomized by the nature and level of the writing, by organization, and by style. Those who occasionally face the unhappy task of helping graduate students approximate to clear, scientific prose have reason to be worried by the prevalence of models similar to the present volume. The net impression created by the majority of chapters is that of turgid, imprecise, often vacuous writing. The fact that the personal flavor of the prose varies from the pseudo-Jamesian "Harvard-psychologese" ("memories seem . . . linked . . . like kaleidoscopic sausages in a giant string" (p. 395); "I have . . . stretched the immensely complex past on a Procrustean bed" (p. 405)) of the chapter on learning, through the shabby rhetoric of the chapter on social psychology, to the inflated verbosity of the chapter on abnormal psychology, is secondary. More serious is the frequent failure to meet minimal standards of scientific or scholarly writing. At its worst, this may involve the failure to communicate anything; at its best the failure to communicate anything worth while (p. 358).

Whether or not one agrees with Professor Koch's evaluation, he obviously agrees that serious linguistic problems exist in psychology--exist despite any amount of lip-service to operationism. Would the mode-use classification help make such book reviews obsolete? The answer to this question would require a long-range research program. This thesis is only a beginning.

If scientists are not always clear as to the signification of

their own signs, what of the sign-behavior of that celebrated citizen, John Q. Public? Once, while delivering what the writer considered to be a very mild lecture on child psychology to a night-school class, he noticed the fidgeting of one girl student. Her face was flushed, and each of his words seemed to redden it the more. Finally she could stand it no longer and blurted out, "My pastor doesn't say THAT!" The writer hastened to assure her, with just a bit of success, that the pastor and the teacher of psychology were not in the same business--that they did not speak the same type of discourse, and that this was perhaps as it should be. The possibility of any large number of us becoming semioticians, philosophers of science, scientists, mathematical logicians, or general semanticists is slim indeed, but we could be taught to "read."

This writer is at times disturbed by the suspicion that many of us have a vested interest in language, and that often we neither want to disseminate semantic information nor even to communicate. While a student of Chinese at Cornell University, William H. Shererts and the present writer developed a system for Romanizing the Chinese language. Naively, we thought that the alphabetizing of Chinese would be welcomed. By the use of combinations of two--from a list of thirteen--unvoiced consonants at the end of each word, we were able to distinguish what character (of several similarly pronounced) was intended, and to indicate the tone without the use of any diacritical mark. We felt that ours was neither the best nor the only system, but that we had answered the main objections which had been offered to other systems. The writer lost interest in the system because no matter how many arguments were

countered, those who had mastered written Chinese did not want any simplification. It is interesting to note that the Chinese Communists, prior to coming into power, advocated and used the Harvard University system of Romanization, but they very suddenly dropped it. Why they did so is left to the reader to decide. But it should be stated that it is difficult NOT to detect nonsense when it is written in an uninflected, limited-vocabulary, and referential language like Chinese.

Even if there are breaks in the scientific chain of communication, doesn't this make for progress? Didn't the intellectuals achieve their semantic capital the hard way and thus merit the profit? Perhaps Man is the "ambiguous animal." Surely any need for another language study must spring from the needs of the investigator. Burke states:

> Indeed, on the Symbolic plane, as a likely hunch that may or may not be verified on closer examination, I think one is always justified in looking for tender apprehensions behind the apparent toughnesses of materialist debunking, as Bentham in his childhood had an abnormally intense fear of ghosts, and in adult life developed a critique of language particularly zealous in disclosing kinds of words that named merely fabulous or fictitious entities having but the semblance of reality; or as he aimed to dispel the moral pretense in idealistic words, by treating them in terms of the material interests they cloak, thus translating spirit back into body, which as regards the childhood pattern equalled the transformation of ghosts back into corpses (11, p. 162).

Burke's analysis of Bentham's motivation may be accurate enough; but of equal clinical significance are those who would <u>prevent</u> any spirit-to-body translation--those who follow Humpty Dumpty and make words mean what they want them to mean--at the same time anxiously resisting any attempt at the analysis of "meaning." One should not expect gratitude if he tells a child there is no Santa Claus.

Why Charles Morris

If another language study *is* needed, why select the semiotic of Charles Morris for experimental investigation? It seemed that the core of his system might lend itself to straightforward empirical testing. No semantic background would be demanded of the subjects, and a summary presentation of the system might indicate the teachability of it and thus encourage any attempt to correct the summarization and/or expand it. Though it will be obvious that authors do not use his method in preparing their scripts, nevertheless, it appeared likely that some agreement among raters could be reached by employment of the mode-use classification. Even if agreement were low, the method might still function as a semantic sieve; i.e., we could at least tell what a given unit of discourse was NOT. Moreover, the system seemed to suit the average reader, and if so, it should not be the exclusive property of "experts." In the state of our present knowledge, it seemed appropriate to test a system that did not value one type of discourse more highly than another. The rudiments of the system would seem capable of being grasped rather readily, and--if this should be the case--the student would be encouraged to pursue this or similar analyses of signs. Moreover, if first tests warranted it, the system would seem to lend itself to numerous additional tests, the implications and ramifications of which would be of interest to psychologists no matter what their area of interest or theoretical persuasion. Finally, Morris was selected because his metalanguage is framed behaviorally, in psychological terms, and thus lends itself to easier communication between psychologists themselves and between Professor

Morris and psychologists. There are other reasons, but they will be discussed in another section.

It may have occurred to you to ask why the writer has not followed this system more carefully in the writing of this thesis; you are not alone in this. The writer had thought to make the first chapter more clearly appraisive-systemic and Chapter II more obviously designative-valuative, but--as yet--we lack any experimental data that would indicate the advisability of employing any such procedure. Possibly this chapter is already too prescriptive-insitive to suit many of you. In Chapter IV we promise to try our best to write discourse that is designative in mode of signifying and informative in usage and thus scientific.

CHAPTER IV

METHOD

The experiments to be described were designed to test whether or not a simplified and somewhat modified version of Charles Morris' mode-use classification of types of discourse could be taught to subjects and employed reliably by them in one testing session. (Hereafter Professor Charles Morris will be referred to as "C. M.," while the present writer is "E," and "S" or "Ss" will be used for the subject or subjects who participated in the experiments.) The total time for Experiments I, II, III, and V varied from 45 to 75 minutes. A preliminary tryout conducted in the summer of 1951 served to point out some of the difficulties but gave promise that the plan was feasible. In this tryout, as in the actual experiments, Ss were asked to read a paragraph, determine the mode of signifying and the use, study the matrix, and write down the name of the type of discourse they had selected. (Originally E had thought to present the Morris system to selected groups of graduate students after the manner of a seminar which would seem the best way to introduce difficult and abstract material in a short space of time. Dr. E. E. Cureton of the University of Tennessee suggested the possibility of group training and testing and a wider range of subjects. The writer is indebted to him for outlining this approach, which provided a more rigorous and objective test of the method.)

Because of time and space restrictions, C. M. could not be consulted at any length re the conduct of the experiments. Thus the

preparation of the materials and the development of the scoring system were the responsibility of E. An original pool of 97 paragraphs, varying in length from approximately 108 to 135 words, was assembled from journals, magazines, and books; all paragraphs selected were of psychological import though not necessarily written by psychologists. Since we were primarily interested in the language of psychology and the social sciences and since this area does not offer the most precise or obvious types of discourse, it was felt that this biased sample would serve as an adequate, even severe, test of the system. The difficulties experienced in classifying fiction, news-reporting, and the like are very probably fewer than those to be encountered in evaluating the paragraphs of social scientists. Flesch (23) feels that standard academic writing is less readable and comprehensible than popular writing. The original total of 97 paragraphs was reduced to 21 by use of a crude paired-comparison rating and an over-all judgment; certain paragraphs seemed to be poorer examples of types of discourse already represented in the sample, etc. E rated these paragraphs from 1 to 4 re appropriateness and submitted them to a judge who used the same scoring scheme but operated with no established criteria beyond the knowledge of the use to which the paragraphs were to be put. From these two judgments 12 paragraphs were selected, and finally 6 were chosen to be used in the first three experiments. The test items were not scaled according to difficulty as estimated by any of the systems now in use. They are, however, of approximately the same length; they were judged to vary from fairly difficult to very difficult re readability for an undergraduate

population. Very high-level items were selected not only to avoid the possibility of perfect scores but also to see if some of the Ss could classify the discourse even if they did not understand it. We know that the reverse is true, namely, that familiarity with every word in a paragraph does not guarantee that it will be understood. The use of a paragraph as a unit was dictated by (1) the amount of time available and (2) the fact that it could be considered a unit of discourse. (Words can rarely be so considered, and a sentence--even though a possibility--would have even more limited generality and be a more limited sample than a paragraph.) For obvious reasons the authors' names were not added to the paragraphs. Neither E nor the Ss were judging WRITERS. It should be repeated here that one of the values of the C. M. system lies in the fact that no value judgments are intended or implied, although it is assumed that, for example, a scientist will choose to write DESIGNATIVELY and INFORMATIVELY if he intends to write scientific discourse. (He may in fact be writing by intention critical or logico-mathematical discourse, etc.) Appendix C contains the 12 paragraphs, together with the authors' names, the changes made by E, the scoring, and the sources. To the right of each paragraph will be found additional choices as to scoring; i.e., "2" would be E's second choice in labeling the paragraph, "3" E's third choice, and so on. The first 6 paragraphs were used in Experiments I, II, and III. The 2 paragraphs marked with asterisks were used in Experiment V.

Next, three experimental treatments were devised. Readability was not determined beyond limiting each treatment to approximately

2,000 words and 100 sentences. Flesch (22) has shown that the average sentence length in words is one important index of readability. On the basis of average sentence length (20 words per sentence), the reading-difficulty level of the paragraphs is not high. However, specialized vocabulary makes some of the paragraphs difficult for undergraduates to comprehend. Four judges were asked to read the treatments as though they were going to be asked certain questions about the material at the end of their reading. Naturally, reading times varied, but each judge read each treatment at approximately the same rate. The treatments were about 5 1/2 pages in length when the lines are separated by 1 1/2 spaces mimeographed on 8 1/2" by 11" paper.

The first or "Morris Treatment," for the main experimental groups, consisted of a summary of C. M.'s treatment of modes of signifying and uses of signs AS INTERPRETED BY E and was taken from his book, Signs, Language and Behavior (64). This treatment appears in Appendix B, and the reading of it will give the reader a better understanding than any explanatory words.

Second was the "Feigl Treatment" (see Appendix B), which consisted of a simplification, using the author's own words plus minimal and necessary changes by E, of Herbert Feigl's "Operationism and Scientific Method," which appeared in the Psychological Review in September of 1945 (18, pp. 250-258). E does not claim that his summary has not done violence to the clarity and accuracy of his presentation; it has not, however, been done intentionally. It seemed that it was a good article for OUR purposes in that it should help Ss to think more rigorously about

scientific and logico-mathematical discourse. It should go without saying that Professor Feigl did not intend that his article should serve the purpose to which E put it.

The third or "Osgood Treatment" was an abstract of the first paragraph and the final section, titled "The Semantic Differential," of "The Nature and Measurement of Meaning" by Charles E. Osgood, which was published in the May, 1952, edition of the Psychological Bulletin (62, pp. 197, 222-232). Obviously no comment on this article is intended when we state that even though the writing is concerned with meaning and its measurement, it was felt that it would be of little use to the Ss in solving the problems which were set for them. The summary may be found in Appendix B. (The names of the authors who wrote the articles from which the three treatments were abstracted were not attached to the summaries.)

The basic design for all the experiments was a modification of the classical transfer-of-training type and was suggested by Dr. William O. Jenkins of the University of Tennessee. Solomon (79) writes that Thorndike and Woodworth were the first (1901) to employ the control-group design in studying transfer of training. The design of E's experiments, particularly Experiment V where the effect of the pretest is being measured, draws on the work of Hovland, Lumsdaine, and Sheffield which was done in the early '40's. In particular one should cite the "before--after" design which they used to measure the effects of the film, "The Battle of Britain"; measurements were taken on the experimental (film) and control groups both before and after the introduction of the

experimental variable (57, p. 28). Solomon (79) has offered several extensions of control-group design in transfer-of-training experiments and has suggested the importance of measuring the effect of pretests.

Strictly speaking, control groups were not used. Both the Feigl and the Osgood treatments have to do with language; the pure control case (a "nothing" condition or treatment) might well have produced greater differences between the experimental (Morris) and the control (Feigl and Osgood) groups. There is also the problem of degrees of "nothingness" in that it is difficult to determine the part that boredom, proactive inhibition, etc., may be playing when the Ss of control groups are asked, for example, simply to sit or to read a story from the New Yorker. Many studies indicate that improvement takes place in the pretest-posttest design no matter what the intervening condition, and the fact that both the Feigl and Osgood articles deal with the meaning of words should--if anything--have stacked the cards AGAINST the "Morris Treatment." The Osgood I groups (which will be discussed later), however, were--a priori--considered as being experimentally "neutral," since it was felt that this treatment would be of no assistance in solving the test problems. The design for Experiments I and II is given in Figure 2, page 57. Improvement or change scores could be obtained for each S; this score is equal to the posttest score minus the score on the pretest. In addition, the amount of time in minutes required to complete the posttest was recorded by each S.

The two Osgood groups were not used in Experiment III, but the experimental design remained the same other than this.

	Experimental Groups			
	Morris	Feigl	Osgood I	Osgood II
Pretest	Yes	Yes	Yes	Yes
Training	Yes (M)	Yes (F)	Yes (O)	Yes (O)
Posttest	Yes	Yes	Yes	Yes

Figure 2

Diagram for the Experimental Design of Experiments I and II

The design outlined in Figure 2 has certain advantages: Since the paragraphs were the same for both pretest and posttest, the stimulus situation while taking the pretest was similar to that of the posttest, and thus we don't have to be concerned with comparable forms of the test; a test of the "learning-by-doing" hypothesis (the effect of the pretest) could be made (see Experiment V); and in all experiments Ss were assigned to the groups at random, but we do have an initial measure (pretest) and time scores which make the application of covariance techniques feasible and also permit each S to act as his own control, thus offering a potential cutback of the effects of individual differences.

A pretest was given to all Ss in Experiments I, II, and III. They were asked to determine the types of discourse in each of 6 paragraphs. (The first 6 paragraphs in Appendix B were used in these tests.) The pretest instruction sheets for all experiments and treatments are given in Appendix B. E read the instructions aloud while the Ss read them from the instruction sheet.

Certain changes were made in the original C. M. mode-use matrix (see Figure 1, page 27). E's alterations are in Appendix B. The revision used in Experiment I contained only the following changes: "valuative" was changed to "evaluative," "fictive" to "fictional," and "logico-mathematical" to "logical-mathematical." It was felt that these changes were minor and that the new terms might be more familiar to undergraduates.

Upon completion of the pretest, all papers were collected. Verbal instructions were then given for Part II of the test (treatment

and posttest). Ss were told (1) to read the material about to be given them and to read it in the same order in which it was presented to them; (2) to follow all the instructions as given in the written material; (3) to note the clock which was on the desk in front of them (everyone using the same clock) and to mark the time when they began and the time they finished the final task (this point about timekeeping was carefully illustrated); and (4) to raise their hands at any time if they had any questions. (Monitors were available for answering questions; however, very few were asked.) Prior to the meeting with the Ss, the three treatments had been intermixed in a serial fashion so that the monitors could pass them out and E would not know which S had which treatment--in an attempt to insure that each group would be a random sample from the class.

After the instructions for this part of the test had been presented and discussed, the monitors passed out two booklets and a single page. The first booklet contained ONLY the treatment (Morris, Feigl, or Osgood), and no titles, instructions, or names of authors appeared in this booklet. The second booklet contained a page of instructions and--attached to the instructions--the 6 pretest paragraphs, which were to be rejudged; there were 3 posttest pages, i.e., 2 paragraphs per page. One S from a preliminary tryout reported that he felt that E was actually interested in testing recall, and so he had tried to give the same responses (answers) in the test that he had given in the pretest. So a single sheet covering this point was inserted between the treatments and the posttest in an attempt to avert such misconceptions. A copy of this

sheet will be found in Appendix B. This material was grouped in the following order: Booklet I was on top, the memory sheet was second, and Booklet II was on the bottom--so that the Ss could begin by reading the treatment. (Monitors made certain that these instructions re order of reading were carried out.)

In Experiments I, II, III, and V, three experimental and four control Ss were lost. In Experiment I one S ("Osgood Treatment") was not used because she was unable to follow the instructions--re types of discourse--in the posttest. In Experiment II six Ss were lost--three "Morris Treatment," one "Feigl Treatment," one "Osgood Treatment"--No. I, and one "Osgood Treatment"--No. II--because of failure to keep time correctly.

It will be recalled that the posttest was made up of the same 6 paragraphs that were used in the pretest. Re "Osgood Treatment" II above, those Ss who received this treatment were divided into two groups; the treatment was not changed in any way, but two sets of posttest instructions were used. One of the subgroups ("Osgood Treatment" II) received instructions that contained the matrix. It was felt that the matrix alone might be of some help to the Ss in determining the types of discourse. Please consult Appendix B for the instruction sheets and revisions.

The paragraphs were scored in the following manner: 1--if MODE was correct; 1--if USE was correct; and 4--for TYPE OF DISCOURSE. After scoring the pre- and posttest papers for Experiment I, it seemed that the Ss grasped the concept of use more readily than mode of signifying, and so a 1 (MODE), 4 (USE), and 16 (TYPE OF DISCOURSE) scoring scheme was

adopted, and the papers were rescored. Only the 1-4-16 scoring plan was used for the other four experiments. Other scoring methods are defensible. However, the important points were to give partial credits and to score type of discourse highest. In any event, no change in the scoring system will change the direction of the experimental results.

Experiment I

Twenty-eight students, 10 males and 18 females, who were members of a class in general psychology for teachers took part in this experiment. (Recall that one S was lost.) The pretest, training, posttest design previously mentioned was used. The pretest instructions, three posttest instruction sheets, and the three treatments used in Experiment I may be found in Appendix B. The complete testing session lasted 50 minutes. (The total time of testing for all experiments should be interpreted as indicating the maximum time for any S.)

Experiment II

In addition to the two changes already mentioned (the elimination of the three definitions given in the pretest instruction sheet and discontinuance of the 1-1-4 scoring), three other alterations were made for this and the remaining experiments: (1) It was found that the Ss could not ignore the titles of the types of discourse; e.g., they had a conception of "legal discourse" which would prevent their assigning this label to a given paragraph even if the mode-use classification seemed to

warrant its use. Therefore, the names of all the types of discourse save "mythical" were changed, changed to more general and abstract terms. (Obviously these changes had to be made in the pretest and posttest instruction sheets and in the "Morris Treatment." These and all other alterations may be found in the illustrative material given in Appendix B.) A list of the titles for Experiment I and the new titles is given below. (2) The illustrations of the four modes of signifying at the bottom of page 3 of the "Morris Treatment" were omitted (see Appendix B). (3) There were slight changes in the posttest instructions. Everything else was the same as Experiment I.

Types of Discourse

Experiment I Titles	Revised Titles
cosmological	universal
critical	judgmental
fictional	imaginative
grammatical	exemplary
legal	statutory
logical-mathematical	rigorous
metaphysical	suppositional
moral	ethical
mythical	mythical
poetic	idealistic
political	documental
propagandistic	doctrinal
religious	imperative
rhetorical	persuasive
scientific	testable
technological	procedural

The Ss who served in this experiment were young females enrolled in nurses' training; all were members of a class in psychology for nurses. There were 23 members of this group not counting five whose tests could not be used--four because they had included the time required to read the treatment as the "Starting Time" for the posttest, and one who did not

understand the time instructions, and there was no way to adjust for this. No systematic attempt to control for intelligence was made in any of the experiments, except by randomization. Many of the paragraphs were inappropriate (in terms of comprehensibility) for these student nurses, but E was interested in seeing if Ss could use the classification system even if "understanding" of the paragraphs was limited. Paragraph No. 6, dealing with chi square, is an example of an item that possibly no member of this group could be expected to understand. Total time for this experiment was 75 minutes.

Experiment III

There were 18 Ss in this experiment, 5 female and 13 male. Six Ss were psychology students (5 graduate and 1 advanced undergraduate); 6 Ss were advanced graduate students in physics; 3 were graduate students and 1 an undergraduate in philosophy. The 6 psychologists were divided into two groups of 3 each by Dr. E. E. Cureton, who used scores on the Miller Analogies test as a criterion for the division. The 4 students of philosophy were divided into two groups, on the basis of their class work, by Dr. Willis Moore. The physicists, all candidates for the Ph. D. and nearing completion of their work, were assigned at random to one of two groups. These divisions gave 9 Ss for the experimental group ("Morris Treatment") and 9 for the control group ("Feigl Treatment"). Because the sample was small and because the first two experiments led E to believe there were no significant differences between the two control treatments, no "Osgood Treatment" groups were used. The main purpose of this

experiment was to further test the significance of the "Morris Treatment"--in this case at a high intelligence level--in using the mode-use classification. With the exception of the omission of the Osgood groups, the experimental procedure was exactly the same as in Experiment II. (Minor additions were made to the "Morris Treatment" and the corresponding posttest instruction sheet. These changes may be noted in Appendix B.) Again intelligence was not controlled rigorously, but it seemed safe to assume that the Ss of Experiment III would have a higher mean intelligence than either those of Experiment I or II. Also the reading difficulty of the paragraphs would not be an important factor in that most of the Ss were familiar with material similar to that used in the test. Total testing time was 60 minutes.

Two additional experiments (IV and V) were run. These will be discussed in Chapters VI and VII. Experiment IV investigated the agreement between C. M. and E in scoring (labeling) the paragraphs. Experiment V was primarily directed towards estimating the effect of the pretest--in conjunction with the "Morris Treatment"--on posttest scores.

CHAPTER V

RESULTS OF EXPERIMENTS I, II, AND III

Experiment I

Before presenting the data, certain points should be mentioned. Prior to conducting the experiments, it had been decided that two statistical analyses would be employed--the first a t test of mean change-score differences (posttest score minus score on the pretest) between the experimental and the control groups; and the second a multiple regression analysis which would adjust differences between groups on mean pretest scores and the amount of time spent on the posttest. Summaries of these tests will be presented in this chapter; the raw data for the t tests are given in Appendix C, and an extended discussion plus the summary raw data for the multiple regression analyses are presented in Appendix D.

The three control groups (Osgood I and II and Feigl) were evaluated to ascertain if there were any significant differences between the groups, and there were none. The means for pretest, posttest, and time were very comparable. Presentation of control-group data and a discussion of the differences will be reserved until the analysis of the data for Experiment II is undertaken. E felt that the combination of these three groups to form one control group was justified.

It will be recalled that the original score plan was to give 1 point if a S's selection was correct as to mode of signifying, 1 point if use was correct, and 4 points for type of discourse. On a chance

basis (please consult the matrix), this would seem to be the reasonable system. The mean change score was 1.10 for the experimental (Morris) group and 0.39 for the control (Others) group. When the 1-1-4 score plan was applied to the change scores of Experiment I, a t of .624 was obtained; this gave a P value of .27 (one tail) for 26 df. (The raw data for Experiments I, II, III, and V will be found in Appendix C.) The P value for this analysis was considerably above the point of significance that we had agreed to accept.

As previously mentioned, inspection of the data led us to believe that the concept of use, though theoretically less objective than mode of signifying, was being grasped more readily by the Ss. Thus the 1 point for mode, 4 points for use, and 16 points for type of discourse score plan was applied to this experiment. The difference between change scores favored the Morris group--4.40 to 1.56. This scoring gave a t of .75 and a P value of .23 (one tail) for 26 df. Admittedly this is an after-the-fact procedure, and possible objections might be raised about the use of the 1-4-16 weights; nevertheless, this change seemed justified by the evidence at hand and by the fact that E was committed to the use of this scoring plan (no matter what the statistical test) for the remainder of the experiments. Actually the revised scoring did decrease the P value--but only slightly and not to an acceptable level--while the variance difference was increased. However, an F test of the variances was not significant. (When all the data were in, the differences between mode of signifying and use were substantial, but difficult re interpretation. The Ss scored more substantially on mode for C. M.'s key and for

use with E's key. Since the number of items was small, the differences may represent biased sampling as well. Future experiments will have to determine the best scoring plan.) A summary of the results of the two scoring plans is given in Table I.

It was thought that for a pioneering study such as this and one in which the reliability of the test proper was not known, a P value of .10 for one tail of the distribution would be accepted as significant. This figure was further justified by the fact that replications were to be made. These replications, though not exact duplicates of Experiment I, should add generality to the hypotheses being tested. That is, Experiments II and III were similar enough to Experiment I to be considered as replications and different enough to add generality to any conclusions.

Since the t based on the revised scoring did not reach significance because there were INTERGROUP variations (not fully revealed by comparison of means alone) in both pretest and time mean scores, E decided to use an analysis of covariance. The rationale, summary raw data, and a discussion of the particular analysis that was used are given in Appendix D. The writer will present only a general description of the technique and a summary of the results in this chapter; this procedure is justified by the fact that--as we shall see--the statistic proved to be incorrect for showing differences between the experimental and control treatments, but it was of great value in pointing out the nature of the predictor variables with which we were dealing and their differential effects on the two groups. Also it was felt that the reader might lose the main thread of the argument and the actual significance of the

TABLE I

RESULTS OF THE 1-1-4 AND 1-4-16 SCORING FOR EXPERIMENT I

	1-1-4 Scoring		1-4-16 Scoring	
	Morris	Others	Morris	Others
N	10	18	10	18
$\bar{X}_c$[a]	1.100	0.389	4.400	1.556
s_c^2	7.878	8.606	66.711	132.026
	$t(\bar{X}_{c_{M-O}}) = .624$		$t(\bar{X}_{c_{M-O}}) = .750$	
	df = 26		df = 26	
	P = .27 (one tail)		P = .23 (one tail)	

[a] $\bar{X}_c$ is the mean change score or mean posttest score minus pretest mean score.

"Morris Treatment" if he were unduly burdened with the actual statistic (computation, etc.) in this chapter.

The technique employed was a multiple regression analysis of covariance which led to a t test of the adjusted criterion (posttest) scores. The total and within-group multiple correlations were computed; these correlations expressed the relationship between pretest and posttest scores and between time spent on the posttest and the posttest score. That is, pretest mean and time mean were adjusted for both the experimental and the control groups and expressed in terms of a common regression line. The actual regression equations do not have to be computed in order to apply the test of significance which is made for the predicted mean criterion scores on the basis of the total and within-group multiple correlations; a t^2—in this case equivalent to F—is computed, and its square root or t is tested like any other t with degrees of freedom equal to the total number of Ss minus the number of groups and predictors, or N - 4 for our experiments. The chief factor that led E to employ this statistic was the intergroup variations in pretest scores. Regression towards the mean seemed to be affecting the change scores; i.e., those Ss who had above average scores on the pretest showed smaller change scores, while those who started below the mean obtained the larger change scores. Another contaminating factor was the differential time scores; those Ss in the Morris group—in particular—who spent more time on the posttest showed larger average change scores.

Another positive feature of multiple regression analysis stems from the fact that prediction equations, if needed, are easily obtained;

i.e., the predictors can be weighted differentially, and the adjusted or expected criterion scores can be obtained. An important corollary of this fact deals with the intragroup correlations. The three zero-order correlations can be computed for each separate group and combined to give the intragroup multiple correlation. This statistic is difficult to interpret; no exact test known to the writer is available for testing the significance of a difference between multiple correlations. However, prediction formulae can be obtained, and thus we can determine whether or not pretest score and time predict DIFFERENTIALLY for the two groups; moreover, the direction and the differences between the zero-order correlations can be statistically tested. The intragroup correlations are given on page 79, but we are reserving further discussion of them until the results of Experiments I, II, and III have been presented.

Table II gives a summary of the multiple correlations and the results of the regression analysis; more complete data will be found in Appendix D.

Thus, even though the obtained posttest mean for the experimental (Morris) group is 2.92 points larger than that for the control group, the corrected mean-difference is not in the same direction; i.e., were we to compute the beta coefficients and the b' weights and use the latter to obtain a predicted mean, this corrected mean would be smaller for the experimental group than that actually obtained and larger for the control group; this is what the negative t implies. In view of the positive t for the mean change-score differences which was previously yielded, this reversed t for the multiple regression analysis should make one suspicious

TABLE II

EXPERIMENT I: SUMMARY OF THE RAW DATA AND RESULTS OF THE MULTIPLE REGRESSION ANALYSIS[a]

	Morris	Others	
N	10	18	
$\bar{X}_0$	19.70	16.78	
s_0^2	121.79	55.71	
$\bar{X}_1$	15.30	15.22	
$\bar{X}_2$	15.90	13.94	
$\bar{X}_{e_{2.1}}$	2.92		
$r_{0.12}^2$	.084		
$r_{0.12}$	.290		(within-group multiple correlation)
$R_{0.12}^2$	.140		
$R_{0.12}$	.374		(total multiple correlation)
t	-1.122		(multiple regression covariance analysis)
df	24		
P	.275		(two tails)
F_0	2.186		
n_1	9		
n_2	17		
P	$>.05$		

[a] 0 = the posttest or criterion; 1 = the pretest; 2 = time spent on posttest; e = posttest minus pretest.

of curvilinear regression. The variances are not significantly different. Additional comments on the negative t and its interpretation will be reserved until the data for Experiments II and III are presented.

Experiment II

This is the experiment in which the 23 student nurses served as Ss; the main procedural change involved the use of the revised titles for the types of discourse. Later a breakdown of the pretest choices (item-by-item) will be made and the data discussed—after the results of Experiment III have been presented. This analysis, among other things, will clarify—if not justify—our thinking that some change was necessary.

Again the differences between the three control groups (posttest and change scores) were not STATISTICALLY significant, but the data should be discussed. Below is a summary of the results for both Experiments I and II. "0" is the criterion or posttest; "1" and "2" are the pretest and time predictors; and "e" is the change score or posttest minus pretest:

		N	$\bar{X}_0$	$\bar{X}_1$	$\bar{X}_2$	$\bar{X}_e$
Experiment I:	Feigl	9	18.11	16.11	15.33	+2.00
	Osgood I	5	14.60	13.00	13.00	+1.60
	Osgood II	4	16.50	16.00	12.00	+0.50
Experiment II:	Feigl	8	9.62	10.62	13.75	-1.00
	Osgood I	5	14.40	15.80	13.80	+1.40
	Osgood II	4	12.50	5.50	12.50	+7.00

All control-group mean scores for Experiment I are very comparable. Comparing change scores between experiments, the Feigl group has the largest mean change score in Experiment I and the smallest in Experiment

II, while the Osgood II group is smallest in Experiment I and largest in Experiment II. If the three groups for the two experiments are combined, the mean change score for the Osgood II group is somewhat larger than means for the other two groups. At first glance it might seem that the matrix had been of some assistance to the Osgood II Ss, even though the observed difference is not significant. However, if we look at the pre-test means, we find that the score for the Osgood II group in Experiment II is 50 per cent smaller than the next smaller score, and over 50 per cent smaller than the theoretical chance mean score of 11.625. IF the chance mean score were established for all groups, with the post-test scores remaining the same, virtually all differences between the groups' mean change scores would be eliminated. In other words, we interpret the Osgood II group's change score in Experiment II to be primarily the result of regression towards the mean. Since an evaluation of the worth of the matrix was in no way a major concern of these experiments, it was decided that no further test would be made. Again the three control groups were combined for the multiple regression analysis, and only the "Feigl Treatment" was used in Experiment III.

The group means and variances for this experiment are presented in Table III.

Again we have a negative t which does not reach our criterion of significance. In addition, the F test for the variances gives a significant P of less than .01. Thus we have further indication of the fact that the two samples are drawn from a population with a common mean. The negative t and the significant F cast even further doubt as to linearity

TABLE III

EXPERIMENT II: SUMMARY OF THE RAW DATA AND RESULTS OF THE MULTIPLE REGRESSION ANALYSIS[a]

	Morris		Others
N	6		17
$\bar{X}_0$	19.00		12.12
s_0^2	218.40		40.86
$\bar{X}_1$	17.00		10.53
$\bar{X}_2$	13.50		13.47
$r_{0c \cdot e}$		0.41	
$r_{0.12}^2$		.033	
$r_{0.12}$		.181	(within-group multiple correlation)
$R_{0.12}^2$		.119	
$R_{0.12}$		.345	(total multiple correlation)
t		-1.179	(multiple regression covariance analysis)
df		19	
P		.253	(two tails)
F_0		5.345	
n_1		5	
n_2		16	
P		$< .01$	

[a] 0 = the posttest or criterion; 1 = the pretest; 2 = time spent on posttest; c = posttest minus pretest.

of regression and--by implication--the advisability of the use of the covariance analysis. For purposes of comparison, let's take a look at the change-score mean difference and the straightforward t test of this difference. Here the mean for the Morris group is 0.41 points LARGER than the Others mean, and the t is POSITIVE although insignificant (a P of .48 for one tail). An F test of the corresponding change-score variances which equals 6.05 almost reaches the one per cent significance point. This is the same story that the multiple regression analysis is telling, but the t for the noncorrected mean change score favors the Morris group so that it is positive.

As we have already seen, sampling played a more important role in the pretest means of this experiment. The mean change score for the Morris group is very slightly larger than that for the control group, but the mean pretest scores show a difference of 6.47 with the Morris group the larger (38 per cent) of the two means. Again regression towards the mean is working against the experimental group, while the control group pre- and posttest means hover close to the theoretical chance mean score of 11.625. The mean pretest score for both groups combined very closely approximates that for Experiment I; this fact gives some reason for believing that the changed titles for the types of discourse did not depress scores on the pretest. This finding is more impressive when it is recalled that the Ss of Experiment I had considerably more academic training than those of Experiment II, and they should have been more familiar with material similar to that used in the test paragraphs.

Experiment III

It will be recalled that the purpose of this experiment was to determine the effects of high intelligence and long academic training on test performance, and thus to add generality to the results. Recall that only the "Morris Treatment" and the "Feigl Treatment" were used. A summary of the multiple regression analysis is given in Table IV. In this experiment the criterion mean (O) is slightly larger for the Feigl group. The t equaled 1.111 for a P value of .286 for two tails. Re the multiple regression analysis, this is the first positive t and the first instance in which the Morris mean criterion score was the smaller of the two. The F test of the criterion variances was 1.780, which does not reach the five per cent significance point. Though this t for the corrected mean difference is positive, theoretically it should be interpreted in the same way as the first two t's, i.e., as favoring the control group. Despite two of the three obtained posttest means favoring the Morris groups--and the third virtually a tie--if the predictor scores were adjusted by means of covariance (the regression line for the adjusted pretest and time scores), the predicted posttest mean scores would be higher for the control groups. The P values for the multiple regression analyses could be combined by chi square to yield a smaller probability value. The combination of the change-score P's would be the same, but would tell a diametrically opposed story, since now--for the first time--the t for the change-score mean difference is negative; i.e., the control-group mean is the larger. In the light of the three F values, it would be the variances as well as the means which would differ.

TABLE IV

EXPERIMENT III: SUMMARY OF THE RAW DATA AND RESULTS OF THE MULTIPLE REGRESSION ANALYSIS[a]

	Morris		Others
N	9		9
$\bar{X}_0$	31.56		31.89
s_0^2	269.03		151.11
$\bar{X}_1$	29.56		28.44
$\bar{X}_2$	18.44		11.22
$\bar{X}_{c_{M-O}}$		-1.45	
$r_{0.12}^2$		.218	
$r_{0.12}$		.467	(within-group multiple correlation)
$R_{0.12}^2$		.145	
$R_{0.12}$		.381	(total multiple correlation)
t		1.111	(multiple regression covariance analysis)
df		14	
P		.286	(two tails)
F_0		1.780	
n_1		8	
n_2		8	
P		$> .05$	

[a] 0 = the posttest or criterion; 1 = the pretest; 2 = time spent on posttest; c = posttest minus pretest.

In sum, the results of the multiple regression analyses are not significant, although they lead us to suspect that the samples have not been drawn from a common population as far as pretest mean score and/or time spent on the posttest are concerned. In other words, the predictors are not working uniformly between groups. Let us examine the intragroup correlations to see if we can determine more exactly the effects of the predictors. This should tell us what is wrong with the previous statistic. The intragroup correlations are presented in Table V.

Table V reveals that the lowest multiple correlation for the Morris groups (.681) is larger than the highest (.386) for the three control groups. (This same relationship holds when the C. M. scoring is applied to Experiment I; see Chapter VI.) As previously mentioned, it is not possible to test the differences between multiple correlations. We could test each correlation against zero, i.e., as being significantly different from a hypothetical correlation of zero, but--for our purposes--this would not be too meaningful. However, one can determine the probability value for the observed ordering of the correlations. Using the formula $_nP_r = \frac{n!}{(n-r)!}$ for r number of events taken n at a time, for this case we have 6 events and thus $_nP_n = n!$ and $_6P_6 = 6!$ or 720--the total number of combinations for 6 such events. If the correlations were to be grouped into top and bottom halves, the number of combinations would be $_3P_3 = 3!$ or 6 per each half and 36 for the observed ordering. Thus 720/36 = 20, and there is 1 chance in 20 that an arrangement of 6 multiple correlations would yield a split such that the smallest value of one group would be higher than the highest value for the other group;

TABLE V

THE INTRAGROUP COEFFICIENTS OF CORRELATION

		Morris		Others	
Experiment I:		1	2	1	2
	0	.679	.083	.129	-.078
	1		-.046		-.528
		$r_{0.12}^2$ = .464		$r_{0.12}^2$ = .017	
		$r_{0.12}$ = .681		$r_{0.12}$ = .130	
		Morris		**Others**	
Experiment II:		1	2	1	2
	0	-.479	.842	.091	-.001
	1		-.543		.348
		$r_{0.12}^2$ = .710		$r_{0.12}^2$ = .009	
		$r_{0.12}$ = .842		$r_{0.12}$ = .096	
		Morris		**Feigl**	
Experiment III:		1	2	1	2
	0	.207	.723	.383	-.128
	1		.410		-.427
		$r_{0.12}^2$ = .532		$r_{0.12}^2$ = .148	
		$r_{0.12}$ = .730		$r_{0.12}$ = .386	

thus we have a significant P of .05 for this event. (This P of .05 represents an upper limit of difference or the largest P value; i.e., combining P values or the use of any other test of significance would necessarily yield a smaller probability figure.) What is the meaning of this finding? Primarily it indicates that we could determine the regression line for the Morris group, and by knowing the pretest score and posttest time score for a given S, we could decrease the magnitude of the errors of prediction from what they would be if we used only the group means for our prediction. But--since the multiple correlations are so low for the control groups--we could do little better than predict the group mean score for each individual S. In short--the combined predictors PREDICT for the experimental groups; they do NOT predict for the control groups. Thus we know that the three t's obtained for the multiple regression analyses were not only not significant but might lead to misinterpretation.

Examining the intragroup correlations in more detail, we find that the criterion (posttest score) and time correlations (r_{02}) are fairly high and positive for the Morris group and low and negative for the control group in each separate experiment. How do we interpret these correlations? Shouldn't they predict larger criterion mean scores for the Morris groups, and wasn't just the opposite of this the case when multiple covariance was used? The answer to this latter question would have to be yes, BUT the answer to the riddle seems to be that time works differentially--only for those groups that had the "Morris Treatment"--and so time is a resultant and NOT a predictor; i.e., Ss who had the "Morris

Treatment" spent more time on the posttest very probably because the treatment gave promise of helping them with the experimental task. Since the control treatments did NOT offer as much help--in the opinion of the Ss--they did not spend so much time on the posttest, AND the correlations indicate that additional time would not have aided them. The multiple regression analyses, on the other hand, imply (give credit re the expected posttest scores) that the Ss would have had higher scores if they had worked longer at the task. We will find later that this is not the case. If the differences between the z scores for Morris versus Others for r_{02} are computed and divided by the standard error, P values of .3632, .0268, and .0351 are obtained. Combining these values gives a chi square of 16.068, which leads to a P of .014 for these three events with six degrees of freedom. These between-group comparisons give further support to the argument that time spent on the posttest is a result of having had the "Morris Treatment."

What of the correlation between pretest and criterion scores? Inspection of the intragroup correlations will point to no consistent order, and in two cases the correlation is higher for the control groups. Over-all the correlations are fair, but how can the r_{01} of -.479 for the Morris group in Experiment II be explained? The best interpretation would seem to be that this relationship is confounded; i.e., to the extent that pretest score is related to positive regression, intelligence, motivation, or unknown factors, it correlates positively with the criterion; if the high pretest score is a chance phenomenon or the result of some unknown factor, negative regression towards the mean (posttest) will

occur and produce a negative correlation. The mean pretest score for the experimental group is larger in every case, and in Experiment II this difference amounts to 6.47 points or 38 per cent; thus the posttest scores are very apt to be confounded (for the Morris group) by regression downward and so lead to the negative correlation between pre- and post-test scores. So--to sum up--pretest score does seem to be a predictor (positive or negative), but downward regression, if actually a factor, always works against our experimental groups.

Thus the conclusion is reached that the multiple regression analyses and the intragroup correlations and tests of significance tell us a lot about the nature of the pretest and time variables and something of the differences between the groups, BUT covariance is the WRONG statistic for determining the effects (direction and magnitude) of the three treatments. The change-score t tests indicate that the effect of the "Morris Treatment" cannot be demonstrated for all Ss experiment by experiment, though--if we chose--we could throw all Ss into one experimental and one control group and get a t which favored the Morris group.

Some might ask, "If curvilinearity of regression is suspected, why not test this and--if it is found to be significant--make the proper allowances and compute new multiple regression t's which are corrected for curvilinearity and establish P values for these t's?" The answer to this is twofold: (1) E was not interested in predicting posttest scores for the data at hand but rather in determining the effectiveness of the "Morris Treatment" when freed from contaminating factors. (2) Another test (Experiment V) was planned which might not only add further

generality but might circumvent some of the problems which were uncovered in Experiments I, II, and III. This experimental test should prove more satisfactory than any statistical manipulation.

The obvious next step was to experimentally match the experimental and control groups on pretest and time scores. To minimize the effects of regression, it was decided to use only those Ss whose pretest scores fell in the middle part of the distribution. The role of regression can be shown in another way. For example, in Experiment I the median pretest score for both groups combined is 12.5; those Ss above the median lose 1.60 points on the average, while those below gain an average of 6.64 points. In other words, the treatments, etc., minimize downward regression and maximize upward regression. To the extent that the pretest scores correlate with intelligence and/or motivation, etc., Ss with large pretest scores can resist the effects of downward regression. Those Ss with below-average pretest scores who "understand" the treatment facilitate upward regression. (This is true for all experiments. For another demonstration of this phenomenon, see page 128 of Chapter VII.) Since E was interested in testing the usefulness of the "Morris Treatment" and because one of the results of this treatment was to produce longer posttest working times, E wanted to use only those Ss who worked at the task. But how define working? The posttest instructions asked the Ss to spend approximately 5 minutes per paragraph, but if this definition of working were used, we would have NO Ss. G. M. (page 100, Chapter VI) spent 21 minutes judging the same paragraphs; were this the definition, we would have but 5 experimental Ss and 1 control S. The mean working

time for all Morris Ss is 16.24 minutes; this number is approximately 1 minute more than half the amount of time that the Ss were asked to spend. This figure seemed reasonable enough, and it gave a pool of 23 Ss--13 experimental and 10 control.

Objections can be raised to ex post facto studies. Freedman (24) has pointed out certain cautions re the interpretation of the results of matched-group experiments when the percentage of matched cases is not large. To anticipate these fears, and try to allay them, it should be stated that the criteria of selection (working and mid-range pretest score) are independent of the data. This case is in no way parallel to drawing inferences from a questionnaire in which there was only a 40 per cent return, for we do have ALL the data. Our case is more analygous to Stouffer's (83), who was trying to evaluate the effectiveness of a film in changing soldier attitudes. Stouffer pleads for a design similar to E's, namely:

	Before	After	
Experimental Group	x 1	x 2	d = x 2 - x 1
Control Group	x'1	x'2	d' = x'2 - x'1

The question to be asked is whether or not d is significantly greater than d'. Stouffer adds, "One of the troubles with using this careful design was that the effectiveness of a single film when thus measured turned out to be so slight" (83, p. 356). For the actual test those Ss who remembered seeing the film were compared with those who did not remember it. In other words, he was not interested in the effects of a film on those soldiers who slept through it. It is the same with us; we are

not concerned with evaluating the effectiveness of the "Morris Treatment" on those Ss who couldn't or wouldn't use it. Parenthetically, the group of nonworkers should be studied, but recall that the Ss had one approximately 20-minute exposure to the treatment; it is not surprising that some of them should not be able to employ the system. Most teachers would be highly gratified if 50 per cent of their students grasped reasonably well a 20-minute lecture on abstract and difficult material! Moreover, E was not asking questions about the "Morris Treatment" but was giving the Ss the more difficult task of using it.

That the "Morris Treatment" influenced Ss to spend more time on the posttest can be demonstrated in another way. By our definition of working, 13 of 25 experimental Ss were workers (52 per cent), while only 10 of 44 control Ss were workers. Analysing these frequencies by chi square (4.901) for one degree of freedom leads to a P value of .0385; the Ss were not drawn from a common population re work.

The mean and standard deviation of the pretest scores for these 23 Se--workers by definition--were computed; plus and minus one standard deviation from the mean was used, giving a range from 8 to 31 points. When only those Ss whose pretest scores fell within this range were selected, we were left with 16 Ss--8 experimental and 8 control. Table VI presents the data for these Ss in summary form. Inspection of Table VI will show extremely comparable matching on pretest and time. A t test of the difference between change-score means ($\bar{X}_{c_{n-e}}$) equaled 2.416, which for 14 degrees of freedom gave a P value of .015 for one tail of the distribution; this is considerably smaller than the significance point that

TABLE VI

COMPARISON OF WORKERS FROM EXPERIMENTS I, II, AND III[a]

	Morris	Others
N	8	8
$\bar{X}_1$	17.375	17.750
$\bar{X}_2$	19.375	19.625
$\bar{X}_c$	9.375	-4.000
$\bar{X}_0$	26.750	13.750
s_c^2	132.839	112.286
s_c	11.526	10.597
s_0^2	151.357	83.375
s_0	12.304	9.131

$\bar{X}_{c_{M-O}}$ = 13.375 $\quad$ $\bar{X}_{c_{O-O}}$ = 13.000

$t_{\bar{X}_{c_{M-O}}}$ = 2.416 $\quad$ $t_{\bar{X}_{c_{O-O}}}$ = 2.400

P = .015 (one tail) $\quad$ P = .0154 (one tail)

df = 14 $\quad$ df = 14

[a] 0 = the posttest; 1 = the pretest; 2 = time spent on posttest; c = change score.

E had agreed to accept. Because pretest scores are related to posttest scores, the independence of the change scores and thus the advisability of employing the t test might be questioned by some students. So the t test was applied to the posttest mean difference ($\bar{X}_{e_{post}-c}$); this t of 2.400, giving a P value of .0154 for 14 df, is not different from the P for change scores. Actually the variances for change scores are more comparable than those for posttest scores. Some investigators might wonder about the assumption of normality as applied to this data; for this reason it was thought advisable to use a nonparametric test. The Wilcoxson Sum (95) was applied; this test led to a P of .0175 for the change scores and .0225 for the posttest scores; these results are very similar to those obtained with the t tests.

The Ss, the experiments from which they were drawn, and the raw data will be found in Appendix G. One question that might be asked would inquire as to the significance of the fact that three of the Morris workers were taken from Experiment III and none of the control-group workers came from this experiment. This was the experiment which supposedly provided the most intelligent and educated of our Ss. The most striking difference between these and other Ss is found in the pretest scores; the mean for Experiment III is 29.000, while for I it is 15.250 and for II, 12.217. Recall that the names of the types of discourse were changed for Experiments II and III; the pretest mean for Experiment II is comparable to that for Experiment I. A t of 5.84 (df 44) is obtained when the pretest mean of Experiment III is compared with that for Experiment I, and a t of 6.99 (df 39) for Experiment III versus Experiment II;

both of these P values are highly significant. At least we are certain that the Ss were not drawn from the same pretest population as those for the other two experiments. For another point it should be stated that most of the Ss for Experiment III were willing volunteers; two-thirds of the Morris group were workers according to E's definition; this is a considerably higher percentage than was produced by the other two experiments. The Ss for Experiments I and II were tested during class time; the motivation for the Ss of Experiment I may have been low because they had served as Ss in several other experiments and at the time of our test were preparing for the final course examination. The nurses of Experiment II had little at stake and may not have been highly motivated. On the other hand, the Ss of Experiment III may have spent more time on the "Morris Treatment" because they were better able to understand it. Whatever the reason the three Morris workers from Experiment III, despite very high pretest scores (and potential effect of regression), showed greater average change scores than other Morris workers. It very well may be that high intelligence and long academic training do produce more workers and higher change scores. (The actual percentages of workers were 50, 33, and 67 for Experiments I, II, and III respectively.) The control group for Experiment III had the smallest mean time of any of the three control groups. (A t test of the difference between the Morris and Feigl means for time was significant at the .052 point in Experiment III.) There are at least two possible explanations for this: (1) The Ss were experimentally sophisticated and thus realized that they were in a control group; however, they did have the "Feigl Treatment,"

which--if not of definite assistance in solving the experimental problem--would at least serve to sharpen the question at issue. (2) Those Ss who did work longer on the control treatments only became confused. The control workers' pretest mean is approximately four points larger than that of all control Ss in Experiments I and II, and interestingly enough their change score is minus four; for these Ss the change score would seem to be more the result of regression than confusion.

The problem remains of analysing the data for the nonworking Ss. Perhaps the simplest method--if not the most dramatic--would be to compare all experimental and control Ss who were not used in the "worker" analysis. This way no data will be wasted; i.e., we will have tests involving all 69 Ss instead of 16. This analysis will include the four workers from Experiment III (three Morris and one Feigl) who were not used in the worker analysis proper. Summaries of the data and the tests of significance for both posttest scores and change scores for these 53 Ss are given in Table VII. Though the difference between mean change scores favors the control group, this difference is not significant. The F test of the variances, however, is highly significant. When the t test is applied to the posttest mean difference--which now favors the experimental group--the P value is again not significant, while the P value for the F test is less than .05; the means are not significantly different, but the variances are. A glance at $\bar{X}_1$ gives reason to believe that the samples are not drawn from a population with a common pretest mean. In short, there are no important, consistent differences between these two groups that cannot be accounted for by regression, sampling,

TABLE VII

RAW DATA SUMMARY AND TESTS OF SIGNIFICANCE FOR ALL SS NOT USED IN WORKER ANALYSIS[a]

	Morris	Others
N	17	36
$\bar{X}_1$	22.118	15.750
$\bar{X}_2$	14.765	12.471
$\bar{X}_e$	-0.059	3.278
$\bar{X}_0$	22.059	19.028
s_e^2	303.809	106.378
s_e	17.430	10.315
s_0^2	257.059	125.571
s_0	5.070	11.207
	$r_{e_2 \cdot e} = .731$	$r_{e_2 \cdot e} = .703$
	$P \approx .80$ (two tails)	$P \approx .50$ (two tails)
	df = 51	df = 51
	$F_e = 2.856$	$F_0 = 2.047$
	$P < .01$	$P < .05$
	$n_1 = 16$	$n_1 = 16$
	$n_2 = 35$	$n_2 = 35$

[a] 0 = the posttest; 1 = the pretest; 2 = time spent on posttest; e = change score.

or unknown and nonsystematic factors. At any rate, the selection of the 16 workers has not biased the data given by the remaining Ss. Various other tests could have been made. For example, the posttest means could be tested for a difference significantly higher than the chance mean of 11.625; this might very well be significant for the Morris group. However, these three experiments were not designed with that test in mind. At this point E is interested only in justifying the test involving the 16 workers and in demonstrating the effectiveness of the "Morris Treatment" for those Ss in the middle of the pretest distribution who work at the task.

It would be informative if some method were available to determine the combination of factors (intelligence, dullness, motivation, etc.) that lead (differentially) to working at the task. It would be useful, also, to know the factors (including chance) and their respective weights which go to make up a given pretest score. In the future an attempt to control these variables may be made, but their effect can only be surmised at this time. So that time, material, and data need not be wasted, E might--for example--discard all Ss who failed to work a specified amount of time on the pretest. Be this as it may, the worker analysis does seem to indicate the effectiveness of the "Morris Treatment," and--for the moment--there is no need to generalize beyond this specified target population.

In conclusion, we should take a summary glance at the separate test items--the six paragraphs. Table VIII presents the choices of all Ss of Experiment I, i.e., the labels they attached to the pretest

TABLE VIII

EXPERIMENT I: LABELS GIVEN BY ALL Ss ON PRETEST[a]

Types of Discourse	Paragraphs 1	2	3	4	5	6	Number Ss Selecting
Cosmological			2	2		11	15
Critical	1	13	5		1		20
Fictional	1	2					3
Grammatical	1		2				3
Legal		1					1
Logical-mathematical			1	*1	*6	14	22
Metaphysical	10	6	1	6	4		27
Moral	1		2		1		4
Mythical	1	*1	*1		1	1	5
Poetic					1		1
Political			1				1
Propagandistic					1		1
Religious			1				1
Rhetorical	*2	4	5	1	3	1	16
Scientific	7	1	4	9	6		27
Technological	4		3	9	4	*1	21
	N = 28	28	28	28	28	28	Σ=168

[a]Correct E answer = *

paragraphs. E proposes no specific analysis of Table VIII. Instead some general comments will be offered, and E hopes that the reader will have some speculations of his own as to the underlying state of semantic affairs which contributed to the obtained data.

Inspection of Table VIII shows that seven of the types of discourse received 60 per cent of the total possible choices. As far as the test was concerned, certain types of discourse had very low drawing power. A glance at the titles for these types of discourse will indicate that the labels themselves probably have very definite connotations for the Ss; they appeared to know that a certain paragraph was not--say--religious discourse. In light of the definitions given for "cosmological," "metaphysical," and "rhetorical" in the pretest instruction sheet for this experiment, the selection of one of these labels by a given S in all probability indicates that he has no idea what to title that paragraph. As we shall see later, item No. 3 proved to be one of the best for demonstrating the effectiveness of the "Morris Treatment," and it produced a wide spread of pretest selections. Reasons like this led E to change the titles of the types of discourse--making them more abstract--for Experiments II, III, and V. Turning to the individual items, E would say that the Ss had no idea how to label item No. 1 save that 7 Ss thought it scientific since it was concerned with neurology. Re item No. 2, 15 Ss thought it critical because the author seemed to be saying negative things about himself and the word "critic" appears in the passage. Item No. 3 shows no consistency. The Ss are rather certain that paragraph No. 4 is either scientific or technological discourse,

while No. 5 somewhat resembles the profile of No. 4. Item No. 6 did at least look like mathematics to 14 Ss, but it looked like Greek ("cosmological") to 11 Ss; (this is doubtless a very proper split which most teachers of statistics could verify).

Since the names of the types of discourse were revised, E does not want to enter into any comparison of all the items across the three experiments; instead we will have a look at item No. 5 and for Experiment III item No. 4 also. Some justifiable reasons exist for calling items No. 4 and No. 5 scientific. (C. M. applies this label to No. 4, and E uses it for No. 5.) Regardless of what the reader's semiotic may be, E is certain that he would hesitate to call any of the other paragraphs scientific discourse. What of the Ss of Experiment I, some of whom hold college degrees? Seven of these Ss thought item No. 1 to be scientific discourse. Four voted for No. 3. E has no evidence to show how such semantic fuzziness affects the behavior of these 11 Ss, nor does he know that they would be helped by courses IN sciences or a course in semiotic. But if these Ss act as they select on this test, their behavior at worst would be confused and indecisive; at best it would exemplify Kenneth Burke's point that rhetoric has dropped out of the literature curriculum only to reappear clothed in other garbs--this time in the robes of science. For what it is worth, the writer would like to venture his personal opinion, which holds that the authors of paragraphs No. 1 and No. 3 KNEW what they were trying to do and had no intention of writing scientific discourse. True Dr. Russel Brain, the author of paragraph No. 1, is both a medical and scientific practitioner. HE IS ALSO A

MASTER OF RHETORIC, and the paragraph in question is the introduction to a speech prepared for the B. B. C.--part of a symposium designed to reach the layman--on the modern scientific conceptions of mind and brain. He succeeded admirably as the reading of his essay will readily reveal. Elton Mayo's words in paragraph No. 3 are well worth writing, but there is more heat than scientific light in them.

For Experiment II "scientific" was renamed "testable." Only 14 of the 23 Ss (versus 100 per cent of the Ss in Experiment I) used this new title; thus we have some reason to believe that the new titles were at least encouraging the use of more of them. Interestingly enough, none of these Ss thought that item No. 4 was "testable." Three applied this label to item No. 1; 1 for item No. 2; 5 used it for item No. 5 (E's correct answer); and 5 gave this title to item No. 6. Again specific determiners in the paragraphs themselves probably played an important role in eliciting these answers. For example, the author is reporting a test in item No. 5, and the word "test" appears in item No. 6.

When we turn to Experiment III, these academically advanced Ss apply the label "testable" in the following manner: Two Ss so label item No. 1; 4 Ss nominate item No. 4 for this label; 6 so label item No. 5; while 1 S uses this title for item No. 6. No significance tests were applied to any of the data listed above, but it would seem that these Ss had learned something about reading or reading matter. There were 18 Ss in this experiment; recall that 8 of the Ss were psychology majors (7 of these being advanced graduate students), 6 were Ph. D. candidates in physics, and 3 were advanced graduate students plus 1 undergraduate in

philosophy. On the pretest 3 of the physicists and 2 of the psychologists did not label any item "testable." Regardless of whether the treatment received was Feigl or Morris, all but 1 S (a philosophy student who called no item testable) labeled either item No. 4 (9 Ss) or item No. 5 (8 Ss) as being testable discourse. The two treatments had altered the responses of the Ss. (This striking change cannot be shown--for testable discourse at least--for the Ss of Experiments I and II.) It is interesting that 8 of the 9 Ss who labeled item No. 4 "testable" were either philosophy or psychology students; only 1 was a physicist. Five physicists and 3 psychologists called item No. 5 "testable." Fisher's exact test might show this split (philosophy and psychology students versus students of physics) to be significant. The experiment was not designed to test any such hypothesis, and so no significance test was applied; but a future experiment could readily test for this, namely, that advanced academic training is not only training to read but training to prefer certain types of discourse. Item No. 4 is Clark Hull's first postulate in his behavior theory and thus might be expected to be valued as being at least a testable proposition by psychologists and philosophers. Item No. 5, which is straightforward laboratory reporting and thus resembles more closely the actual experimentation of physicists, had a more testable form for physics students. E favored labeling item No. 5 "testable," while C. M. chose item No. 4. Either choice satisfies E, BUT do we have in this dichotomy a demonstration of the lack of agreement between scientists as to the criteria for scientific discourse--a lack of agreement that might foster a failure of communication even among scientists? Future experimentation will have to provide the answers to these

questions; still--since it is impossible for any considerable number of scientists to master anything but their own, and closely related, subject matter--would it be profitable to teach them something about the nature of words and semiotic so that they could read and more accurately evaluate the work of others? In time some fundamental agreement re scientific discourse might be reached, and this could lead to increased interdisciplinary communication.

Picking up the threads of this chapter, the writer would suggest that the "Morris Treatment" version of C. M.'s semiotic did aid those Ss who worked at it and who were not overly penalized by regression. Moreover, the sometimes weird labeling that was shown in the pretest of Experiment I would indicate that SOME training re types of discourse and the differences between types is desirable. If we have no panacea to offer the student, we can at least point out the fallacy of "a word is a word is a word." Arguments re the denotational aspects of words aside, the signification of even the same word depends not only on <u>what</u> sort of word it is but also on <u>who</u> is using it and for what <u>purpose</u>! If--as has many times been suggested not only by E but others more qualified than E as well--education is primarily training to read with stress on the favored discourse of a given student's area of specialization, educators are remiss if they fail to outline something of the nature of written and spoken discourse and certain of the limitations of their own favorite brand! (E has closed the two preceding sentences with exclamation points lest he trap you into believing that he has any considerable amount of <u>empirical</u> data to back them up; E hopes that those sentences are <u>behaviorally</u> incitive.)

CHAPTER VI

EXPERIMENT IV: RESULTS AND DISCUSSION (THE C. M. SCORING KEY)

This experiment was designed to determine (1) the amount of agreement (rater-reliability) between C. M.'s scoring of the paragraphs and E's, i.e., the relationship between our judgments as criteria; (2) the effects on the test results of another scoring key (C. M.'s); (3) whether or not a given item was better (more clearly showed the effects of the "Morris Treatment") if C. M. and E were in complete (type-of-discourse) agreement in judging it.

In Experiments I, II, and III, E has not been concerned with the construction of a reliable test. Instead E tried to develop an adequate test, i.e., one which might demonstrate the effects of the "Morris Treatment." One would need (1) to know whether or not it was worthwhile improving a test and (2) to gather data on a test to learn its characteristics before attempting to increase its reliability. Prior to these attempts no test existed. While the methodology for increasing the reliability of a test was readily available, data were needed to warrant its use. One question was this: If C. M. and E did agree on a given paragraph, would this agreement be reflected in the improvement or change scores of the experimental Ss for that item or items?

Upon completion of Experiments I, II, and III, the 12 paragraphs, with no identifying information, were sent to C. M., who was asked (1) to determine the mode of signifying and the use for each paragraph; (2) to state the amount of time (in minutes) required to score each paragraph;

and (5) to express the degree of confidence, from 1 (Highest Confidence) to 5 (Lowest Confidence), in each of his 12 judgments. A chart of C. M.'s scoring of the 12 paragraphs appears on the next page. The column headed "X-Choice" indicates the number of E's choice (see page 53, Chapter IV) which matches C. M.'s in terms of type of discourse. For example, he labeled paragraph No. 1 "prescriptive-incitive" (religious discourse); none of E's three selections were "religious"; therefore, under "X-Choice" the number "0" is entered to indicate that we did not agree completely at any point on this selection. The column "P-Choice" shows the partial or total agreement between his choice and E's FIRST choice; e.g., we agree as to mode on paragraph No. 2. Throughout Table IX the nomenclature of Experiment I is followed (see "Morris Treatment," Experiment I, Appendix B).

It should be remembered that E had the advantage or disadvantage of knowing the authors of the paragraphs, the sources, and what came before and after the actual selections. C. M. worked only with the paragraphs, although he may have recognized certain of them as being the work of authors familiar to him. Recall also that E did not choose obvious examples of discourse, and for a sample as short as a paragraph to be clear-cut would almost require that the author write with the classificatory scheme in mind. Despite these shortcomings and possible sources of error, both C. M. and E consider our agreement high.

On page 101 will be found Table X, in which C. M.'s evaluation of the 12 paragraphs is compared with E's ("Experimental Key"). The 1-4-16 score plan was used. "Score I" is C. M.'s score using the

TABLE IX

C. M.'S PARAGRAPH EVALUATIONS

Paragraph Number	Scoring: Mode	Use	Type of Discourse	Time	Confidence	X-Choice	P-Choice
1	Prescriptive-Insitive		"Religious"	5	2	0	——
2	Appraisive-Evaluative		"Poetic"	5	1	2	Mode
3	Appraisive-Evaluative		"Poetic"	5	2	1	Mode-Use
4	Designative-Informative		"Scientific"	1	1	2	Use
5	Designative-Insitive		"Legal"	3	1	2	Mode
6	Formative-Insitive		"Grammatical"	2	1	•2	——
7	Designative-Evaluative		"Fictional"	5	2	0	Use
8	Designative-Informative		"Scientific"	1	1	1	Mode-Use
9	Appraisive-Informative		"Mythical"	3	1	•3	Mode
10	Designative-Insitive		"Legal"	2	2	•2	Use
11	Designative-Evaluative		"Fictional"	5	4	3	——
12	Appraisive-Insitive		"Moral"	3	2	1	Mode-Use

TABLE X

C. M. VERSUS E: PARAGRAPH SCORING

Paragraph Number	C. M.	Experimental Key	Score I	Score II
1.	Religious	Rhetorical	0	
2.	Poetic	Mythical	1	
3.	Poetic	Poetic	16	
4.	Scientific	Logical-Mathematical	4	
5.	Legal	Scientific	1	
*6.	Grammatical	Technological	0	16
7.	Fictional	Poetic	4	
8.	Scientific	Scientific	16	
*9.	Mythical	Moral	1	16
*10.	Legal	Religious	4	16
11.	Fictional	Metaphysical	0	
12.	Moral	Moral	16	
			Σ = 63	106

Perfect Score = 192 (i.e., 16 for type of discourse times 12 paragraphs)

Chance Total Score = 23.25 (i.e., 1.9375 per paragraph)

"experimental key" as the criterion. The score would be the same for E if C. M.'s judgments were used as the key. The column headed "Score II" indicates our revised judgments after one exchange of letters. The two paragraphs with a single asterisk are those where C. M. changed his opinion and agreed with E's scoring; the double asterisk marks the one paragraph that E changed to agree with C. M.'s evaluation.

To determine the amount of agreement between C. M. and E, a method that is commonly employed in scoring objective exams was followed. The total perfect score was computed; from this sum the total chance score (number of points obtained on the basis of chance) was subtracted. This figure gives the total number of points for a perfect score after allowing for chance. Next the number of points alloted to C. M.'s evaluations was computed, and the chance score was subtracted from it. Then the amount of agreement (per cent correct) over and above chance can be determined by dividing the derived score by the derived perfect score. These results are given below:

63.00 (obtained score)	192.00 (perfect score)
23.25 (chance score)	23.25 (chance score)
39.75	168.75

39.75 ÷ 168.75 = .236 or 24% Agreement

The numbers above show that using the sum of 63, C. M. and E are in 24 per cent agreement AFTER allowing for chance. The sum is approximately three times chance. We agreed completely on 3 of the 12 items (25 per cent); 25 per cent of the items showed no agreement.

After C. M. had sent E the scored paragraphs, E returned them to him along with his scoring and asked him to evaluate E's judgments.

C. M. then explained his scoring and his thinking about our differences and changed his opinion on items No. 9 and No. 10 (see "Score II," Table X) and agreed with E. E went over the items in turn and decided that C. M.'s scoring of item No. 6 was correct. This gave the following agreement data:

106.00 (obtained score)
23.25 (chance score)
82.75

82.75 ÷ 168.75 = .490 or 49% Agreement

Thus after one exchange of letters, our agreement reached 49 per cent with 50 per cent agreement as to types of discourse.

Prior to this experimentation both C. M. and E felt that mode of signifying would be a more objective measure than use. It should be in fact when quantification devices are worked out, but at this stage there are no differences between the two; exclusive of the types of discourse in common, we agreed on mode of signifying three times (items 2, 5, and 9) and use three times (items 4, 7, and 10). The second scoring yielded two mode of signifying (items 2 and 5) and two use (items 4 and 7). The fact that our opinions changed on certain of the items is not necessarily to be interpreted as either lack of character on the part of the judges or as fallibility of judges or classificatory scheme, but rather as one of the advantages of the system. Science is often spoken of as a method of agreement. The basis of this agreement is the language of science; i.e., scientists, for the most part, should be able to agree as to the signification of their signs or be able to set about the task of establishing the signification of their signs. By the same token, employment

of the mode-use classification system should help forward communication between judges and between teacher or faculty adviser and student. If agreement cannot be reached about a given piece of writing, the author can employ the system--if he chooses--so that mode of signifying and use will be more precise.

It is interesting to note items 1, 6, and 11 (the reader is asked to re-examine these paragraphs in Appendix C), where we disagree on both mode of signifying and use, and where no one of E's choices for item No. 1 matches that of C. M. re type of discourse. Each of the three involves the formative mode. This mode of signifying poses some special problems. E offers no claims that the "Morris Treatment" is a precisely accurate presentation of C. M.'s views. After Experiments I, II, and III were completed, the summary was sent to C. M., who wrote this about it:

> The main source of disagreement stems, it seems, from a different meaning we give to the term "formator"--which occurs in 3 of your paragraph analyses. I do not think your account of formators on p. 3 of your document is very clear . . . particularly the words "if a sentence, etc., defies classification, it is probably in the formative mode."
>
> I use "formative" in a narrow sense--perhaps too narrow. Namely, for "tautologies" such as "2 + 2 = 4," "a = a," "a moral person ought to do right."
>
> I think you have a wider use of "formative" than I have--and perhaps also of "systemic."
>
> The latter (meaning E's summary) seems to me a good simplification (56).

C. M. was curious to know if the Ss had difficulty with formative discourse. The preliminary tryout indicated that they did; the concept seemed to be particularly puzzling for those untrained in formal logic. Therefore, in the experimental tests, E tried to lead the Ss to

concentrate on the other three modes of signifying and uses of signs and to select the formative only on the basis of elimination. This procedure seemed expedient and justifiable. On page 3 of the first revision of the "Morris Treatment," E tried to exemplify formative discourse because Experiment I seemed to indicate that despite the elimination "rule" (page 3, also)--which C. M. felt might better be omitted--Ss were still not clear about the formative mode of signifying and the systemic use--of signs. This latter change and the elimination rule were the only criticisms of the "Morris Treatment" that C. M. expressed. However, E's interpretation of formative discourse remains more broad than that of C. M. For one thing, it seems that tentatively labeling a bit of discourse as formative, especially if the writing and the author's purpose are not clear, will at least serve to delay judgment until a larger sample of his writings can be obtained. It is interesting to note that the Ss scored the paragraphs more nearly in agreement with E than with C. M.; this has led E to wonder if this method could not be broadened and used to measure, however indirectly, differences between author and reader in interpretation. That is, despite the fact that C. M. calls it a good simplification, would Ss more closely agree with C. M.'s scoring than with E's if C. M. had written the summary? (This point is discussed again later in this chapter.)

Returning to the paragraphs where agreement was minimal, item No. 1 is the introductory statement of a speech prepared for the B. B. C. Item No. 6 is the one where E now feels his scoring was incorrect and that C. M.'s is right. But item No. 11, from Fenichel's The

Psychoanalytic Theory of Neurosis, caused C. M. to spend his maximum amount of time (5 minutes) and to express his lowest degree (4) of confidence. E shall not recount his trials with the passage. However, to state that the paragraph produces a minimum of communication (interpersonality of signs) seems conservative enough. As a matter of fact, our struggles with this paragraph remind E of an article by Maloney, who quotes a conversation between T. S. Eliot and Nevill Coghill, who were discussing a production of Eliot's play, "Sweeney Agonistes," to this effect:

> . . . Nevill Coghill remarked to Eliot that "I had no idea the play meant what he made of it . . . that everyone is a Crippen. I was astonished." Eliot replied: "So was I." "But," said Coghill, "can the play mean something you didn't intend it to mean, you didn't know it meant?" Eliot: "Obviously it does." Coghill: "But can it then also mean what you did intend?" Eliot: "I hope so . . . yes, I think so." Coghill: "But if the two meanings are contradictory, is not one right and the other wrong? Must not the author be right?" Eliot: "Not necessarily, do you think? Why is either wrong?" Mr. Coghill, who is an Oxford don, thereupon changed the subject (49, p. 267).

Now this state of affairs is perhaps as it should be for poetic drama, but I doubt that Fenichel intended to write poetry! Item No. 7 is interesting also (see Appendix C). On this paragraph our first judgments agree that the use is evaluative, but no one of E's choices agrees with C. M.'s as to type of discourse; i.e., E never considered the passage to be designative in mode of signifying. Are C. M. and E reflecting differences in academic training and values? E admits that the appraisive signs found in the passage do not judge the human characteristic under discussion as being "good" or "bad," and thus they give the impression of being designative. Still, what stimulus-properties are specified by

the predicates? Would you be able to recognize a "cultural system" if you saw one? The author feels that you would not; he labels it an "abstraction." The author, if he chose, could put these statements in the designative mode of signifying; doubtless he has empirical data which justify their use, but they would have to be spelled out more fully for E to be able to recognize the "stimulus objects." (See the Glossary, Appendix A, for definitions of "designator" and "mode of signifying.") As previously mentioned, certain postulates of psychology give the impression of being designative because the derived theorems have been experimentally tested and partially verified. At least certain psychologists consider such statements as designative. What does the mathematician think of them? Perhaps the same conditions hold when an anthropologist looks at anthropological propositions. Perhaps both of us, to paraphrase paragraph No. 7, are operational designators attesting to the fact men indeed are constantly imposing on their environments private constructions and meanings.

At least the discussion of these several items should show some of the ways in which the mode-use classification can be used to further communication and agreement if only by illuminating areas where there has been a failure of communication (lack of agreement as to the signification of signs).

There remain 9 of the 12 paragraphs where E's first, second, or third choice completely agrees with that of C. M. In all 12 instances we are in the same sector of the matrix if not in the same cell. In other words, our agreement as to what a given paragraph is NOT would be

high indeed. Although C. M. had a copy of the "Morris Treatment" when he judged the paragraphs, he obviously was not following THE PRESENT WRITER'S version of HIS OWN system in reaching his decisions! In fairness to C. M. it should be remarked that he did not have in mind the particular use to which E put his semiotic when he was writing it. Further discussion and experimentation should lead to a system in which the quantifications are more explicit. But--for E--this was not the FIRST step. The prior considerations were these: (1) Is the method teachable? and (2) Do we have an initial basis of agreement even though accurate measures have not been worked out and the written samples are brief and difficult? The answer to both of these questions appears to be in the affirmative.

E was desirous of ascertaining whether or not the experimental results depended on the use of his own evaluations, i.e., using his personal judgments of the paragraphs as the scoring key or criterion. Since E independently summarized C. M.'s semiotic and selected and scored the paragraphs, what would happen to the results if C. M.'s evaluations were used as the scoring criterion? This test should serve as an inverse cross validation of the C. M. system and E's interpretation of it. This should be a severe test of the interpersonality of C. M.'s semantic metalanguage. It was felt that Experiment I was most suitable for using the C. M. scoring key, since the names of the types of discourse used in this experiment were most like his original terms. The multiple regression analysis, therefore, was used on this one experiment only. Table XI gives a tabulation of the results of this analysis. (More complete data are given in Appendix C.)

TABLE XI

EXPERIMENT I: MULTIPLE REGRESSION ANALYSIS WITH THE C. M. SCORING KEY[a]

	Morris	Others
N	10	18
$\bar{X}_1$	12.100	9.500
$\bar{X}_2$	15.900	13.944
$\bar{X}_0$	14.300	11.889
s_0^2	68.676	90.928
s_0	8.287	9.536
$R_{0.12}$[b]	.947	.489

$R_{0.12}$ = .628 (total multiple correlation)

$r_{0.12}$ = .648 (within-group multiple correlation)

t = 1.755 (multiple regression covariance analysis)

df = 24

P = .093 (two tails)

[a] 0 = the posttest (criterion); 1 = the pretest (predictor 1); 2 = time spent on posttest (predictor 2).

[b] Intergroup multiple correlation.

Table XI indicates that the use of the C. M. scoring key on the data of Experiment I gave results not quite comparable to those obtained with E's scoring key as the criterion. Again the intragroup multiple correlation is higher for the "Morris Treatment" group. The t test is in the positive direction, and the P value reaches our significance level. How are these differences to be explained? When E's scoring was used, the t was negative. For one thing, combining pretest scores for the two groups gives an $\bar{X}_1$ of 10.429, which is lower than the $\bar{X}_1$ of 15.250 which was obtained when E's scoring was used for this same experiment. The theoretical chance score per item is 1.9375, giving a total chance score per S of 11.625. No tests of significance were made for these pretest differences, but it is interesting to at least note the outcomes: Combining pretest scores for both experimental and control groups--for Experiment II, E's scoring key yields a pretest mean of 12.22, while that of C. M. gives 13.30; for Experiment III, E's mean is 29.00 and C. M.'s is 13.28. One can do little more than speculate about these differences, but it will be noted that the C. M. mean pretest scores hover consistently about the chance mean, while the mean for Experiment III (doubtless the most intelligent group) when E's scoring is used is almost three times the chance score: (1) E's key may reflect the "common-sense" evaluation of the paragraphs; (2) it hardly seems likely--since dictionaries and the writings of C. M. were consulted when the names of the types of discourse were revised--but the titles selected may suit the test paragraphs particularly well (it will be recalled that the new titles were used in Experiments II and III, but C. M. offered no

objections to these title changes); (3) it may be the result of sampling. Whatever the explanation, the revised titles--if anything--enhanced the pretest scores when E's key was used, and consequently the factor of regression worked against E as far as any E versus C. M. comparisons are concerned. Later in this chapter analyses of the data for "workers," using the C. M. key, will be made. Although these tests reach satisfactory points of significance, the improvement scores and the P values are not as great as those obtained when E's scoring was used. E is not concerned with asserting that his scoring is the more accurate. Instead E is suggesting again that the differences may result from his interpretation of C. M.'s semiotic and--since E prepared the summary statement--that the Ss' scores reflect their understanding of the summary and not their understanding of Signs, Language and Behavior. It should be possible to design an experiment in which the changes in meaning resulting from an interpretation of original statements can be measured. We know how rumors become altered and embellished after many repetitions, but as far as E knows, we have no tests which are designed to show the differences in interpretation that result, for example, from reading an article in the original German or a translation of it.

One point is obvious, namely, the C. M. key is better than E's for demonstrating the effectiveness of the "Morris Treatment" for all Ss because--by depressing mean pretest scores to approximately the chance level--the effects of regression are minimized if not completely eliminated. The reader can compare Table XII with the results for the same experiments when E's key was used.

TABLE XII

C. M. SCORING: EXPERIMENTS I, II, AND III

		Morris			Others
Experiment I	N	10			18
	$\bar{X}_1$	12.100			9.500
	$\bar{X}_0$	14.300			11.889
	$\bar{X}_e$	2.200			2.389
	s_0^2	68.678			90.928
	s_e^2	7.289			82.487
			$F_0 = 1.324$	$F_e = 11.317$	
			$n_1 = 17$	$n_1 = 17$	
			$n_2 = 9$	$n_2 = 9$	
			$P > .05$	$P < .01$	
Experiment II	N	6			17
	$\bar{X}_1$	8.167			15.118
	$\bar{X}_0$	6.000			13.353
	$\bar{X}_e$	-2.167			-1.667
	s_0^2	69.367			115.493
	s_e^2	63.367			136.566
			$F_0 = 1.383$	$F_e = 2.155$	
			$n_1 = 16$	$n_1 = 16$	
			$n_2 = 5$	$n_2 = 5$	
			$P > .05$	$P > .05$	
Experiment III	N	9			9
	$\bar{X}_1$	11.778			14.778
	$\bar{X}_0$	25.000			16.111
	$\bar{X}_e$	13.222			1.333
	s_0^2	135.250			68.611
	s_e^2	176.444			71.750
			$F_0 = 1.971$	$F_e = 2.459$	
			$n_1 = 8$	$n_1 = 8$	
			$n_2 = 8$	$n_2 = 8$	
			$P > .05$	$P > .05$	

Table XII again demonstrates how close the C. M. scoring for the pretest means approximates the chance mean of 11.625. Indeed for Experiments I and II the posttest means approximate the chance score. Regression is not an important factor if the C. M. key is used. Moreover, since every one of the pretest means is not significantly different from chance, we have no cause to speculate re I. Q., chance, etc., affecting the pretest scores. According to our previous reasoning--if these factors are minimized and if the "Morris Treatment" truly operates--the Ss of Experiment III should show the most substantial change-score gains. This proves to be the case. In short, the C. M. key is a less sensitive measure, but--by the same token--less influenced by factors that we cannot analyze and that we are not interested (at this time) in measuring; e.g., it is not too germane to this study to know that the pretest mean for the Ss of Experiment III (E's scoring) is significantly above the means for Experiments I and II. Returning to the C. M. data, the reader can ascertain with a glance that little is happening--beyond sampling, etc.--in Experiments I and II. However, in Experiment III a t computed for the posttest mean difference equals 1.868, which for 18 df reaches the .039 point (one tail of the distribution). E used the posttest scores because the difference between the variances was smaller than that for change scores and the pretest scores were not significantly different from chance. WHEN THE C. M. KEY is used, the Ss must have the "Morris Treatment" in order to show posttest gains. The factors of working and pretest scores are still significant; for example, in Experiment III the change-score mean for the workers is five points larger

than that for nonworkers. The C. M. key thus yields results in the same direction as E's key. An interesting side issue of the application of the C. M. key lies in the fact that the variances are larger for the experimental group ONLY when the treatment is effective; E has been told (and his limited reading in the area substantiates it) that this is a common finding when the effects of drugs are being tested; i.e., if a drug does not bring about the PREDICTED result, it--nevertheless--will decrease the variability of the EXPERIMENTAL group.

Table XIII gives an analysis of the same 8 experimental and 8 control Ss, using the C. M. scoring, who were used in E's "worker" analysis. The P value for these Ss, when E's scoring key was used, was .015 for one tail of the distribution. Again the groups were well matched on the two predictors, namely, pretest and time.

Next the original pool of 13 "Morris" workers and 10 "Others" was re-examined. The mean pretest score for both groups combined was determined, and only those Ss whose pretest scores (C. M. scoring) fell within the approximate range of plus and minus one standard deviation from the mean were used. (These conditions parallel those used in selecting the Ss for the previous analysis.) The results are presented in Table XIV.

A glance at Table XIV will show that the variances are quite different. (This has not been a major factor in the previous "worker" comparisons.) Therefore, this difference was tested and an F of 4.95 was obtained; this P is below the .05 point. In this test, even though the matching on pretest scores and time is not as precise as when E's groups were used and though both regression and time favor the "Morris Treatment"

TABLE XIII

THE C. M. SCORING KEY APPLIED TO E'S "WORKER" GROUP[a]

	Morris	Others
N	8	8
$\bar{X}_1$	15.375	15.62
$\bar{X}_2$	19.375	19.62
$\bar{X}_c$	4.875	-2.25
s_c^2	89.554	109.071
s_c	9.463	10.443

$t_{c_{2-1}}$ = 1.434

df = 14

P = .0875 (one tail)

[a] 1 = the pretest; 2 = time spent on posttest; c = change score

TABLE XIV

C. M. SCORING: "WORKERS" PLUS AND MINUS ONE STANDARD DEVIATION FROM THE MEAN OF THE PRETEST FOR THE COMBINED GROUPS[a]

	Morris	Others
N	10	6
$\bar{X}_1$	8.60	11.33
$\bar{X}_2$	20.60	18.83
$\bar{X}_c$	8.60	2.83
s_c^2	109.82	22.17
s_c	10.48	4.71

$t_{\bar{X}_{c_1}-\bar{X}_{c_2}} = 1.51$

df = 14

P = .077

$F_c = 4.954$

$n_1 = 9$

$n_2 = 5$

$P < .05$

[a] 1 = the pretest; 2 = time spent on posttest; c = change score

group, the P of .077 is about the same as the P of .088 which was obtained for the C. M. scoring of E's "worker" groups. In other words, workers are more important than judges' ratings (scoring keys) of paragraphs re demonstrating the effectiveness of the "Morris Treatment."

It will be recalled that one secondary purpose of this particular experiment was to determine what--if any--were the effects of items wherein the scoring of C. M. and E agreed completely. Of the 6 paragraphs used in Experiments I, II, and III, we were in complete agreement on only item No. 5, which we labeled "appraisive-evaluative" ("poetic" or "idealistic" discourse). Comparisons were made between this item and all items for Experiments I, II, and III, and thus the differences between paragraph No. 5 and the others are minimized. A tabulation comparing change scores of item No. 5 and others for the separate groups and experiments is given in Table XV. In three of the four comparisons of Table XV, item No. 5 shows a larger mean change score than the mean change score for all items for the "Morris Treatment" groups; in three of four comparisons for the "Others" groups, it gives a smaller mean change score than that for all items. All the difference scores favor the "Morris Treatment" group for item No. 5, while there is one reversal for all items. Thus it seems to be an effective item--particularly so in Experiment III. Thus item No. 5 is good; i.e., it shows the effect (change score) of the "Morris Treatment" more clearly than all items combined. For the "workers" it is almost three to one better (for the "Morris" group) and at the same time is a poorer than average item for the "Others" group. We have good reason to believe that the selection

TABLE XV

COMPARISONS BETWEEN ITEM NO. 3 AND ALL ITEMS FOR EXPERIMENTS I, II, AND III EMPLOYING E'S SCORING KEY[a]

		Morris	Others	$\bar{X}_{e_{M-O}}$
Experiment I:	N	10	18	
(Per item)	$\bar{X}_e$[b]	0.74	0.26	0.48
(Item No. 3)	$\bar{X}_e$	1.30	0.83	0.47
Experiment II:	N	6	17	
(Per item)	$\bar{X}_e$	0.33	0.26	0.07
(Item No. 3)	$\bar{X}_e$	-0.17	-0.35	0.18
Experiment III:	N	9	9	
(Per item)	$\bar{X}_e$	0.33	0.57	-0.24
(Item No. 3)	$\bar{X}_e$	4.78	-0.33	5.11
All "Workers":	N	15	10	
(Per item)	$\bar{X}_e$	1.56	0.10	1.46
(Item No. 3)	$\bar{X}_e$	4.23	-1.50	5.73

[a] O. M. and E agree re Item No. 3

[b] e = change score

of better paragraphs (paragraphs that are independently scored in exactly the same way by C. M. and E, for example) will yield experimental data that more clearly illustrate the effectiveness of training in the mode-use classificatory scheme.

In conclusion, it should be said that it is not possible to state how much of the disagreement in scoring (C. M. versus E) springs from the difficulty of the paragraphs themselves, how much from the unreliability of C. M. and E as judges, and how much from lack of precision in the mode-use system (or more accurately, E's interpretation of the system). C. M., on more than one occasion, has written to E of the need for quantification devices which could be used in evaluating samples of written or spoken signs. The writer's studies have only scratched the surface, and--like all first attempts--E can see now how the "Morris Treatment" might possibly be improved; several quasi quantifiers have been evolved (such as "frequency," "prominence," etc., that have been mentioned in Chapter II), and future research may prove them to be useful. However, the writer wants to repeat that our agreement in scoring the 12 paragraphs is high when the difficulty level, etc., of the items is considered.

Not only do the experiments described in this chapter indicate that the mode-use classification produces considerable interjudge reliability and that certain items are more reliable than others, but in addition it demonstrates that E's experimental results are not the result of E's scoring key. When C. M.'s key is applied to the same data, the P value for the multiple regression analysis (Experiment I) favors the

Morris group and reaches our significance point. The factors of working at the task and pretest score regression hold up for the C. M. scoring, further decreasing the probability that the importance of these factors is the result of sampling. Though rigorous methods to insure test reliability were not used, we know that at least item No. 5 was functioning well, and also that the agreement of two judges as to scoring of a given item improves the reliability of that item. An indirect compliment to the worth of the mode-use system can be derived from the fact that E's key showed the results of the "Morris Treatment" more clearly than C. M.'s own key in the "worker" analysis. C. M.'s system would seem to be communicable or interpersonal and not private or dependent on the author for its use. This seems to E a tremendous achievement, especially when one recalls the literature on interjudge differences in the scoring of essay exams, mathematics tests, OBJECTIVE TESTS, etc.; i.e., items seemingly straightforward and having face validity are scored in a very different way by different judges when working independently (80; 81, pp. 212-14). It should be recalled that E applied only the first and independent evaluations of both C. M. and E as scoring criteria. It is difficult to define the limits within which the agreement between the two keys shall be considered as "high," "medium," or "low." Our original agreement for types of discourse was only 25 per cent after correcting for chance, and no matter what the selected scoring plan, the data of this chapter will show that both keys pointed to the same sector of the matrix. In other words, C. M. and E had to be in high agreement as to what (types of discourse) a given paragraph was NOT!

CHAPTER VII

EXPERIMENT V: RESULTS AND DISCUSSION OF THE EFFECT OF THE PRETEST

Up to this point we have found that the "Morris Treatment" is effective at least for those Ss from the middle of the distribution of pretest scores who--according to our definition of working--work at the posttest. We know that working time and regression are important no matter whether the C. M. or E scoring key is used. Moreover, we are satisfied with the interjudge reliability in scoring--particularly so when the reading difficulty of the paragraphs is taken into consideration. We have some evidence that high intelligence raises the mean pretest score if E's scoring is used; but the C. M. key indicates that Ss with high intelligence and/or increased motivation do yield higher than average change scores. There is reason to believe that more reliable items would more clearly show the superiority of the experimental treatment. BUT we do NOT know the effect of the pretest on the experimental groups. Consequently, Experiment V was designed to test the effect of the pretest and as a replication which could take advantage of some of the information gleaned from Experiments I, II, and III, and thus strengthen some of the above surmises and accomplish this without the loss of Ss. In addition, it was hoped that a group could be tested in which motivation would not be as questionable as it was for the Ss of Experiments I and II--a group that would be a better sample from the target population in which we were--at this time--most interested. Finally the writer wished to avoid

any complications arising from between-group variations in mean pretest and mean working-time scores; if these distractors could be avoided, simple and straightforward tests of significance could be used.

The group selected for this test was made up of undergraduates (University of Tennessee) who were members of a second-quarter class in introductory psychology. This group had not been overtested, and the professor in charge of the class was available throughout the experiment. The presence of the teacher--it was felt--might motivate the Ss more highly by giving them the impression that they had something at stake in the experiment.

Of the 29 Ss who took part in this experiment, 11 were females and 18 were males. Intelligence, again, was not controlled beyond random sampling. The test consisted of only two items (paragraphs No. 5 and No. 8); these were two of the items where C. M. and E agreed completely as to the scoring. The two paragraphs represented very different types of discourse, the difficulty level of the vocabulary was not high, and one of the items was judged to be scientific discourse--in which E was most interested. E hoped that by the use of only two items--thus shortening the task--more of the Ss would work at the problem--at least for a period longer than two and one-half minutes per item which was the unit used to define working in Experiments I, II, and III. Since all of our experiments were conducted in classrooms and--except for Experiment III--during class periods, it may be that students budget their time in the usual fifty-minute blocks (though the nurses' class--Experiment II--was normally a one-hour-and-forty-minute session). If this should be the

case, there would be ample time for even the pretested group of Experiment V to complete the task within fifty minutes.

The Ss were assigned at random--but according to the time of arrival at the classroom--to one of two rooms by their professor. Only the "Morris Treatment" was used. One of the groups was to have the pretest and follow the procedure previously employed; the other group was not to be pretested. An outline of the design follows:

	Pretest	Treatment	Posttest
Experimental Group:	Yes	Morris	Yes
Control Group:	No	Morris	Yes

E was in charge of the group which had the pretest. Minor changes, mostly grammatical, had to be made in the test materials; the materials used in this experiment will be found in Appendix C. The revised titles for the types of discourse were employed. The procedure followed the same outline as that for the other experiments, but we will present again its main outline. After some introductory words, the pretest containing the instructions and the two test paragraphs were passed out by monitors. The Ss read the instructions while E read them aloud. Five minutes were alloted to the pretest; this number of minutes for two paragraphs compares with the fifteen minutes which were allowed for the experiments in which six paragraphs were used. Upon completion of the pretest, the posttest procedure (order of reading, timekeeping, etc.) was explained. Then the treatment, memory sheet, and posttest were passed out in the order mentioned. The Ss then read the treatment and the memory sheet before evaluating the posttest paragraphs. The starting and finishing time for

the posttest were recorded by each S using the clock on the front desk. Monitors were available for answering questions, etc. The longest any S worked was 50 minutes.

The second group--which did not have the pretest--was in charge of an assistant. He made some introductory remarks from an outline which E had prepared, and which dealt for the most part with generalities about types of discourse such as those which appear in the first paragraph of the revised pretest instruction sheet. Next he told about the order of reading and how the starting and finishing times were to be noted and recorded. Monitors then passed out the two booklets--the "Morris Treatment" and the posttest instructions and test paragraphs. Monitors were available to make certain that the Ss began by reading the treatment, and to answer questions re timekeeping, etc. Maximum time for any S in this group was 40 minutes.

First we shall present a summary comparison of the results for the two groups. This tabulation is given in Table XVI. (The complete data for this experiment will be found in Appendix C.)

The first statistical test of the posttest mean differences shows a clear superiority for the pretested group (P = .017 for both tails of the distribution). The variances are quite comparable. The pretest significantly enhanced posttest scores; the actual means differ by a factor of two. Recall that both groups, experimental and control, showed positive change scores in Experiments I, II, and III. Thus the conclusion that no one of the treatments depresses the change scores seems justified. We have good reason to believe that "learning-by-doing" was an important

TABLE XVI

EXPERIMENT V: SUMMARY OF THE RESULTS FOR BOTH GROUPS[a]

	Pretest	Nonpretest
N	15	14
$\bar{X}_0$	21.000	11.571
$\bar{X}_1$	7.400	
$\bar{X}_2$	7.733	9.071
$\bar{X}_c$	+13.600	
s_0^2	110.571	83.956

$\bar{X}$ Theoretical chance = 3.875

$\bar{X}_{0(p-n)}$ = 2.567

df = 27

P = .017 (two tails)

[a] 0 = posttest; 1 = pretest; 2 = posttest time; c = change score.

factor re the superiority of the pretested group. Also the factors of "warm-up," exercise, positive transfer, etc., probably are contributing. A variant on Thorndike's concept of "belonging" may well be operating. We know that psychological experiments often seem to be more than a bit silly to college sophomores. Perhaps the experience of taking the pretest makes the posttest seem more reasonable than it does to a nonpretested group.

The mean chance score per item is 1.9375 giving a total chance mean for the experiment of 3.875. Thus we do not need to compute a t for the pretested group because the t for this group versus the nonpretest group is significant. The group that was not given the pretest had a posttest mean 7.695 points above chance; for 13 df this t of 3.142 gives a P of .0013 for one tail of the distribution.

If the obtained mean pretest score is used as the best estimate of the mean for the nonpretest group, the estimated mean change score for this group is plus 4.171 (posttest mean 11.571 minus estimated mean pretest 7.400); this difference yields a t of 1.70 which for df 13 leads to a P value of .067 for one tail of the distribution. Even this t fulfills our criterion of significance. This may be compared with a t of 4.22 for the pretested group where the mean change score was plus 13.600; the one-tailed P for 14 df was .0005.

A simpler and perhaps more striking method of showing the effectiveness of the "Morris Treatment" is the one we used to show the relationship between the C. M. and E scoring keys. This is the method often followed in grading multiple-choice objective exams. A perfect

score equals 32 points per S. The chance score equals 3.875 per S:

$$\frac{\text{Obtained Score} - \text{Chance Score}}{\text{Perfect Score} - \text{Chance Score}} = \text{Corrected Total Score}$$

Using the above formula for the data of the separate groups in Experiment V gives the following results:

	Pretest Group	Nonpretest Group
N	15	14
Sum Perfect Score	480	448
Sum Chance Score	58.125	54.250
Sum Obtained Posttest Score	315	162
	$\frac{315 - 58.125}{480 - 58.125} = .609$	$\frac{162 - 54.250}{448 - 54.250} = .274$

The tabulation above indicates that the pretested Ss showed (as a group) 61 per cent improvement after the correction for chance. Expressed in another way, they obtained 61 per cent of the total possible credit points. If the sum of the scores on the pretest is substituted for the sum of chance scores, the pretested group of Experiment V shows a gain of 55 per cent. For purposes of comparison and using only item No. 3, the "Morris Treatment" group of Experiment I (N equals 10) gave a group improvement score of 8.3 per cent above chance.

If we ask the question whether or not the pretest mean is significantly greater than chance, the statistical answer is that the difference is not significant. (Pretest mean minus chance mean equals 3.525; the t is 1.00 and for df 14 the P for one tail is .167.) One might ask, "Still, the pretest mean score IS higher than chance; why all the fuss about regression towards the mean in the first three experiments? Isn't this an illusory factor?" The tabulation which follows will show that regression is still very much at work. If we divide the pretested group with

15 Ss into TOP, BOTTOM, and MIDDLE thirds, according to their pretest scores with five Ss for each third, the results are as follows:

		TOP	MIDDLE	BOTTOM
	N	5	5	5
(Pretest)	$\bar{X}_1$	15.0	6.6	0.6
(Change Score)	$\bar{X}_2$	6.8	11.8	22.2

Again—the larger the pretest mean score the smaller the gain or change score.

Did the shortened test produce longer posttest work-time per S? The answer to this question is yes also. The previous definition of working was more than two and one-half minutes per item. By this definition all but two of the 29 Ss were workers (more than five minutes), and even these two Ss worked exactly five minutes each. Interestingly enough each showed a mean change score of zero. (Complete details of this analysis will be found in Appendix C.) Thus it is unnecessary to carry out a worker analysis for this experiment.

However, working on the posttest beyond the defined amount of time does not seem to be helpful, for the nonpretest group averaged 9.071 per S, while the pretested group averaged 7.755 per S; this mean difference gives a t of 1.36, and for df 27 reaches the .19 level or the .095 point. Thus we can't be too sure that the observed difference in time is important. The fact remains that the pretested group, which had the smaller of the two mean scores for time, did produce the larger posttest scores.

Were there differences between the two items? There were differences, but only in the pretest scores, and each group was homogeneous as

the following summary will show:

		Pretest Group	Nonpretest Group
ITEM No. 1			
(Pretest)	$\bar{X}_1$	1.53	
(Posttest)	$\bar{X}_0$	10.07	5.71
ITEM No. 2			
(Pretest)	$\bar{X}_1$	5.87	
(Posttest)	$\bar{X}_0$	10.93	5.86

The pretest mean score for item No. 2 is considerably larger than the mean score for chance. Item No. 1 (paragraph No. 5 in Appendix C) is keyed as "idealistic" discourse; item No. 2 (paragraph No. 8) is scored "testable." Since the posttest scores (intragroup) are approximately the same for both items, we can only assume that the explanation of the difference lies in the fact that the title "testable" is the more suitable; i.e., it "fits" the paragraph better than "idealistic" fits its paragraph or else the word "testable" is better understood by the Ss. Be that as it may, future experiments--planned by E--will probably employ a matrix with blank center cells; i.e., the modes of signifying will be on the left side of the matrix and the four uses across the top, but the 16 cells which now contain the titles for the types of discourse will be left blank. The S will be asked to select the mode of signifying and the use for a given passage and then merely record the number of the passage in the appropriate cell of the matrix. E wanted to follow this procedure for these experiments, but the necessity of having control groups ruled out its use.

The writer in Chapter VI showed that different items were not

contributing equally to the change scores; some were better than others for showing the effectiveness of the "Morris Treatment." Specifically, it was demonstrated that item No. 3 (which is item No. 1 in Experiment V and keyed as "idealistic" discourse) was a better item than the average for all items INCLUDING No. 3; e.g., in the "worker" analysis the Morris group showed an average change score of 4.25 versus a change score of negative 1.50 for the Others group. We resume this analysis now by inquiring whether or not this item yields comparable results for Experiment V. In order that the groups be comparable, only the workers from Experiments II and III will be used, since 27 of the 29 Ss of Experiment V were workers by definition; the 5 workers from Experiment I were eliminated, because this experiment did not use the revised titles for the types of discourse. (This group's mean change score for workers is 2.60 versus 5.25 for Experiments II plus III; this may be further evidence that the revised titles were helpful.) Only the pretested group from Experiment V will be used. The data of Table XVII show that the two groups are rather comparable with respect to pretest, time spent on posttest, and change scores. The mean change score is somewhat larger for the Experiment V group. The additional time per item might account for this; it is more apt to spring from the fact that these Ss had a simpler task--two items versus six; another possibility is that the Ss from Experiments II and III had already used "idealistic" to identify one of the five other paragraphs--since they were instructed that no two types of discourse were alike in the test paragraphs. At least we cannot conclude that the Morris workers are drawn from a different

TABLE XVII

ITEM NO. 3 ONLY: COMPARISON OF THE WORKERS, EXPERIMENTS II AND III WITH PRETESTED GROUP OF EXPERIMENT V[a]

(Workers)	Experiment II (2 Ss) and III (6 Ss)	Experiment V (pretest only)
N	8	15
$\bar{X}_1$	1.38	1.53
$\bar{X}_c$	5.25	7.73
$\bar{X}_2$	3.40	3.87

[a] $\bar{X}_1$ = mean pretest

$\bar{X}_c$ = mean change score (posttest minus pretest)

$\bar{X}_2$ = time spent on posttest

population than that of the Experiment V Ss; i.e., the impressive change scores obtained in Experiment V were produced by work and more reliable items and NOT by the fact that the Ss were drawn from an elite population. What E assumed to be an elite population was the one from which the working Ss of Experiment III were drawn. Actually the means for this group are very comparable to those of Experiment V (pretest group). The 6 working Ss of Experiment III show a pretest mean of 1.67 for item No. 5 and a mean change score of 7.00. On the basis of these samples, we would conclude that--beyond a certain point--further increments of time spent on the posttest plus intelligence and/or academic training do NOT increase the change scores.

Parenthetically, the writer would like to add that when the papers for the pretested group are scored according to the original 1-1-4 score plan, the P value for the 1-4-16 plan is more significant--but not to a major degree. Again there is doubt that the Ss more readily grasped the concept of use. The differential in favor of use probably represents biased sampling from either the item or discourse-title universes. Any future experiments can employ the 1-1-4 scoring and--with the blank matrix--avoid any use of titles for the types of discourse. However, it should be repeated that our experimental results are IN NO WAY dependent--in their direction--upon the use of any specific scoring scheme.

It was felt that the reader might have received the impression that many loose threads were left dangling after the discussion of Experiments I, II, III, and IV; that is why the writer has tried to gather them in this chapter with the data of Experiment V at his disposal. However,

this gathering of loose threads may have caused the reader to lose sight of the main purpose of this experiment--which was to determine the effect of the pretest in relation to the "Morris Treatment." The data show that the average posttest score is twice as large for the group which had the pretest, and this difference leads to a highly significant P value for the t test. Moreover, the nonpretested group gave a mean posttest score significantly above both the theoretical chance mean AND the obtained pretest mean of the other group.

CHAPTER VIII

DISCUSSION

We have seen that very little experimentation has been directed towards meaning in any one of the many philosophical connotations of the term. Our studies indicate that college students can—after approximately 20 minutes' training--show great improvement over their pretest scores (if they work a reasonable amount of time) in labeling paragraphs in terms of the 16 main types of discourse as developed in Charles Morris' book, Signs, Language and Behavior. Moreover, if time spent on the posttest is held constant, the supposedly more intelligent Ss do not show appreciably greater change scores; higher intelligence, however, may be related to higher pretest scores. This finding may very probably support the hunch that most of us apply something very much like the mode-use classification whenever we read. But the modus operandi is doubtless not based on such well-defined (explicit) categories and is not as consciously employed as was the case with these students. The mode-use system, nevertheless, must be closely related to our habitual systems if its main outlines are to be so well mastered in so short a space of time.

One experiment needs to be done. (1) We should determine the effect of intensive training (treatment) in the application of the mode-use system (labeling paragraphs), and (2) this group should be compared with a "nothing" treatment group. If this test were made, we would not have to be concerned with either the effects of the pretest or the

nature of control-group treatment(s). This test would employ two groups matched on verbal fluency or verbal intelligence and amount of schooling. With good matching we would need very few Ss. The design would take the following form:

	Treatment	Test	N
Experimental Group	Yes (intensive)	Yes	10
Control Group	No	Yes	10

Projected experiments will require the use of a large pool of reliable paragraphs--preferably shorter than the ones used previously and covering all the main types of discourse. Judges selected from an English department, for example, could be trained in the mode-use method, combine this training with their literary sophistication, and then label certain paragraphs in mode-use terms. Only those items which showed high inter-judge reliabilities--like those of Experiment V--would be selected for further tests.

What do our present experiments "mean?" For one thing they indicate--whether the Ss know it or not--that Ss are aware of the significative conditions of the signs of the test paragraphs, and--for at least some of the signs--may be able to specify the conditions under which they denote. The writer is sophisticated enough to state that this is the meaning of meaning which his experiments attempted to elucidate; he is not naive enough to state that this is THE definition of meaning. He does feel that no matter what the definition of meaning, it should be clearly separated from verbal understanding. If one hears the words, "Brother, can you spare a dime?", the hearer's behavior will give the ready answer

as to whether or not he has understood the request, or we could ask him to write down what he had heard. Our version of meaning is more closely related to the everyday concept. The average financially-able man will not immediately respond positively to the above request. Instead some sort of cogitation will ensue; e. g., does this question mean that the man is broke, a panhandler, or what? Or in terms of mode-use, what sort of words made up this plea, and for what purpose were they used? The answers to these questions will partially determine the meaning--whether or not the man gets the money, if no questions are asked re the potential benefactor's personality, value system, and economic solvency. In addition, the mode-use system can be reversed; i. e., if agreement can be reached as to the signification and conditions of denotation for the dime-plea words, we can experimentally determine the hearer's economic state (income tax report), apply certain psychological tests, and infer certain aspects of his personality. Of course if he did not understand the request, he can if necessary be referred to an etologist or a psychiatrist. With no attempt at definitional rigor at this point, this area of meaning (which has just been outlined) is the one with which the writer is concerned.

Actually the term "meaning" is not basic. Attempts to make it precise are not often successful. A recent article illustrates this point. In referring to his own study, Noble writes:

> Various possible logical "meanings" of meaning--e. g., signification, denotation, connotation, equality, equivalence, definitional equivalence, material implication, strict implication--are not at issue here. Throughout the present analysis one must clearly and persistently distinguish between logical and psycho-

> logical notions about meaning; the former class of notions is conceptual (hypothetical), the latter empirical (categorical) (60, page 483).

The present writer submits that the above author's actual study does in fact investigate the logical notion of connotation except that it is defined in terms of the number of S-R connections established. Actually Noble counted the number of free associations of a specified class defined by set rules. But because a study derives from Hull's habit-strength postulation certainly does not rule out the possibility that it is "clearly and persistently" logical! In addition, the writer would assert that his experiments indicate that the signification and denotation of signs--within the framework of the Charles Morris semiotic--can be empirically investigated in a laboratory setting, realizing at the same time that his study is only a beginning. One possible implication of Noble's comments is that such a study as the present writer's could not be carried out.

One does not need to become embroiled in a definitional controversy; all that needs to be said is...this is the area of language and language behavior that is under investigation. And any investigation may be framed in terms of semiotic as well as learning theory. The present writer agrees with Charles Morris when he writes that

> Signs, Language and Behavior . . . was an attempt to show that semiotic could be given a biological foundation such that scientific methods could be applied to the whole domain of meaning. The suggestion was made that the analysis of any sign might be given by stating the things to which it directs behavior and the behavior which it prepares to these things. (55, page 57).

Indeed Professor Morris' semiotic--admittedly--owes much to the work of Clark Hull! We have merely brought the mode-use classification into an

experimental setting in an attempt to verify the statement that "scientific methods could be applied."

One might question the present results thus: "If your experiments do indicate that a large percentage of students can be taught to use this method in a short space of time, WHAT OF IT? For what is the system valid?" The frank answer to the above is simply that--as yet--we do not know. Certain data support various hunches--which have already been expressed, and we do have a quasi study of validity; at least Charles Morris and the writer agree to a reasonable extent in the use of it; i. e., the experimental scoring key predicts Charles Morris' responses to the test paragraphs considerably above the chance figures. The writer felt that any questions re the usefulness of the system would have to be postponed until he knew whether or not his summary treatment of it could be taught, used, and the results reliably scored. However, with the experimental data at hand, he would like to offer--for what they are worth--some of his speculations re the useful applications of the technique.

Many of the writer's hunches spring directly from the data. For example, recall that the pretest mean for the combined groups in Experiment III (our most educated group) was virtually twice as high for the experimental scoring key, as over and against the Charles Morris key. It was suggested previously that this finding might be the foundation for an experimental test of literary interpretation; or better, differences between an author's intended meaning and an interpretation of his writing. That is, if Charles Morris would prepare a summary of his mode-use classification, present it to experimental Ss, score their responses to the passages, and then have the present writer presumably use the same

system to score the paragraphs, WOULD THE SUBJECTS OBTAIN LARGER SCORES WITH THE AUTHOR'S KEY? If items of known reliability were used, we would have a measure of how well the Ss understood the technique, AND the comparison of the mean score of the subjects with the score of the present writer would be one measure of differences in interpretation (whether or not the writer's interpretation agrees more closely with that of the Ss or that of the author). If the writer's was the larger or if he and the Ss were in close agreement, the author could--if he chose--clarify the exposition of the system. If the writer's score was much smaller than that of the Ss, semantic pathology might be suspected; i. e., norms could be established for non-pathological mode-use behavior.

The following hypotheses follow less directly from the obtained data. Many of the tests require no training in the mode-use system. In the area of general psychology and learning theory, the method might tell us more of the organismic variables involved in language behavior; i. e., we could measure a different response from those usually recorded. (The method might prove useful in validating attitude and opinion polls.) An important concept in Charles Morris' semiotic is that concerned with post-language symbols (see the Glossary). These symbols should be sharply differentiated from signs per se; they are producible by their interpreters. They may be interpersonal or personal, and they may become elements in a language or not. In short, post-language symbols approximate very closely Watson's subvocal talking. Let's illustrate: Once, just prior to presenting himself for an oral exam, the writer was greeted with the words, "We are going to grill you." The straightforward S-R

measurement would indicate that the words were understood. The writer subjectively classified the discourse as being in the designative mode of signifying and named the use valuative AND, therefore, fictive or fictional. The post-language symbols were something like this: "I have had a considerable number of pleasant discussions with Professor X and did very creditable work in one of his courses. His behavior towards me has never been anything but friendly and helpful. He has gone out of his way to assist me with any number of academic problems. Therefore, he and the other professors will ask me a number of questions, but they won't flip me into an intellectual frying pan." What of the person who interprets the grilling statement as scientific discourse or--worse--as the gospel (religious discourse)? The writer fears that--if this were the case--anxiety would run high. Thus one straightforward experiment would be to determine what a S thought of certain passages in terms of types of discourse and then determine the behavioral correlates. Will the person who interprets the statement above as "scientific" fact, in terms of a type-of-discourse answer, show any increase in heart-rate, PGR, etc., in comparison with the person who labeled the statement "fictive?" Here we are getting at--however indirectly--individual differences; we are measuring something more than communication--in any usual interpretation of this term. The use of this or a similar method might go a long way towards explaining how, for example, parents, whose verbal behavior may seem permissive and kind to a clinician, may seem threatening to their child. Obviously the direct opposite of this follows also. (Later the writer will develop this concept a bit more--in terms of his "metasign

multipliers.") Of course the test suggested above could be hypothecated in terms of motivation as well as learning, or cast in the form of a memory, thinking, or feeling experiment.

The mode-use method would seem to lend itself to many studies in social psychology. For example, one could determine the role of differing types of discourse in the breakdown or forwarding of group communication in a specified social situation. E could determine if different socio-economic backgrounds lead to different type-of-discourse preferences. The mode-use method in no way obviates the necessity for using Flesch counts, etc., to determine readability (understanding, in the present terms) and for investigating the information-gathering problem.

The applied psychologist might find the system useful in aptitude testing and selection; e. g., does the successful engineering student prefer different types of discourse to those preferred by the failing student or a student of political science? Educational advisers could approach the curriculum with the view of making certain that a student is exposed to several types of discourse. Despite his having obtained an A. B. degree, the writer feels that he would have been more liberally educated if he had been more intimately associated with scientific discourse.

The method might be used to measure certain aspects of intelligence. The writer submitted three rank-orders of six graduate students to three judges who were familiar with the students' work. One rank-order was a random affair; the second was derived from scores on a well-known (high-level) intelligence test; and the third consisted of the students' scores on the experimental pretest. Each judge--with no knowledge of the criterial basis

of the rank-orders--thought that the pretest order was the most accurate with the intelligence test scores second. Ability to successfully use (after training) the mode-use system might indicate yet another aspect of verbal intelligence.

The clinical psychologist might be interested in analysing the types of discourse favored by neurotics or psychotics. The catatonic schizophrenic might demonstrate a loss of confidence in any type of discourse; the hebephrenic--a preference for his own language; and the paranoid--serious deviations only in his post-language symbols. Moreover, he could evaluate changes in discourse given by the improving patient; or differences re discourse between successful and unsuccessful therapists; or between different therapeutic procedures. What type of discourse is "umh, huh"? Certainly it is a noise made by even the most non-committal of the non-directive therapists. In our terms the Rogerian theory holds that formative (non-lexicative) discourse forwards the therapeutic process. If mode-use methodology is followed, lexicative versus non-lexicative discourse questions can be experimentally answered. TAT protocols could be scored in terms of types of discourse. If a mode-use intelligence test were used, we might be able to determine the type-of-discourse profiles of those who are most prone to emotional illness (mental hygiene).

All but one of the experimental possibilities previously mentioned do not require that the person tested have any training in semiotic. If the mode-use method is potentially valid for certain experimental purposes, may it not also be valid for the Ss as well? What would training in analysing types of discourse DO FOR a man--child, student, or clinical

patient? It would be interesting to determine the earliest age at which children could make mode-use distinctions. Training in semiotic may be a reinstatement of the formal-discipline concept but with at least one difference; students would not study a particular subject matter to "train their minds," but would add mode-use analyses to whatever subjects made up the curriculum. The writer has found that--if the matrix is kept in mind--it is possible to write a more precise type of discourse, though he has not tried to fill in every cell of the matrix. However, to the extent that--say--the writing of scientific discourse is a more difficult task than the creation of moral discourse, the more imperative is the use of some extragrammatical scheme. To advocate such a method is not to invalidate or render obsolete the suggestions for scientific writing previously made by Boring (9), Hebb and Bindra (35), Stevens and Stone (82), and others. It is possible that semiotic training would help therapists write more precise case histories, etc. Though the individual's post-language symbols or interpretants (dispositions to respond) are often private, this does not make the study of them impossible. Moreover,

> . . . to have scientific knowledge of something does not require that each individual be able to respond directly to the something in question, but merely that he be able to obtain evidence that statements which are made about it do denote (54, page 229).

Training in semiotic might assist in keeping men somewhere near the center of a linguistic distribution which is anchored at one end by paranoia, with gullibility at the other. Such training might rescue us from the grasp of intellectual as well as economic hucksters. The mode-use analysis might also cast new light on the "great books" and "great speeches" --past and present--and indicate usable differences between those that

changed behavior and those that did not.

Serious doubts have been expressed about our system of public education--some stressing its academic failures, some decrying its psycho-social invalids. The classical mold of American democracy must turn out citizens and not semantic wolves and lambs. Charles Morris' "democratic commitment" should force all educators to give their students a critical appreciation of all types of discourse. The writer has often toyed with the idea of composing an essay on John Dewey; the thesis would be that the confused educational theory often attributed to him results from the mistaken belief that he wrote only one type of discourse. His popular, humanistic articles were designed to alter educational values. The present writer, who has read Dewey's journal articles primarily, would often have no idea just what person was being discussed if the "Deweyites" did not use the master's name. I hasten to add that certain disciples have not altered my respect for Dewey's work.

Throughout this discussion we have suggested that semiotic training might be of help to the patients of a psychological clinic. Little empirical evidence supports this belief. Still language behavior is behavior; and to study linguistic productions is only to study the effect of this on other behaviors. Many patients would seem to be exhausted by the search for denotational chimeras; if the signification of the word "success" could be more adequately specified, they could better know its denotata; they might find that they had already achieved it. Other patients are oppressed because--to them--every statement is in the desig-

native mode of signifying.

The writer repeats that he is aware of the limitations of his experimental data; he has tried but to sketch some of the propositions that seem to be open for empirical test. In presenting these hypotheses, he may have given the impression that Charles Morris' semiotic metalanguage is either the best or the only one; this is not known to be the case. Feigl (19, p. 7), for example, classifies language functionally into three cognitive and three non-cognitive meanings; one could experiment with these as six types of discourse or combine them with the four modes of signifying and test the resultant twenty-four types of discourse. The writer is content if this thesis stands as evidence for the fact that something does go on in terms of types of discourse and that it is an experimentally testable something. He is neither proclaiming the semiotic to be used nor specifying the methodology.

The writer has done some preliminary thinking about a different mode-use classification, which involves an alteration of the Charles Morris matrix. The classification is an attempt to make explicit (but in no way to unduly limit) the conditions which scientific discourse must satisfy. The writer's revised matrix is given in Figure 3.

In examining Figure 3, note that the formative mode of signifying has been divided. We have previously mentioned the experimental difficulties with this mode and the writer's differences with Charles Morris re its interpretation. Feigl writes thus of the philosophical analysis of language meaning:

> . . . logical analysis is usually directed analysis. That is to say, it is either postulational codification (as in

USE / MODE	Informative	Valuative	Incitive	Positivistic	Logical	General Purpose
Designative	X			X	X	
Appraisive						
Prescriptive						
Epistemological				Thorndike	Wm. James	
Postulational				Skinner	Hull	
Mixed						

Figure 3

Projected Mode-Use Matrix

> the mathematical and the exact empirical sciences) or epistemological reduction (the reconstruction of factual terms and propositions on a basis of observational evidence) (19, page 8).

This division adds nothing new to the original matrix. The additional mode ("mixed") is inserted to take care of everyday discourse in which humor, fancy, and fact are interwined in such a way that the primary mode of signifying cannot be determined. The "positivistic" and "logical" uses correspond to the epistemological and postulational modes, while "general purpose" is related to "mixed." In this brief summary no attempt to label the 36 cells of the matrix will be made. Examples of mixed discourse are not hard to find; these types comprise the bulk of our everyday behavior.

The writer has entered the names of four psychologists in cells of the matrix to exemplify these types of discourse. When Thorndike (87) summarizes his work and spells out its experimental inferences, he is using epistemological-positivistic discourse. William James's numerous books draw out philosophical implications and thus would be classified epistemological-logical. Skinner's (73) Behavior of Organisms--while being grounded in the literature of conditioning--does not reconstruct factual terms; it offers new primitive terms or axioms for a different psychological frame of reference. This is done with the view to constructing an experimentally fruitful theory (postulational positivistic). Hull's (38) Principles of Behavior, despite certain differences, would fall in this same cell; but when his theory is expressed in the form of symbolic logic (39), it would be evaluated as "postulational-logical" discourse. The specific works of the four investigators just mentioned would repre-

sent four types of scientific discourse. There would be three other types (marked "X" in the matrix), namely, the designative-informative straightforward experimental reporting found in most scientific journals; the designative-positivistic discourse used in presenting experimental summaries and conclusions; and the designative-logical type which might be exemplified by any implications-for-future-research discussion. Thus--in all--we would have seven types of scientific discourse, all equally valid but not all valid for the same purpose. Moreover, a reader or student should be entitled to request that a given author translate certain concepts from one type of discourse to another (or else be permitted to make his own translation) so that experimental or theoretical implications could be more precise and interpersonal. This sort of linguistic manipulation might do much to solve the riddle posed by experimental rats "acting like" the experimenter!

Something like the above is implicit in the instructions offered by the manual on publication of the American Psychological Association (99). The request to avoid italics, exclamation points, etc., is made; use of such devices--which try to represent spoken language--is interpreted as scientific immaturity. The editors are requesting authors to use an acceptable form of scientific discourse. This is a very reasonable request but to refrain from the use of certain devices does not guarantee the production of scientific communication.

In discussing post-language symbols, the writer mentioned his conception of a metasign analysis. Once a given unit has been evaluated in terms of mode and use, one should retain the "message" and apply one of

the following metasign multipliers: GAIN, PRESTIGE, CREATIVITY, SOCIOLOGICAL, PSYCHOLOGICAL. The reader's (or hearer's) entire background with all the attendant dangers of projection will enter into this evaluation. The best he can do is be conscious of this for he in fact actually employs something very like the multipliers. (The reader will note the similarity between the multipliers and context.) The multipliers are not in themselves value judgments but rather aids in estimating value. For example, a given unit of scientific discourse is best evaluated by a search for denotata or an experimental replication. If we say "X" did this to make money (GAIN), the need for verification may be imperative. "Prestige" is not as demanding a multiplier, neither is the joy of intellectual manipulation ("creativity"), but such scientific productivity may have to be translated by the author or someone else into designative-informative terms. The "sociological" multiplier is obvious in its implications. This discourse might be used at a wedding reception, polite conversation, etc. The "psychological" multiplier handles questions of personality and temperament, distinguishes humor from fact, and science from fiction.

In conclusion, no matter how completely the structure, content, and meaning aspects of language are specified, we remain forever stuck with the philosophical evaluations that are implicit in the very structure of our own language. We can be conscious at least of this logical structure if--following Whorf-- we remind ourselves quite often that it is the EMPTY gasoline drum that explodes, the empty words of future Hitlers that lead-- not only to failures in communication--but to war!

CHAPTER IX

SUMMARY AND CONCLUSIONS

In reviewing the literature concerned with language and language behavior, we found a considerable number of provocative philosophical and logical analyses of sign behavior. A certain amount of experimentation had been aimed at linguistic learning (acquisition), the mechanics and content of communication, and propaganda analysis. The phonemic and morphemic structure of language was being carefully investigated. In the realm of meaning--as defined by the writer in terms of the connotational and denotational characteristics of language (signification)--very little experimentation had been carried out, and this medium of empirical investigation dealt with very restricted units of discourse (a single word, nonsense syllables, etc.). The majority of the laboratory studies assumed the form, "If you can carry out the instructions, you have understood; you know the meaning of the given unit of discourse." Experimentation with syllogisms does approach more closely the writer's definition of connotation. But three articles in the most recent Psychological Review--including one titled "An Analysis of Meaning"--are deceptive unless one is willing to accept a very restricted and/or theoretically derived definition of meaning.

Moreover, few of the analytic, linguistic postulations seemed--to this writer--to lend themselves to experimental treatment. But one of the semiotic analyses impressed the writer as being behaviorally grounded and designed to make laboratory studies of its chief axioms possible; this

was the semiotic, or science of signs and sign behavior, of Charles Morris. In addition, Professor Morris' fully developed semantic meta-language was readily available in his book, Signs, Language and Behavior. The writer was primarily interested in testing the teachability and reliability of the core of Charles Morris' system, namely, the four modes of signifying and the four uses. Re modes of signifying, signs are differentiated in terms of that event, object, or characteristic which would serve as evidence that a given sign denotes. For example, the word "dog"--in its usual connotation--symbolizes a furry quadruped that barks, etc., etc. If one is able to identify a representative of this species when some interpreter uses the word, the actual dog would indicate (denote or designate) that the word "dog" does indeed signify in an interpersonal way; i. e., the word would have designative significance or --the word would be in the designative mode of signifying. There are four modes of signifying--designative, appraisive, prescriptive, and formative--each of which is determined in a manner similar to the designative example above. There are four uses--informative, valuative, incitive, and systemic. The uses are not directly related to the words themselves but rather to the purposes to which they are put; i. e., the interpreter of a sign might ask himself the question, "What purpose do these words or signs serve? Do they give information, place a value on something, exhort some action, or merely organize future behavior?" The answer to the four queries above would determine which of the uses was being stressed.

A summary explanation of the four modes of signifying and the four

main uses was prepared--the "Morris Treatment." This abstract explained the above concepts and gave instructions as to their value in determining types of discourse--a type of discourse being roughly defined as a specialization of language for the accomplishment of a given purpose.

Two control treatments, which paralleled the "Morris Treatment" re number of words and sentences, were devised: The "Feigl Treatment"--an abstract of an article on operationism by Herbert Feigl--and the "Osgood Treatment"--a digest of this author's article which dealt with an investigation of the connotational meaning of single words or concepts.

The design of the first three experiments was a modification of the classical transfer-of-training paradigm, in which subjects were given a pretest--consisting of six paragraphs of approximately 120 words each--and asked to determine the type of discourse for each of the six items. The paragraphs were taken from writings of psychological import. From a pool of 97 paragraphs, 12 paragraphs were finally selected by a crude paired-comparison technique. Then the subjects were randomly assigned to one of the three treatment groups. They read the assigned treatment, and --upon finishing--took the posttest, which consisted of the same paragraphs used in the pretest.

Experiment I followed the above design. Experiment II used this design with one main procedural change--the names of the types of discourse were revised and more abstract terms were purposely substituted for the original titles. The Osgood group was divided (Experiments I and II) into two parts for the posttest; one half received a four-by-four matrix dealing with the modes and uses--which was also employed by the "Morris Treat-

ment" groups. Experiment III, designed to test the role of (assumed) high intelligence and (known) long academic training, was similar to Experiment II except for the fact that only one control group ("Feigl Treatment") was used. It had been established previously that there were no significant differences between the two control treatments.

Preliminary analysis of the results of the first three experiments were suggestive, but equivocal. The t tests of mean change score (posttest score minus score on pretest) differences were not significant, but two of the three t's were positive (favoring the experimental group). Multiple regression analyses of covariance--with two predictors (pretest score and time spent on the posttest) and the criterion of posttest score--were not significant and actually were diametrically opposed to the direction of the mean change score t's. Intragroup coefficients of correlation indicated that the predictors were effective only for the "Morris Treatment" groups. Pretest scores were confounded by regression towards the mean for both groups and seemed to be correlated with intelligence. Thus subjects with high pretest scores could show at best only a small gain in the posttest. The regression line based on the pretest seemed to be nonlinear. Posttest working time seemed to show approximate linearity but with an upward regression line for the experimental groups and a downward line for the control groups. Instead of trying to correct for nonlinearity, the writer defined "working" as 16 minutes spent on the posttest (median time of the Morris groups or mean time for all groups). With this criterion 13 experimental workers and 10 control workers were obtained. The mean pretest for these 23 workers was

determined, and cutting points of plus and minus one standard deviation from this mean gave 8 experimental and 8 control subjects. The means for the two groups showed good matching re posttest working time and pretest score. Parametric and nonparametric tests for both mean posttest score differences and mean change score differences yielded highly significant P values. Mean change scores were more than three times as great for the Morris group.

Experiment IV was concerned with the agreement between Charles Morris and the writer in scoring 12 paragraphs--6 of which were those used in the first three experiments. This experiment indicated substantial interrater reliability in scoring the 12 items. Since agreement was not perfect, the test papers were scored according to the Charles Morris evaluation. The results were in the same direction as those obtained with the writer's scoring key. Pretest scores, however, were depressed for all groups when Professor Morris' key was used. This did lead to certain differences; e. g., the change-score mean difference for Experiment III was highly significant and favored the experimental group. When the writer's key was applied to this data, the mean change score was very slightly larger for the control group, and pretest means were very large. However, this experiment showed (1) that the mode-use system could produce rather high agreement between judges and (2) that the results of the previous worker analyses, etc., were not the result of a particular scoring key.

Experiment V was aimed at testing the effect of the pretest on posttest scores ("Morris Treatment" only). Two items were used, and

Charles Morris and the writer agreed completely in scoring these two paragraphs. This shortened the task for the subjects while giving the promise of being a more reliable test. The posttest scores for the group which had the pretest were larger by a factor of two; the P for this difference was quite small. Moreover, the nonpretested group mean change score was significantly above both chance and the mean pretest score obtained by the experimental group. Regression towards the mean was still a factor in determining the change score. Only 2 of the 29 subjects in this experiment were not workers by our definition.

In sum then the experiments indicated that an abstract of the Charles Morris semiotic could be understood rather well after only 20 minutes of training and employed by college students with considerable success if they spent more than a specified amount of time on the posttest. When subjects (Morris groups only) were equated on the working-time variable, differences in mean posttest or change scores for two reliable items were small and not statistically significant.

These experiments have barely scratched the surface. There remains the task of intensively training subjects in the use of this system for analyzing types of discourse and then testing them in conjunction with a comparable control group who have had no training of any kind. The system should be tested with a wider range of written material and with different age, intelligence, and educational ranges. Since Charles Morris and the writer are not in complete agreement as to the interpretation of formative discourse, attempts to improve the present system of classification and/or to develop a new system should be made.

As yet we know nothing as to the usefulness of learning to classify

linguistic material into categories based on the 16 main types of discourse. Is this a meaningful aspect of meaning? If a person can detect the differences in signification between scientific writing and fiction, will he actually use the system in his daily life and would its use make any behavioral difference? We know that the present method IS teachable. If a pool of reliable items covering a wide range of subject matter were available, could these items be used in testing aptitudes, in differentiating neurotics and psychotics? What sort of discourse does the schizophrenic prefer--and which type or types does he use and how often--in speaking and writing? What type of discourse do Rogerian therapists use? Do successful psychological therapists use a type of discourse that is consistently different from that of nonsuccessful therapists? Are there semantic differences among clients of a psychological clinic that would be helpful in diagnosis and in predicting the outcome of therapy? These and many other similar questions are open to experimental test. Most of the problems outlined above have to do with the labeling of written or spoken material or the stating of preferences. Preference testing would be valuable clinically and psychometrically, and the practical implications are obvious. For example, the <u>same</u> short message could be written in several types of discourse; subjects could be asked to check which they liked best, etc., and/or asked to answer questions designed to test their understanding of the message. Numerous validity studies could be made employing the data. All of these problems--of course--would require the construction of a reliable test.

It must be understood that the writer is NOT suggesting that his studies prove the possibility of obtaining answers to all of these

questions. Neither is he saying that the semiotic of Charles Morris is the only or even the best system. He does believe that his studies open up the territory.

Still another wide area is open for investigation, namely, what behavioral changes result from intensive training in and practice with the mode-use system? All questions of the reliability and worth of the method aside, would such a course train students to think critically about the nature of words and sign behavior--about the different sorts of words, who uses them, and for what purposes? And would this training put us all on our mettle and eventually lead, in fact, to individual and social sanity and science?

Since our studies have dealt with signification and the denotation of signs and not with the denotational world, we should add that any "how-is-it-with-the-world questions" are the special concern of general semanticists and experimenters. In the intimate relationship of language and life, one is constantly hung on either the significative or denotational horn of this dilemma. The writer conceives of this predicament in terms of three types of error--"the unholy three." There are those who think that the word IS the thing (Error I); i. e., their semantic matrix has but one cell. Many achieve linguistic sophistication, but they feel that one cell of the matrix SHOULD be the only cell, and/or that the conditions of signification for the terms of their own area are so well verified that these words are somehow different from other words (Error II). Many become so hostile or semantically confused by verbiage that they feel no word denotes--no matter how thoroughly its conditions of

signification have been specified; they may even become mute or develop their own language--schizophrenia (Error III). No one of us can avoid all of these errors all of the time. Perhaps the best we can do is achieve some dynamic equilibrium by use of a semiotic-empirical dialectic. That is, if you feel that "the word" has NO meaning, turn to semiotic for help; if you know that "the word" means EVERYTHING, turn about and say to yourself, "How is it with people--how is it with the world?"

BIBLIOGRAPHY

BIBLIOGRAPHY

1. Adler, M. How to read a book. New York: Simon and Schuster, 1940.

2. Ayer, A. J. Language, truth and logic. London: Victor Gollancz, 1936.

3. Baldwin, A. L. Personal structure analysis: A statistical method for investigating the single personality. J. abnorm. soc. Psychol., 1942, 37, 163-183.

4. Bavelas, A. Communication patterns in task-oriented groups. J. acoust. Soc. Amer., 1950, 22, 725-730.

5. Bloomfield, L. Language. New York: Holt, 1933.

6. Boder, D. P. The adjective-verb quotient; a contribution to the psychology of language. Psychol. Rec., 1940, 3, 309-343.

7. Bontrager, O. R. Letter to a teacher of reading. ETC., 1944-1945, 2, 126-128. (P. 128, quoting Adolf Hitler, Mein Kampf.)

8. Boring, E. G. A history of experimental psychology. New York: Appleton-Century Co., 1929.

9. Boring, E. G. Another note on scientific writing. Science, 1936, 84, 457-459.

10. Bridgman, P. W. Logic of modern physics. New York: Macmillan, 1927.

11. Burke, K. A grammar of motives. New York: Prentice-Hall, 1945.

12. Burke, K. A rhetoric of motives. New York: Prentice-Hall, 1950.

13. Cameron, N. Experimental analysis of schizophrenic thinking. In (Various), Language and thought in schizophrenia: proceedings of the round table on schizophrenia held at the meeting of the American Psychiatric Association in Chicago, May 1939. Berkeley: Univer. of California Press, 1943.

14. Carnap, R. Introduction to semantics. Cambridge, Mass.: Harvard Univer. Press, 1942.

15. Carroll, J. B., et al. Report and recommendations of the interdisciplinary summer seminar in psychology and linguistics. Unpublished manuscript. Ithaca, N. Y.: Cornell Univer., August 10, 1951.

16. Cobb, S. Personality as affected by lesions of the brain. In J. McV. Hunt (Ed.), Personality and the behavior disorders. Vol. 1. New York: Ronald, 1944. Pp. 550-581.

17. Eaton, Helen S. Semantic frequency list for English, French, German, and Spanish. Chicago: Univer. of Chicago Press, 1940.

18. Feigl, H. Operationism and scientific method. Psychol. Rev., 1945, 52, 250-260.

19. Feigl, H. Logical empiricism. In H. Feigl and W. Sellars (Eds.), Readings in philosophical analysis. New York: Appleton-Century-Crofts, 1949. Pp. 3-26.

20. Feigl, H., and Sellars, W. (Eds.) Readings in philosophical analysis. New York: Appleton-Century-Crofts, 1949.

21. Flesch, R. The art of plain talk. New York: Harper, 1946.

22. Flesch, R. A new readability yardstick. J. appl. Psychol., 1948, 32, 221-233.

23. Flesch, R. Psychological writing, easy and hard for whom? Further comment on the "peak stress" method. Amer. Psychologist, 1952, 7, 593-594.

24. Freedman, R. Incomplete matching in ex post facto studies. Amer. J. Sociol., 1950, 55, 485-487.

25. Freud, S. New introductory lectures on psycho-analysis. (Trans. by W. J. H. Sprott.) New York: Norton, 1933.

26. Freud, S. The basic writings of . . . . (Trans. by A. A. Brill.) New York: Modern Library, 1938.

27. Gagne, R. M., Foster, Harriet, and Crowley, Miriam E. The measurement of transfer of training. Psychol. Bull., 1948, 45, 97-130.

28. Gardiner, A. H. The theory of speech and language. Oxford: The Clarendon Press, 1932.

29. Gesell, A., and Ilg, Frances L. Child development. New York: Harper, 1949.

30. Goldstein, K. L'Analyse de l'aphasie et l'etude de l'essence du langage. J. de Psychol., 1933, 30, 430-496.

31. Harris, Z. S. Methods in structural linguistics. Chicago: Univer. of Chicago Press, 1951.

32. Hartmann, G. W. Propaganda. In D. H. Fryer and E. R. Henry (Eds.), Handbook of applied psychology. Vol. 1. New York: Rinehart, 1950. Pp. 23-30.

33. Hayakawa, S. I. Semantics, general semantics: An attempt at definition. ETC., 1944-1945, 2, 116-120.

34. Hayakawa, S. I. Language in thought and action. New York: Harcourt Brace, 1949.

35. Hebb, D. O., and Bindra, D. Scientific writing and the general problem of communication. Amer. Psychologist, 1952, 7, 569-573.

36. Heidbreder, Edna. Seven psychologies. New York: Appleton-Century-Crofts, 1933.

37. Hovland, C. I., Lumsdaine, A. A., and Sheffield, F. D. Studies in social psychology in World War II. Vol. 3. Experiments on mass communication. Princeton, N. J.: Princeton Univer. Press, 1949.

38. Hull, C. L. Principles of behavior. New York: Appleton-Century-Crofts, 1943.

39. Hull, C. L., Hovland, C. I., Ross, R. T., Hall, M., Perkins, D. T., and Fitch, F. B. Mathematico-deductive theory of rote learning. New Haven: Yale Univer. Press, 1940.

40. Jørgensen, J. The development of logical empiricism. Vol. 2, No. 9, International Encyclopedia of Unified Science. Chicago: Univer. of Chicago Press, 1951.

41. Johnson, W. People in quandaries. New York: Harper, 1946.

42. Koch, S. Review of Theoretical foundations of psychology (Harry Helson, Ed.). Psychol. Bull., 1952, 49, 349-355.

43. Korzybski, A. Science and sanity. (2nd Ed.) Lancaster, Pa.: Science Press, 1941.

44. Lasswell, H. D. The analysis of political behaviour. London: Routledge and Kegan Paul, 1948.

45. Lasswell, H. D., Leites, N. C., et al. Language of politics. New York: Stewart, 1949.

46. McGranahan, D. V. The psychology of language. Psychol. Bull., 1936, 33, 178-216.

47. Maier, N. R. F. Reasoning in humans: I. On direction. J. comp. Psychol., 1930, 10, 115-143.

48. Malinowski, B. The problem of meaning in primitive languages. Supplement in C. K. Ogden and I. A. Richards, The meaning of meaning. (Rev. Ed.) New York: Harcourt Brace, 1947.

49. Maloney, M. "Communication" and the arts. ETC., 1952, 9, 263-269. (P. 267, quoting Nevill Coghill.)

50. Masserman, J. H. Behavior and neurosis. Chicago: Univer. of Chicago Press, 1943.

51. Mead, G. H. Mind, self and society. Chicago: Univer. of Chicago Press, 1934.

52. Miller, G. A. Language and communication. New York: McGraw-Hill, 1951.

53. Miller, G. A. Speech and language. In S. S. Stevens (Ed.), Handbook of experimental psychology. New York: Wiley, 1951. Pp. 789-810.

54. Morris, C. Signs, language and behavior. New York: Prentice-Hall, 1946.

55. Morris, C. The science of man and unified science. In Contributions to the analysis and synthesis of knowledge. Proc. Amer. Acad. Arts Sci., 1951, 80 (1), 37-44.

56. Morris, C. Personal communication. 1952.

57. Murphy, G. Historical introduction to modern psychology. (Rev. Ed.) New York: Harcourt Brace, 1949.

58. Neurath, O. Einheitswissenschaft und Psychologie. Wien: Gerold and Co., 1933.

59. Neurath, O. Foundations of the social sciences. Vol. 2, No. 1, International Encyclopedia of Unified Science. Chicago: Univer. of Chicago Press, 1944.

60. Noble, C. E. An analysis of meaning. Psychol. Rev., 1952, 59, 421-430.

61. Ogden, C. K., and Richards, I. A. The meaning of meaning. (Rev. Ed.) New York: Harcourt Brace, 1947.

62. Osgood, C. E. The nature and measurement of meaning. Psychol. Bull., 1952, 49, 197-237.

63. Osgood, C. E. Review of Language and communication (George A. Miller).

64. Peirce, C. S. Collected papers of . . . . Vol. 5. Pragmatism and pragmaticism. Cambridge, Mass.: Harvard Univer. Press, 1934.

65. Piaget, J. Children's philosophies. In C. Murchison (Ed.), A handbook of child psychology. (2nd Ed.) Worcester, Mass.: Clark Univer. Press, 1933. Pp. 534-547.

66. Quine, W. V. Semantics and abstract objects. In Contributions to the analysis and synthesis of knowledge. Proc. Amer. Acad. Arts Sci., 1951, 80 (1), 90-96.

67. Richards, I. A. How to read a page. New York: Norton, 1942.

68. Sapir, E. Speech as a personality trait. Amer. J. Sociol., 1927, 32, 892-905.

69. Sapir, E. Language. In Encyclopedia of the social sciences, Vol. 9. New York: Macmillan, 1933. Pp. 155-169.

70. Sapir, E. Nootka texts. Philadelphia: Linguistic Society of America, Univer. of Pennsylvania Press, 1939.

71. Shannon, C. E., and Weaver, W. The mathematical theory of communication. Urbana: Univer. of Illinois Press, 1949.

72. Shen, E. A generalized formula for testing the significance of experimental treatments. Harvard educ. Rev., 1940, 10, 70-75.

73. Skinner, B. F. The behavior of organisms. New York: Appleton-Century, 1938.

74. Skinner, B. F. The alliteration in Shakespeare's sonnets: A study in literary behavior. Psychol. Rec., 1939, 3, 186-192.

75. Skinner, B. F. A quantitative estimate of certain types of sound-patterning in poetry. Amer. J. Psychol., 1941, 54, 64-79.

76. Skinner, B. F. Verbal behavior. William James Lectures, Harvard University, 1947. (To be published by the Harvard Univer. Press.)

77. Smith, H. L., Jr., and Trager, G. L. Foreword. ETC., 1952, 9, 163-164.

78. Snedecor, G. W. Statistical methods. (4th Ed.) Ames, Ia.: Iowa State College Press, 1946.

79. Solomon, H. L. An extension of control group design. Psychol. Bull., 1949, 46, 137-150.

80. Stalnaker, J. M. Question VI: The essay. English J. (Coll. Ed.), 1933, 22, 561-567.

81. Starch, D., Stanton, Hazel M., and Koerth, Wilhelmine. Psychology in education. New York: Appleton-Century, 1941.

82. Stevens, S. S., and Stone, Geraldine. Psychological writing, easy and hard. Amer. Psychologist, 1947, 2, 230-235.

83. Stouffer, S. A. Some observations on study design. Amer. J. Sociol., 1950, 55, 355-361.

84. Sullivan, H. S. The importance of a study of symbols in psychiatry. Psyche, 1926, 7, 81-93.

85. Thorndike, E. L. The teacher's word book. New York: Teachers Coll., Columbia Univer., 1921.

86. Thorndike, E. L. The Thorndike-century senior dictionary. New York: Appleton-Century, 1941.

87. Thorndike, E. L. Selected writings from a connectionist's psychology. New York: Appleton-Century-Crofts, 1949.

88. Trager, G. L. The field of linguistics. Studies in linguistics: Occasional Papers, No. 1. Norman, Okla.: Battenberg Press, 1949.

89. Warner, W. L. A black civilization. New York: Harper, 1937.

90. Watson, J. B. Behaviorism. New York: People's Institute Publishing Co., 1924.

91. Whitlow, M. Little puffs of air. Review of Journey to the end of the night (Louis-Ferdinand Celine). ETC., 1948, 5, 127-132. (P. 130, quoting Celine.)

92. Whorf, B. L. Four articles on metalinguistics. Washington: U. S. Department of State, Foreign Service Institute, 1950.

93. Whorf, B. L. Language, mind, and reality. ETC., 1952, 9, 167-188.

94. Wiener, N. Cybernetics. New York: Wiley, 1948.

95. Wilcoxson, F. Some rapid approximate statistical procedures. (Rev. Ed.) New York: American Cyanamid Company, Agricultural Chemicals Division, July, 1949.

96. Winthrop, H. Logos, telos, and psychos: The semantics of value disorder and personality disintegration. ETC., 1948, 5, 252-262.

97. Woodworth, R. S., and Sells, S. B. An atmospheric effect in formal syllogistic reasoning. J. exp. Psychol., 1935, 18, 451-460.

98. Zipf, G. K. Human behavior and the principle of least effort. Cambridge, Mass.: Addison-Wesley, 1949.

99. Publication manual of the American Psychological Association. Supplement to Psychol. Bull., 1952, 49 (4, Part 2), 1-61.

100. The whole truth? (Article on Elmer Davis.) Time, 1952, 60 (4), 51.

APPENDIX A

GLOSSARY

Taken from Charles Morris (54, pp. 345-356)

This glossary includes the main terms of the present system of semiotic. A few terms from general behavior theory are included, but since they are not defined in terms of signs, they really operate as undefined terms in this system. This glossary is not an index; it does not include many of the topics treated in the text nor the discussion of terms as employed by other semioticians. Its sole purpose is to help the reader keep at hand the main terminology adopted in this work. The terms that are starred are the most important and provide the basis for the definitions of the other terms.

*Appraisor . . . . . A sign that signifies something as having a preferential status for behavior. A significatum of such a sign is called a valuatum. Positive and negative appraisors are distinguished; also, utilitators and consummators, depending upon whether what is signified is a means-object or a goal-object.

*Ascriptor . . . . . A sign complex (or combination of sign complexes) in which something is signified in the identificative mode of signifying and in some other mode of signifying (designative, appraisive, prescriptive, or formative). Ascriptors are accordingly designative, appraisive, prescriptive, and formative.

Behavior . . . . . This term is presupposed by semiotic and not defined within it. Roughly speaking, behavior consists of the sequences of responses (actions of muscles and glands) by which an organism seeks goal-objects that satisfy its needs. Behavior is therefore "purposive" and is to be distinguished from response as such and from the even wider class of reactions. Behavior is individual or social, and when social may be co-operative, competitive, or symbiotic.

*Behavior-Family . . . . . Any set of response-sequences that are initiated by similar stimulus-objects and that terminate in these objects as similar goal-objects for similar needs.

Belief . . . . . An interpreter is said to believe that an ascriptor is a T-ascriptor or is reliable to the degree that he is disposed to act as if the ascriptor denotes or is reliable. Belief in general is not here defined.

Communicatee . . . . . An organism that interprets a sign produced by a communicator.

Communication . . . . . When restricted to signs, communication is the arousing of common significata by the production of signs. It is language communication when the signs produced are language signs. Not all communication is language communication.

Communicator . . . . . An organism that produces a sign that is a stimulus in social behavior to some organism (the communicatee).

***Denotatum** . . . . . Anything that would permit the completion of the response-sequences to which an interpreter is disposed because of a sign.

***Denote** . . . . . A sign that has a denotatum or denotata is said to denote its denotatum or denotata. All signs signify, but not all signs denote.

***Designator** . . . . . A sign that signifies characteristics or stimulus-properties of stimulus-objects. Such a significatum is called a discriminatum. Monadic, dyadic, and triadic designators are distinguished; also, object-designators and character-designators.

***Disposition To Respond (Response-Disposition)** . . . . . The state of an organism at a given time such that under certain additional conditions a given response takes place.

Dominant Sign . . . . . The sign in an ascriptor whose interpretant is not subordinated to other interpretants; its mode of signifying determines the mode of signifying of the ascriptor.

***Formative Ascriptor** . . . . . A compound ascriptor such that the denotation of one or more of the component ascriptors (called the antecedent ascriptors) is, because of the signification of the ascripter, a sufficient

condition for the denotation or lack of denotation of the remaining component ascriptor or ascriptors (called the consequent ascriptors), and hence, for the denotation or lack of denotation of the compound ascriptor itself. Formative ascriptors are accordingly either analytic or contradictory.

*Formator . . . . . A sign signifying how something is signified in an ascriptor. Stated behaviorally, formators are signs that dispose their interpreters to modify in determinate ways the dispositions to response occasioned by other signs in the sign combinations in which the formator appears. The significatum of a formator is called a formatum. Determiners, connectors, and modors are kinds of formators.

Goal-Object . . . . . An object that partially or completely removes the state of an organism (the need) which motivates response-sequences.

*Incitive Use of Signs . . . . . When signs are produced to call out more or less specific ways of responding to something. Signs incitively adequate are called persuasive.

*Informative Use of Signs . . . . When signs are produced to cause someone to act as if something had, has, or will have certain characteristics (discriminata, stimulus-properties). Signs informatively adequate are said to be convincing; they need not be true.

*Interpersonal Sign . . . . . A sign is interpersonal to the degree that it has the same signification to a number of interpreters; otherwise personal.

*Interpretant . . . . . The disposition in an interpreter to respond, because of a sign, by response-sequences of some behavior-family.

*Interpreter . . . . . An organism for which something is a sign.

Interpreter-Family . . . . . The interpreters for whom a sign is interpersonal.

*Language . . . . . See Lansign System.

*Language Sign . . . . . See Lansign.

*Lansign . . . . . A sign that is a member of a lansign-system. . . .'language sign' is often used in place of 'lansign'; strictly speaking, only the latter term is defined.

*Lansign-System . . . . . A set of plurisituational signs with interpersonal significata common to members of an interpreter-family, the signs being producible by members of the interpreter-family and combinable in some ways but not in others to form compound signs. Stated in terms of comsigns: a lansign-system is a set of comsign-families the members of which are restricted in the ways in which they may be combined. . . .'language' is often used in place of 'lansign-system'; strictly speaking, only the latter term is defined.

*Lexicative Ascriptor . . . . . Any ascriptor that is not a formative ascriptor.

*Lexicator . . . . . Any sign that is not a formator; hence any identifier, designator, appraiser, or prescriptor.

Linguistics (General) . . . . . A possible term for that part of semiotic which studies lansign-systems. Linguists are free to restrict the term 'linguistics' to some sub-set of such systems (such as those whose sign-vehicles are sounds produced by organisms).

Logic . . . . . A possible term for that part of semiotic composed of analytic formative ascriptors. Logical analysis would then consist of those sentences about signs which are analytic formative ascriptors in the semiotical metalanguage.

Means-Object . . . . . An object that serves as a means toward the attainment of a goal-object.

*Mode of Signifying . . . . . A differentiation of signs in terms of the most general kinds of significata. Five modes of signifying are distinguished (identificative, designative,

appraisive, prescriptive, and formative), and signs signifying in these modes are called respectively identifiors, designators, appraisors, prescriptors, and formators.

Need . . . . . Here used as roughly synonymous with 'organic state motivating behavior,' and hence, distinguished from what is necessary for an organism to survive.

*Pragmatics . . . . . That branch of semiotic which studies the origin, the uses, and the effects of signs. It is distinguished from semantics and syntactics.

*Post-Language Symbols . . . . . Symbols producible by their interpreters and synonymous with language signs (lansigns). Such symbols may be personal or interpersonal, and may or may not themselves become elements in a language (lansign-system).

*Preparatory-Stimulus . . . . . A stimulus that influences a response to some other stimulus. A preparatory-stimulus necessarily causes at the time of stimulation a reaction in the organism for which it is a stimulus, but this reaction need not be a response (an action of a muscle or gland).

Prescription . . . . . A prescriptive ascriptor considered as produced by some interpreter.

*Prescriptor . . . . . A sign that signifies the requiredness of certain response-sequences. A significatum of such a sign is called an obligatum. Hypothetical, categorical, and grounded prescriptors are distinguished.

*Reliable Sign . . . . . A sign is reliable to the degree that members of the sign-family to which it belongs denote; otherwise unreliable.

*Response-Sequence . . . . . Any sequence of consecutive responses whose first member is initiated by a stimulus-object and whose last member is a response to this stimulus-object as a goal-object (an object that partially or completely removes the state of the organism that motivates the sequence of responses).

*Semantics . . . . . That branch of semiotic which studies the signification of signs. It is distinguished from syntactics and pragmatics.

Semiosis . . . . . A sign-process, that is, a process in which something is a sign to some organism. It is to be distinguished from semiotic as the study of semiosis. The terms 'semiosical' and 'semiotical' may be distinguished in a similar way.

*Semiotic . . . . . The science of signs. Its main subdivisions are semantics, syntactics, and pragmatics. Each of these, and so semiotic as a whole, can be pure, descriptive, or applied. Pure semiotic elaborates a language to talk about signs, descriptive semiotic studies actual signs, and applied semiotic utilizes knowledge about signs for the accomplishment of various purposes.

*Sign . . . . . Roughly: something that directs behavior with respect to something that is not at the moment a stimulus. More accurately: If A is a preparatory-stimulus that, in the absence of stimulus-objects initiating response-sequences of a certain behavior-family, causes in some organism a disposition to respond by response-sequences of this behavior-family, then A is a sign. Anything that meets these conditions is a sign; it is left undecided whether there are signs that do not meet these conditions.

*Signal . . . . . A sign that is not a symbol, that is, not produced by its interpreter and not a substitute for some other sign with which it is synonymous.

Sign-Behavior . . . . . Behavior in which signs occur.

*Sign-Family . . . . . A set of similar sign-vehicles that for a given interpreter have the same signification.

*Significatum . . . . . The conditions such that whatever meets these conditions is a denotatum of a given sign. The formulation of what a sign signifies is called a formulated significatum. No attempt has been made

to differentiate 'signification' and 'significatum.'

*Signify . . . . . To signify is to act as a sign in a process of semiosis. 'To have signification' and 'to have a significatum' are synonymous with 'to signify.' A sign is said to signify (but not denote) its significatum, that is, the conditions under which it denotes. All signs signify; not all signs denote.

Stimulus . . . . . Any physical energy that acts upon a receptor of a living organism. A stimulus causes a reaction in an organism, but not necessarily a response (a reaction of a muscle or gland).

Stimulus-Object . . . . . The source of a stimulus. The properties of the object that produce stimuli are called stimulus-properties.

*Symbol . . . . . A sign that is produced by its interpreter and that acts as a substitute for some other sign with which it is synonymous; all signs not symbols are signals. Symbols may be pre-language, language, and post-language symbols.

*Systemic Use of Signs . . . . . When signs are produced to organize behavior that other signs tend to provoke. Signs systemically adequate are called correct.

*Syntactics . . . . . That branch of semiotic that studies the way in which signs of various classes are combined to form compound signs. It abstracts from the signification of the signs it studies and from their uses and effects; hence, it is distinguished from semantics and pragmatics.

*Type of Discourse . . . . . A specialization of language for the accomplishment of specific purposes. The types of discourse are here classified on a mode-use basis, that is, in terms of the combination of some dominant mode of signifying with some dominant use. Thus, there are sixteen major types of discourse: designative-informative, designative-valuative, appraisive-incitive, formative-systemic, and so forth.

*Use of a Sign . . . . . A sign is used with respect to some goal if it is produced by an interpreter as a means of attaining that goal; a sign that is used is thus a means-object. Four main uses of signs are distinguished: the informative, valuative, incitive, and systemic. The use of a sign is not to be confused with its mode of signifying.

*Valuative Use of Signs . . . . . When signs are used to cause preferential behavior to something. Signs valuatively adequate are said to be effective.

APPENDIX B

Pre-test Instruction Sheet
Experiment I

NAME: _____

DATE: _____

We all know that members of various professions, etc., use different words to get their messages across. The spoken and written words of lawyers and physicists, for example, are not alike--either in type or organization. Often the meaning of the same word will change--depending on WHO is using it. Such specializations of language have been called "types of discourse." Thus we might have a legal type of discourse, often employed by lawyers, and a scientific type, used by the physicist in preparing technical papers, etc.

Our understanding increases if we can distinguish various types of discourse. Professor Charles Morris has classified 16 main types of discourse. (See the complete list at the bottom of this page.) Three of the terms in the list may not be familiar to you: "rhetorical" means to emphasize style in speech or writing, i.e., artistic, persuasive, or skillful use of words; "cosmological" means to treat the universe as an orderly system, i.e., theorizing about the nature of the universe, etc.; "metaphysical" is a term applied to abstruse or abstract reasoning.

INSTRUCTIONS:

On the following 3 pages you will find 6 brief paragraphs.

(1) Read each paragraph carefully.

(2) Select from the list below the name of the type of discourse which you think the paragraph represents.

(3) Write the name of the type of discourse you select on the line to the left of the paragraph.

(4) DO EACH AND EVERY PARAGRAPH!

(5) You will have no more than 15 minutes to complete this task.

(6) EACH PARAGRAPH REPRESENTS A DIFFERENT TYPE OF DISCOURSE.

(We realize that this is a difficult assignment, and probably no one could do it satisfactorily without special training. Therefore, don't worry about your answers; merely record your best estimate on the appropriate line.)

(LIST OF THE 16 MAIN TYPES OF DISCOURSE)

1. COSMOLOGICAL
2. CRITICAL
3. FICTIONAL
4. GRAMMATICAL
5. LEGAL
6. LOGICAL-MATHEMATICAL
7. METAPHYSICAL
8. MORAL
9. MYTHICAL
10. POETIC
11. POLITICAL
12. PROPAGANDISTIC
13. RELIGIOUS
14. RHETORICAL
15. SCIENTIFIC
16. TECHNOLOGICAL

Pre-test Instruction Sheet
First Revision--Experiments II and III

NAME: _____ 178

DATE: _____

We all know that members of various professions, social groups, etc., use different words in order to get their messages across. The spoken and written words of prize fighters and doctors, for example, are not alike--either in type or organization. Often the meaning of the same word will change--depending on WHO is using it! Such specializations of language have been called "types of discourse." Thus we MIGHT have a "fistic" type of discourse, often employed by boxers and fight managers, and a "medical" type, used by doctors in discussing diseases.

Professor Charles Morris has classified 16 main types of discourse. (See the complete list at the bottom of this page. Some of the names have been changed for the purpose of this study.) Our understanding of what we read and hear increases IF WE CAN DISTINGUISH VARIOUS TYPES OF DISCOURSE!

INSTRUCTIONS:

On the following 3 pages you will find 6 brief paragraphs.

(1) Please read each paragraph very carefully.

(2) Select from the list below the name of the type of discourse which you think the paragraph represents.

(3) Write the name of the type of discourse you select on the line to the left of the paragraph.

(4) DO EACH AND EVERY PARAGRAPH!

(5) Please judge each selection thoughtfully; you will have 15 minutes to complete this task.

(6) EACH PARAGRAPH REPRESENTS A DIFFERENT TYPE OF DISCOURSE! No two of your answers should be alike when you are finished. However, you may change any answers prior to handing in your paper.

(We realize that this is a difficult assignment, and probably no one could do it perfectly without special training. Therefore, don't worry about your answers; merely record your best estimate on the appropriate line. RAISE YOUR HAND IF YOU HAVE A QUESTION AT ANY TIME.)

(LIST OF THE 16 MAIN TYPES OF DISCOURSE)

1. DOCTRINAL
2. DOCUMENTAL
3. ETHICAL
4. EXEMPLARY
5. IDEALISTIC
6. IMAGINATIVE
7. IMPERATIVE
8. JUDGMENTAL
9. MYTHICAL
10. PERSUASIVE
11. PROCEDURAL
12. RIGOROUS
13. STATUTORY
14. SUPPOSITIONAL
15. TESTABLE
16. UNIVERSAL

Pre-test Instruction Sheet--Pre-test Group
Second Revision--Experiment V

Two changes were made in this revision:

1. Line 18 now reads: "On the following page you will find 2 brief paragraphs."

2. Line 20 now reads: "Please judge each selection thoughtfully; you will have 5 minutes to complete this task."

Otherwise this is the same as the first revision.

Memory Sheet
Experiments I, II, III, and V--Pre-test Group

THIS IS NOT A TEST OF MEMORY!

We realize that most of you can remember the answers you gave the first time you scored the paragraphs.

The point is to change any of the answers that you now feel were incorrect--ON THE BASIS OF THE MATERIAL YOU HAVE JUST READ!

WORDS (spoken and written) CAN AND DO INFLUENCE BEHAVIOR!

A person on the way to a certain town is driving along a road; he is stopped by another person who says that the road is blocked some distance away by a landslide. The person who hears the words does not continue to the point in question, but turns off on a side-road and takes another route to his destination. It would be commonly said that the sounds made by the one person and heard by the other (and indeed by the speaker also) were signs to both of them of the obstacle on the road, and in particular were language signs, even though the actual responses of the 2 persons are very different; i.e., even though the words (signs) are the same for both persons, one will respond by changing his route, etc., while the other may stay to warn the next motorist, etc. The person addressed behaves in a way to satisfy a need, i.e., arrival at a certain town. The driver has various ways of achieving his goal; if the obstacle is encountered, he reacts differently than when spoken to at a distance from the obstacle. The spoken words are not responded to as the obstacle (the man may continue to drive for a time down the blocked road before turning off), and yet the words control or direct the course of behavior toward a goal in a way SIMILAR to the control which would be exercised by the obstacle if it were present as a stimulus; the words determine the man's behavior to be that of getting to a certain town by avoiding a certain obstacle at a given place on a given road. The words are in some sense a "substitute" in the control of behavior for the control over behavior which would be exercised by the obstacle itself.

Let's look at that opening statement again, namely, "WORDS (spoken and written) CAN AND DO INFLUENCE BEHAVIOR!" But hasten to add that not all words are of the same type and not all words are produced for the same

purpose! Shakespeare's words (including their organization) in "As You Like It" are quite different from those addressed to the driver of the car. Shakespeare sought to entertain (among other things), while the man who addressed the driver gave some information (if there actually were NO landslide, then his purpose was to deceive). The language of everyday speech is an amazing complex of various types of words which minister to a variety of purposes. In the course of time various specializations of this common language have appeared so that certain purposes may be more adequately met. Such specializations of languages may be called TYPES OF DISCOURSE. Professor Charles Morris has developed a method by which at least 16 types of discourse can be recognized. ONE TYPE OF DISCOURSE IS NEITHER BETTER NOR WORSE THAN ANOTHER: each has its rightful place in communication. However, to function intelligently, we must know, or take the proper steps to find out, whether a given "meaningful" grouping of words represents fantasy, fact, etc. For example, your professor may define psychology as "the science of behavior," and later on he may be discussing the growth and maturation of children when a student interrupts him to say, "Shakespeare's Jaques disagrees with you, Dr. Earnest. He claims that man has 7 ages." At least one thing could be said about the student's statement: He has confused the poetic discourse of Shakespeare with the (we trust) scientific discourse of his professor. This is NOT to say that Shakespeare is wrong; man may indeed have 7 ages. The psychologist may get an idea from Shakespeare, thanks to the alert student, and proceed to set up an experiment to try to determine what if any are the ages of man. The simple fact is this: Shakespeare and the professor represent different callings, and different professions employ different types of words for different purposes; the "meanings" of even the same words may change from one man (or profession) to another! Many a man has come to grief from accepting the anecdotal remarks of his barber as sound medical advice. In the next few paragraphs, I shall try to explain (so that you will be able to use it) something of Professor Morris' method of classification.

MODES OF SIGNIFYING: There are 4 modes. A mode of signifying may be thought of as (1) the type of words used, and (2) the type of sentence, etc. For example, what are the distinctions between such sentences as "That is a deer," "What a fine fellow!", "Keep the wind ahead!", "It will rain or not rain" and such terms as "deer," "fine," "keep," and "not"? "Deer" is called a DESIGNATOR; i.e., this word signifies certain characteristics of an object, and when these characteristics are learned, we may know that a certain object is in fact a deer and not a bear. (Designators are not always concrete nouns. A word is a designater if it specifies the "object's" characteristics sufficiently so that we could recognise an example of the word if we ever did experience it, e.g., an atom.) "Fine" is called an APPRAISOR; i.e., it signifies something as having a preferential status for behavior--"We will vote for him because he is a fine fellow; we prefer him." "Keep" is a PRESCRIPTOR; i.e., it signifies the requiredness of certain responses in order to achieve a certain goal--"If you don't keep the wind ahead, you will not reach port." "Or" and "not" are known as FORMATORS; i.e., they themselves give no information (about the environment) but instead modify our tendencies to respond--"It will rain or not rain" merely orients us as to the topic for consideration, etc. Formative sentences are "about" something; they may be true, etc., but they do not designate, appraise, or prescribe! Statements in which designators predominate are said to be in the DESIGNATIVE MODE of signifying; if appraisors predominate, the APPRAISIVE MODE; and so on for PRESCRIPTIVE MODE and FORMATIVE MODE. What of the sentence "That deer is fine!"? "Deer" is a designator and "fine" is an appraisor; what is the mode? As a general rule the predicate of the sentence will determine the mode; in the example above, since "fine" is an appraisor, we would say that the sentence was in the appraisive mode of signifying. The important feature is the dominant sign or word, usually found in the predicate, in determining the mode and not the word-type which is most numerous. (To make the role of formators clearer, see what happens when "$x + (y \times z) = k$" becomes "$(x + y) \times z = k$." If a sentence, etc., defies classification, it is probably in the formative mode. Textbooks on

grammar, mathematics, etc., are often good examples of writing in the formative mode.) Let us summarize:

MODE	EXAMPLE	WORD-TYPE
Designative	That is a deer.	Generally nouns
Appraisive	What a fine fellow!	Generally adjectives
Prescriptive	Keep the wind ahead!	Generally verbs (prescriptions, commands, demands, etc.)
Formative	Women are women.	Connectors, punctuation, "is," "or," "not," "+," "5," etc.

Designative mode is stressed in scientific writing, novels, etc.

Appraisive mode in myths, poetry, etc.

Prescriptive mode in cook books, sermons, etc.

Formative mode in books on logic, grammar, mathematics, etc.

THE 4 PRIMARY USAGES OF SIGNS OR WORDS: We must ask not only what sort of words are we reading (or hearing), but also what is the author's purpose. A writer may write a book to make money; this would be difficult for a reader to determine. However, there are 4 usages which can be ascertained with considerable accuracy: Words may be used to (1) INFORM us of something; (2) lead us to EVALUATE something positively or negatively; (3) INCITE (direct or urge) us to action; and (4) ORGANIZE our disposition to respond (i.e., "2 4 2 + =," means little if anything to you while "4 + 2 = 2" does). These 4 uses are called INFORMATIVE, EVALUATIVE, INCITIVE, and SYSTEMIC. (Remember we must try to determine the author's purpose when judging the use of words.) Let us summarize again:

INFORMATIVE USE: (designators)	"In our school system 7 teachers instruct 490 pupils, and their combined salaries equal $8,400.00 per year."
EVALUATIVE USE: (appraisers)	"They are without doubt the very best teachers in the U.S.A."
SYSTEMIC USE: (formators)	"Either they are going to resign or not resign."
INCITIVE USE: (prescriptors)	"These unfortunate teachers are overworked and underpaid, and we MUST take steps immediately to double their pay or I promise you they will all resign!"

Now go back and read the 4 sentences together as a paragraph. When we

consider them as a unit, we might say that the primary USE was incitive, i.e., the speaker or writer wants the audience to grant the teachers a raise; he is prescribing a course of action. (Remember use is determined by an over-all judgment or estimate. We get 2 additional clues as to use: (1) The author kept the prescriptive sentence for the end after building up to it, and (2) the emphasis on "must.") If we look at the words themselves, particularly the "predicates," those that are dominant are adjectives (appraisors), namely, "best," "unfortunate," "underpaid," "overworked," etc., and thus the dominant MODE of signifying would probably be appraisive. That type of discourse which is primarily appraisive in mode and incitive in use is called MORAL DISCOURSE.

MODE-USE CLASSIFICATION (TYPES OF DISCOURSE): Professor Charles Morris has provided a simple matrix (chart); all one has to do is select the mode and use, and then read the type of discourse directly from the matrix (see below). In using the matrix, the 4 modes are in the left (vertical) column, and the 4 uses are in the top (horizontal) row. To determine the type of discourse, find that cell which is common to the selected mode and use. Note that discourse which is predominantly designative in mode and primarily evaluative in use is called "fictional" discourse; that which is appraisive and systemic is called "critical"; and so on throughout the 16 types of discourse.

MODE ↓ / USE →	Informative	Evaluative	Incitive	Systemic
Designative	scientific	fictional	legal	cosmological
Appraisive	mythical	poetic	moral	critical
Prescriptive	technological	political	religious	propagandistic
Formative	logical-mathematical	rhetorical	grammatical	metaphysical

Later you will be asked to judge the paragraphs that were used at the beginning of this session; determine the predominant mode and primary use.

It will be helpful to refer back to page 4, re "the teachers."
(1) You will not have time enough to judge each sentence separately; try to select the dominant modes from the WORDS of the key sentence and CHOOSE that mode which is most pronounced in your opinion.
(2) Render an over-all judgment as to the use (purpose) which you feel the AUTHOR had in mind when he wrote the paragraph.
(3) When you have completed steps (1) and (2), consult the matrix (chart) and write the name of the type of discourse on the line to the left of the paragraph. The intersection (common cell) of the mode and use you select contains the name of the "type of discourse."

(Any given paragraph may contain features of several types of discourse. Most writing is subtle; so don't expect to find clear-cut examples of modes of signifying. Often it is easier to establish the correct mode by elimination of the incorrect modes. A unique core is identifiable in each paragraph that you will judge; evaluate the paragraph in terms of that core.

Do not be concerned about the "names" of the types of discourse; the titles are unimportant. Your job is to determine the MODE and USE and then to write the name of the appropriate type of discourse on the line to the left of each paragraph.)

NOW--PLEASE TURN TO THE BOOKLET AND
FOLLOW THE INSTRUCTIONS GIVEN THERE

The following changes were made in the "Morris Treatment", Experiment II:

1. Page 2, line 13: from "poetic" to "idealistic".
2. Page 2, line 14: from "scientific" to "testable".
3. Page 3, four illustrative sentences at the bottom of the page omitted.
4. Page 4, line 32: from "MORAL" to "ETHICAL".
5. Page 5, line 5: from "fictional" to "imaginative".
6. Page 5, line 6: from "critical" to "judgmental".
7. Page 5, the revised titles for the types of discourse were used in the matrix.
8. Page 5, the following footnote was added: "*For the purposes of this study, the names of certain types of discourse have been changed."

The following changes were made in the second revised "Morris Treatment":

1. Page 3, the word "MODEL" was added.

2. Page 3, one sentence was omitted prior to the presentation of the MODEL, but replaced in the following paragraph which appeared below the MODEL: "If a sentence, etc., defies classification, it is probably in the formative mode, but one should try very hard to fit a sentence into one of the other modes before concluding that it is in the formative mode: textbooks on grammar, mathematics, etc., often exemplify writing in the formative mode. FORMATORS ARE OFTEN MODIFIERS, e.g., "nine men," etc.; here "nine" modifies "men." The word "men" is a designater. Thus the sentence, "Nine men ran onto the field" is in the designative mode, NOT the formative mode! (Obviously formators may modify appraisors, prescriptors, and other formators, also.)"

3. Page 5, the subjects were referred to page 3, as well as page 4.

* * *

"Morris Treatment": Third Revision
Experiment V--Non-Pre-test Group

1. Page 5, the first sentence below the matrix was altered to read: "Later you will be asked to judge paragraphs; determine the predominant mode and primary use."

* * *

The revised titles for the types of discourse were used in the second and third revisions of the "Morris Treatment."

'Operationism' is a new name for recognized aspects of scientific method. It emphasizes the requirements that scientific concepts must meet to be meaningful and fruitful. Descriptions, laws, hypotheses, and theories may be critically examined on the basis of such criteria applied to the concepts they contain. Operationism represents a recent formulation of some of the features of experimental method and empiricism--generally in the direction of pragmatism and instrumentalism (Peirce, James, and Dewey). Bridgman's criteria of meaning resemble those expressed by C. S. Peirce in "How to Make Our Ideas Clear"--1878. We are concerned with two problems: "WHAT DO WE MEAN BY THE WORDS AND SYMBOLS WE USE?" and "HOW DO WE KNOW THAT WHAT WE ASSERT IN THESE TERMS IS TRUE (OR CONFIRMED TO SOME DEGREE)?" Further--"What do we do to find out if a term is applicable?" and "How do we know that it is scientifically useful?" Do implies measurement and experiment, and know reminds us that a term is scientifically worthwhile only if it aids prediction and perhaps practical control. Operationists thus emphasize criteria of scientific meaningfulness and fruitfulness for concepts and of criteria of validity for factual statements. The trend of our age toward a fuller awareness of the presuppositions and guides to our thinking and doing has promoted a more penetrating analysis of scientific method. Interest in operationism arose from the need (1) of purifying scientific method by eliminating metaphysical and prescientific elements, and (2) of understanding the meaning of complex concepts used in constructing scientific theories.

Clarification of meaning is pursued by DEFINITION, and definition may be considered a statement of the RULE CONCERNING THE USE of a term or symbol. The need for definition arises when (a) we are not sure what a given term means; (b) where there is a plurality of meanings; (c) when a term is so vague that greater precision is demanded; (d) when a term is to be enriched in meaning. This wider context may be a system of symbols (a pure calculus) or it may include symbols with empirical reference; in this latter case a 'coordinating definition,' i.e., relating a symbol of

an abstract system to terms which through previous definition already possess empirical reference; (c) definition is needed when a research situation calls for coining a new term as an abbreviation for terms already in use or logico-mathematically constructed. This latter condition arises particularly in the discovery of new elements or relationships in research subject matter.

(Definition as here conceived is NOMINAL, i.e., definition of terms or symbols. REAL definitions--of things, properties, etc.--are empirical descriptions with all terms understood by previous nominal definition or as characterizations of things for identification. Since identification occurs through labelling things on the basis of observational test, it amounts to NOMINAL definition of the label.)

To demand definition of every term used in scientific discourse would be unduly pedantic, utopian, and unnecessary. The sane use of common language is sufficiently definite for intersubjective communication and intelligibility regarding terms representing things and their observable properties. Doubts re meaningfulness or precise meaning usually arise with higher-order constructs and/or terms of fairly clear subjective meaning that nevertheless lack determination for intersubjective testability. Doubts of this second sort are rare in physics but plentiful in psychology, and doubts of the first sort arise in any science above the level of observable-property predicates.

A long series of definitional steps will terminate with definientia beyond the realm of words, i.e., items of direct observation. Is this last step itself a kind of definition? It is fashionable to speak of 'ostensive definitions.' Psychologists should not have much trouble with this: Our use of language is a product of learning, and we learn intra-linguistic relations of words and the extra-linguistic relations (applications) of SOME words. What to the logician is a rule for the use of symbols is to the psychologist a rule-regulated habit re symbolic (verbal) behavior. An ostensive definition then may be a designation rule in a metalanguage

(language about language) or practical drill in learning the 'right use' of words. A logical analysis of the meaning of scientific terms uses the apparatus of syntax and semantics; psychological analysis can be carried out in the object language of behavioristic psychology (emission of sounds) where words are physical events and thus part of general behavior. These two ways of studying definitions and language use supplement each other, are important, and compatible. Logical analysis examines uses of language in terms of critical standards of consistency, etc., while the psychological study is a description and causal analysis of verbal behavior. Each method may reflect the other. The psychologist, studying definitional behavior, uses definitions himself, and thus provides subject matter for logical appraisal, and the logician in appraising definitions is BEHAVING and thus providing material for the psychologist.

Einstein realized that terms like 'absolute length,' 'absolute density,' etc. (Newtonian physics), had no empirical meaning because no experimental or observational procedures could be specified for their application. The operational criterion separates physics from metaphysics. It is obvious to some psychologists that Freud's 'death instinct' is devoid of factual, scientific meaning, though possibly of literary or emotional value. OPERATIONAL ANALYSIS ENABLES US TO DECIDE WHETHER A GIVEN TERM (IN THE WAY IT IS USED) HAS A 'CASH VALUE,' i.e., factual reference. If it does have factual reference, operational analysis is used to show us what it is in terms (ULTIMATELY) of the data of direct observation.

The velocity of a uniformly moving vehicle may be measured by a yardstick and a stopwatch, and the I.Q. may be determined by the Stanford-Binet test. In both cases a simple arithmetical division (s/t; mental age/chronological age) yields the final result. For more complex concepts the mathematical operations which define them are more complicated. The factual reference of physical and psychological concepts, however, depends mostly on mensurational and experimental operations and not the purely mathematical. These operations link the descriptive language of science

and the data of experience. It is merely a terminological question whether or not simple perceptual acts (discriminations) themselves should be labeled 'operations.' The question concerning the POSSIBILITY of operations amounts to asking whether or not the measuring or testing procedures referred to in the definition can be executed. But 'possible' and 'can be' are notoriously ambiguous, and we must distinguish between (a) logical (in principle), (b) empirical, and (c) practical (technical) possibility. No scientist restricts factual meaningfulness to testability within the bounds of the present technical facilities; bigger and better microscopes, etc., are obvious reminders of this.

Operationism has occasionally hypertrophied into a radically anti-theoretical attitude. Ever since Galileo replaced the question "Why?" by the question "How?" and since Newton pronounced his "Hypotheses non fingo" positivistic scientists have been inclined to restrict their endeavors to description and correlation. EXPLANATION is considered a metaphysical misfit (Mach and Ostwald rejected the atomic theory). Some psychologists of our day still reject as non-operational the psychoanalytic theory of the unconscious; others consider neuro-physiological hypotheses, when used as explanations for behavior, so much metaphysical verbiage. If the student of the history and the methodology of science be permitted to mediate in this quarrel, he would say that, as so often, the truth lies somewhere in the region of the golden mean.

Strict operationists have resorted to such easy and dogmatic quips as: "Intelligence is what intelligence tests test" or "temperature is what thermometers measure." These statements may be intended to intimidate inquirers of a mystical or metaphysical bent. But such statements are inadequate. Quantitative concepts are usually the product of long labors of adjustment by repeated definition. Thermometers, etc., did not arise in an historical vacuum but out of a background of qualitative and semi-quantitative knowledge. Problems of description and prediction led to

the search for precise, objective, and fruitful concepts, and these concepts were to represent EMPIRICAL LAWS, i.e., relationships between the various measurable variables. Considering this, it makes sense to ask whether a thermometer measures temperature adequately. There are empirical laws which relate temperature to mechanical energy, and if we are to give the First Law of Thermodynamics its most universal form, we must consider thermometer readings an approximation to be corrected or replaced by the Kelvin scale. Similar considerations apply to psychological magnitudes.

Empirical laws enable us to define the same concept by different operational routes. The fact that length may be defined by the yardstick and by triangulation techniques is a consequence of the empirical laws of (applied) geometry according to which hard and solid bodies as well as light rays exhibit (in terrestial experiments with a high degree of accuracy) Euclidean relationships. Nevertheless, Bridgman's warning to the effect that different operational routes define DIFFERENT concepts is not unjustified: The convergence of operational results is to be taken for granted only until further notice, i.e., until evidence to the contrary emerges.

Since science is a social enterprise, its operations must be repeatable by any properly equipped and trained observer. A statement is scientifically meaningful only if it is intersubjectively testable. Subjective peculiarities, etc., can become scientific subject matter--IF we can verify them in an objective manner. Private experience is only the raw material, NOT the subject matter of science, and if one wishes to convert directly given experience into descriptive subject matter, one is limited to phenomenological introspection; the scientific value of this is insignificant if it is unrelated to a study of physical and physiological processes. If one does relate the phenomenologically given data to these inferred (or conceptually constructed) processes, they can be conceptualized as overt or potential behavior or as hypothetical brain processes.

This happens in the "psychology of the other one," where one constructs the other one's experience on the basis of his own data. Here again we have a choice of various languages for description: mentalistic, behavioral, or--ultimately--physiological. The concepts describing "the other one's experience" are now constructs and not purely subjective symbols.

Though epistemologists and phenomenologists may wish to clarify and formulate the meaning of 'experience' along these lines, there is the alternative of a physicalistic or behavioristic approach right from the start. From the point of view of the methodology of science, this latter approach eliminates the pitfalls of the metaphysical pseudo-problems of solipsism, the mind-body puzzle, etc. If only intersubjective operations are admitted, one's own experience is on a par with that of the other one. Introspection itself is then defined as a response to previous responses. Many of the intra-dermal and cerebral conditions are not yet as directly accessible to outside observers as they are to the subject himself.

Operationism is not a system of philosophy, and it is not a technique for the formulation of concepts or theories. It will not by itself produce scientific results; these result from the labors of the researchers. Operationism is, rather, a set of regulatory or critical standards with which the meaningfulness and fruitfulness of scientific concepts may be appraised. Operationism stresses the following requirements: Valuable concepts in the factual sciences must be definable by operations which are (1) logically consistent; (2) sufficiently definite; (3) empirically rooted and linked with the observable; (4) naturally and, preferably, technically possible; (5) intersubjective and repeatable; (6) aimed at the creation of concepts which will function in laws or theories of greater predictiveness.

The degrees to which these ideals are approximated varies from one science to another. But it would seem that all of these criteria are applicable not only to a well-developed, systematized science such as physics but also to a science still largely in the making such as psychology.

NOW--PLEASE TURN TO THE BOOKLET AND
FOLLOW THE INSTRUCTIONS GIVEN THERE

The language process within an individual may be viewed as a continuous interaction between sequences of central events ("ideas") and sequences of instrumental skills, vocalic, gestural, or orthographic, which constitute the communicative product. One vocalizes, "It looks like rain today; I'd better not wash the car." This is a sequence of skilled movements but not different from tying one's shoes. The organization of these movements into word-units represents skill sequences of relatively high predictability; longer sequences involving syntactical order are also relatively predictable for a given language system. Execution of such sequences brings the communicator to "choice points" where the next sequence isn't highly predictable from the objective communicative product itself. The dependence of "I'd better not wash the car" upon "looks like rain today" (CONTENT) reflects determinants with the semantic system which effectively "load" the transitional probabilities at these choice-points. A method for measuring these skill sequences (meaning) will be proposed.

THE SEMANTIC DIFFERENTIAL: This method is a combination of associational and scaling procedures. It is an indirect method in the same sense that an intelligence test does not directly measure intelligence.

This method had its origins in synesthesia, defined by Warren's "Dictionary of Psychology" as "a phenomenon characterizing the experiences of certain individuals, in which certain sensations belonging to one sense or mode attach to certain sensations of another group and appear regularly whenever a stimulus of the latter type occurs." This implies a "neural short-circuiting" that is present in only a few individuals, and many of the case histories gave credence to this view: a subject reported pressure sensations about his teeth and cheeks whenever cold spots on his arms were stimulated, etc., etc. Here, on the other hand, was a man who imagined the number "1" to be yellow, "2" to be blue, and "8" to be black (anyone who has played pool will recognize the origin of this system).

What modalities are crossed in these cases?

Rather than being a rare phenomenon, Karwoski and Odbert report 13% of Dartmouth College students indulging in color-music synesthesia, often to enrich their enjoyment of music. They varied among themselves as to the modes of translation employed and the vividness of their experiences; their difference from the general population appeared to be one of degree rather than kind. Whereas fast, exciting music might be pictured by them as bright red forms, others would merely agree that terms like "red-hot," "bright," and "fiery" adequately described the music; a slow and melancholic section might be visualized as heavy, slow-moving "blobs" of sombre hue and described as "heavy," "blue," and "dark." The relation of this phenomenon to ordinary verbal metaphor is evident; a happy man is said to feel "high"; a sad man feels "low"; souls travel "up" to the good place and "down" to the bad place; etc. Metaphor in language and color-music synesthesia can be described as the parallel alignment of 2 or more dimensions of experience, defined verbally by pairs of opposites (translations occurring between equivalent portions of the continua).

Interrelationships among color, mood, and musical experiences have been studied more analytically. Subjects listened to 10 musical excerpts and indicated their dominant moods by checking descriptive adjectives. Then, on a second hearing, they listed the colors appropriate to each score. Significant relations between color associations and mood were shown. A portion of Delius' "On Hearing the First Cuckoo in Spring" was judged leisurely in mood and preponderantly green in color; Wagner's "Rienzi Overture" was exciting or vigorous in mood and red in color. When another group of subjects was merely shown the mood adjectives and asked to select appropriate colors, even more consistent relations appeared, suggesting that the unique characteristics of the music had obscured the color-mood relationships. Almost identical findings have been reported for relationships between stage lighting and reported moods of the audience, etc.

Responses to music are themselves too complex for analysis of specific relations between auditory-mood variables and color-form variables; therefore, we used simple melodies recorded by a clarinet as stimuli. The subjects drew their visualizations with colored pencils after listening in a darkened room. The simplest stimulus--the sound merely grew louder, then softer--will serve to illustrate. Subject A indicates increasing LOUDNESS by making the center of his line heavier, subject B by increasing amplitude of vibration, etc. Subject I created meaningful forms--here a little car that comes nearer and then away again. These are meaningfully equivalent responses to the same auditory stimulus--i.e., there are visual continua which parallel auditory continua. This method allows these relationships to show up clearly.

Are visualizers exercising a "rare" capacity or expressing overtly modes of translation that are implicit in our language? Another experiment used subjects who had never thought of "seeing things" when hearing music. The same simple melodies were played, and they were instructed to "force themselves to draw something to represent what they heard." They produced the same types and frequencies of visual forms as the experienced visualizers. Finally, 100 unselected students were given a verbal MEANING-POLARITY TEST, each item of which appeared thus: LARGE-small; SOFT-LOUD, with instructions to circle that word in the second pair which "seems most clearly related to" the capitalized word in the first pair. Large was linked to loud by 96% of these subjects, near with fast by 86%, etc. Thus imagery found in synesthesia is related to metaphor; both represent semantic relations.

Are such semantic relations dependent upon culture or more fundamental human determinants? To get at this question the writer studied anthropological reports on primitive cultures--Pueblo Indian, Australian Bushman, etc.,--to obtain evidence on semantic parallelism. Emphasis was given to nonmaterial aspects of culture (mythology, religion, etc.). There is danger of attributing relations to a primitive group that are actually the observer's projections and borrowings from Western culture.

Nevertheless, GOOD gods, places, etc., were regularly UP and LIGHT (white) in relation to BAD things, which were DOWN and DARK (black). A myth tells how the gods helped the original man struggle "up" from the "dark," "cold," "wet," "sad" world below the ground to the "light," "warm," "dry," "happy" world above. Siberian Aborigines, members of a privileged clan, call themselves "white" bones in contrast to all others who are "black" bones. Among the Uganda Negroes we find a white god being used to ward off evil spirits. Such data suggest a pervasive semantic frame of reference. Further study of the problem could be rewarding.

We adapted the above method to the study of social stereotypes. A set of scales was used to measure the "meaning" of concepts like PACIFIST, RUSSIAN, DICTATOR, and NEUTRALITY. Subjects were tested between April, 1940, and March, 1942. A single item appeared as follows:

PACIFIST: Kind ___:___:___:___:___:___: cruel

with the subject instructed to check that scale position which best represented the direction and intensity of his judgment. The concepts and scales were randomized. The feasibility of using this method to record the changing structures of social stereotypes was demonstrated. That a shift from a pacifistic to a militaristic frame of reference had been accomplished, even before Pearl Harbor, was clearly evident.

The descriptive scale judgments fell into highly intercorrelated clusters. "Fair-unfair," "kind-cruel," etc., correlated .80 or better. This cluster represented, we assumed, a single, general factor--the evaluative (good-bad) dimension. Gradients like "strong-weak," "realistic-unrealistic," etc., were independent of this group and pointed to other dimensions. Having subjects judge the same concepts "as a German" or "as an Englishman" produced gross, appropriate changes in the evaluative dimension but not in the pattern of each stereotype--e.g., GERMANS, when judged by students playing the role of Germans, were still seen as more "strong" and "happy" than "noble" or "kind." This illustrates the difficulty in trying to assume the point of view of another.

The researches described above gave rise to the following hypotheses:

1. THE PROCESS OF DESCRIPTION OR JUDGMENT CAN BE CONCEIVED AS THE ALLOCATION OF A CONCEPT TO AN EXPERIENTIAL CONTINUUM, DEFINABLE BY A PAIR OF POLAR TERMS. These "experiential continua" will be reflections (in language) of the sensory differentiations made possible by the human nervous system. It is assumed that discriminations in meaning cannot be any finer or involve any more variables than are made possible by the human nervous system.

2. MANY DIFFERENT EXPERIENTIAL CONTINUA, OR WAYS IN WHICH MEANINGS VARY, ARE ESSENTIALLY EQUIVALENT AND HENCE MAY BE REPRESENTED BY A SINGLE DIMENSION. This functional equivalence was evident in both the studies on synesthesia and those on the changing structure of social stereotypes. This fact about language and thinking makes the development of a quantitative measuring instrument feasible. If the plethora of descriptive terms we utilize were unique and independent, measurement would be impossible.

3. A LIMITED NUMBER OF SUCH CONTINUA CAN BE USED TO DEFINE A SEMANTIC SPACE WITHIN WHICH THE MEANING OF ANY CONCEPT CAN BE SPECIFIED. This opens the possibility of measuring meaning-in-general objectively and specifies factor analysis as the basic methodology. A limited number of dimensions or factors differentiate among the meanings of randomly selected concepts, and if the technique satisfies criteria of measurement, then the "semantic differential" IS an objective index of meaning. The procedures <u>are</u> an operational definition of meaning, just as the procedures followed in obtaining the IQ score provide an operational definition of intelligence.

The following operations are explicit, and they involve the subject's allocation of a concept within a standard system of descriptive dimensions by means of a series of independent associative judgments. The situation is designed to be simple. Presented with a pair of descriptive polar terms (e.g., <u>rough-smooth</u>) and a concept (e.g., LADY), the subject indicates the direction of this association (e.g., <u>LADY-smooth</u>). We have developed 2 different methods for collecting data: In the <u>graphic method</u>,

a pencil-and-paper technique, the subject indicates the intensity of his association by checking on a 7-step scale. In the judgment-time method, intensity of association is indicated by the latency of the subject's choice reaction toward one or the other of the terms. In both methods each associative judgment of a particular concept against a particular descriptive scale constitutes one item. In successive items, concepts and dimensions are paired in deliberately rotated orders until every concept has been associated with every scale by every subject.

A FACTOR ANALYSIS OF MEANING: A total of 50 descriptive scales, selected in terms of their frequency of usage, have been used in the judgment of 20 varied concepts, yielding a 1,000 item test. One hundred college students used the graphic method. The purpose of this factor analysis is to isolate a limited number of general dimensions of meaning to try to bring some order out of semantic chaos. The larger the proportion of total variance in meaning accounted for by these factors, the more satisfactory will be the measuring instrument. A preliminary estimation of the factors indicates the existence of several roughly independent dimensions. An "evaluative factor" accounts for by far the largest portion of the variance. There is evidence for a "strength factor," an "activity factor," and others. Given such factors it will be possible to select those specific scales (e.g., good-bad, strong-weak, active-passive, smooth-rough, hot-cold, etc.) which best represent them.

Thus the "semantic differential" and factor analysis become useful re:
(a) Semantic norms (connotations).
(b) Individual differences in meaning (CHURCH, LABOR LEADER, ETC.).
(c) Changes in meaning (in "time").
(d) Quantification of subjective language data (test-scoring).
(e) Cross-cultural communication problems.

NOW--PLEASE TURN TO THE BOOKLET AND
FOLLOW THE INSTRUCTIONS GIVEN THERE

"Morris Treatment" Post-test Instruction Sheet
Experiment I

NAME: _____

DATE: _____

STARTING TIME: _____

FINISHING TIME: _____

INSTRUCTIONS

On the next 3 pages you will find the paragraphs that you judged before.

1. Determine the type of discourse (your best estimate) for each selection. BE SURE THAT YOU USE THE MODEL FURNISHED IN THE STORY ABOUT THE TEACHERS ON PAGE 4. YOU MUST USE THE MATRIX ON PAGE 5 TO DETERMINE THE TYPES OF DISCOURSE. You may use anything else in the first booklet that you feel will help you. Turn back to it at any time or, better, keep it before you all the time while you are judging the paragraphs.
2. Write the name (title) of the type of discourse on the line to the left of each paragraph.
3. Remember--each paragraph represents a different type of discourse. NO TWO ARE ALIKE!
4. Do the best you can but work as rapidly as possible; allow yourself about 5 minutes for each paragraph.
5. BE SURE TO LABEL EACH AND EVERY PARAGRAPH EVEN IF YOU HAVE TO GUESS!
6. Raise your hand if you have a question at any time.

"Morris Treatment" Post-test Instruction Sheet
First Revision--Experiment II

1. Line 8, the following words were added: "Please read each selection very carefully."

2. Line 13--instruction No. 4--was changed to the following: "Please judge each selection thoughtfully. Allow 5 minutes per paragraph. You may change an answer at any time."

* * *

"Morris Treatment" Post-test Instruction Sheet
Second Revision--Experiment III

1. Line 3 was changed to read thus: "BE SURE THAT YOU USE THE MODELS FURNISHED ON PAGE 3 AND THE STORY ABOUT THE TEACHERS ON PAGE 4."

* * *

"Morris Treatment" Post-test Instruction Sheet
Third Revision--Pre-test Group, Experiment V

1. Line 1 was altered to read thus: "On the next page you will find the paragraphs that you judged before."

Non-Pre-test Group

1. Line 1 was altered to read thus: "On the next page you will find the paragraphs that you are to judge."

NAME: _____

DATE: _____

STARTING TIME: _____

FINISHING TIME: _____

INSTRUCTIONS

On the next 5 pages you will find the paragraphs that you judged before.

1. Determine the type of discourse (your best estimate) for each selection. Use the material you have just read in any way that seems appropriate. (YOU ARE PERMITTED TO TURN BACK TO THIS MATERIAL AT ANY TIME YOU CHOOSE.)
2. Write the name (title) of the type of discourse on the line to the left of each paragraph.
3. Remember--each paragraph represents a different type of discourse. NO TWO ARE ALIKE!
4. Do the best you can but work as rapidly as possible; allow yourself about 5 minutes for each paragraph.
5. BE SURE TO LABEL EACH AND EVERY PARAGRAPH EVEN IF YOU HAVE TO GUESS!
6. Raise your hand if you have a question at any time.

(LIST OF THE 16 MAIN TYPES OF DISCOURSE)

1. COSMOLOGICAL
2. CRITICAL
3. FICTIONAL
4. GRAMMATICAL
5. LEGAL
6. LOGICAL-MATHEMATICAL
7. METAPHYSICAL
8. MORAL
9. MYTHICAL
10. POETIC
11. POLITICAL
12. PROPAGANDISTIC
13. RELIGIOUS
14. RHETORICAL
15. SCIENTIFIC
16. TECHNOLOGICAL

Revision--Post-test Instruction Sheet
"Feigl Treatment"--Experiments II and III
"Osgood I Treatment"--Experiment II

1. Line 5, the following addition was made: "Please read each selection very carefully."

2. Line 10 was altered thus: "Please judge each selection thoughtfully; allow 5 minutes per paragraph. You may change an answer at any time."

3. In addition, the revised titles for the 16 main types of discourse were used. See page 178.

Experiment I

NAME: _____ STARTING TIME: _____

DATE: _____ FINISHING TIME: _____

INSTRUCTIONS

On the next 3 pages you will find the paragraphs that you judged before.

1. Determine the type of discourse (your best estimate) for each selection. Use the material you have just read in any way that seems appropriate. (YOU ARE PERMITTED TO TURN BACK TO THIS MATERIAL AT ANY TIME YOU CHOOSE.)
2. Write the name (title) of the type of discourse on the line to the left of each paragraph.
3. Remember--each paragraph represents a different type of discourse. NO TWO ARE ALIKE!
4. Do the best you can but work as rapidly as possible; allow yourself about 5 minutes for each paragraph.
5. BE SURE TO LABEL EACH AND EVERY PARAGRAPH EVEN IF YOU HAVE TO GUESS!
6. Raise your hand if you have a question at any time.

(Below is a matrix (chart) that has been prepared by Professor Charles Morris. It may be of some additional assistance to you. The intersection of MODE and USE, the common cell, will give you the name of the type of discourse. Mode means manner of writing. For the purposes of this experiment, it is not possible to tell you more about the matrix; use it or not as YOU SEE FIT.)

MODE ↓ USE →	Informative	Evaluative	Incitive	Systemic
Designative	scientific	fictional	legal	cosmological
Appraisive	mythical	poetic	moral	critical
Prescriptive	technological	political	religious	propagandistic
Formative	logical-mathematical	rhetorical	grammatical	metaphysical

(LIST OF THE 16 MAIN TYPES OF DISCOURSE)

1. COSMOLOGICAL
2. CRITICAL
3. FICTIONAL
4. GRAMMATICAL
5. LEGAL
6. LOGICAL-MATHEMATICAL
7. METAPHYSICAL
9. MYTHICAL
10. POETIC
11. POLITICAL
12. PROPAGANDISTIC
13. RELIGIOUS
14. RHETORICAL
15. SCIENTIFIC

1. Line 5, these words were added: "Please read each selection very carefully."

2. Line 10 was altered thus: "Please judge each selection thoughtfully; allow 5 minutes per paragraph. You may change an answer at any time."

3. The revised titles for the 16 main types of discourse were used; see page 178.

Test Paragraphs
Experiments I, II, and III
(See Appendix C for Experiments IV and V--Test Paragraphs)

I want to attempt to answer this question--are there functions of the nervous system which are at the same time both physiological and psychological? I don't mean by this occurrences like the electrical changes which can be detected in the back of the brain when we see, for since we do not know the connection between the electrical changes and seeing, they take us no further. But, are there events in the brain which we can understand as being at the same time both physiological and mental? Now I believe that there are, and it is in the realm of speech and thought that we shall find them, even though at the moment they are little more than dim outlines seen through the mist of our ignorance.

When I utter a statement a "part" of me, a characterizable subpersonality of mine, listens. My "listener" can use my eyes and ears, and adjust "himself" to signs from the 'real' audience, immediately present. Not so, my "reader," who has to represent a mediate public. "He" is so troublesome that I have had to make a study of the patterns characterizing "his" activity. I conclude that "he" is a queer combination of a high-grade imbecile and a bitterly paranoid critic. Every sentence has to be tortured into such shape that it cannot be misunderstood by the stupid or savagely misconstrued by an intelligent but bitterly hostile critic. My dealings with "him" are simply exhausting, and my writings reflect this and the fact that one writes books for a small audience.

The successful sciences are of humble birth; each had its lowly origin in a simple skill. Some centuries of hard and unremitting labor enabled chemistry and physics to achieve structures of knowledge that are most imposing. In doing this, they have not strayed into other paths, no matter how entrancing the prospect. The social sciences are impressed by this achievement, there is no doubt of that; but the unfortunate effect has been to encourage too much jerry-building of imposing facades in the social area. The pedestrian step-by-step development of a simple unquestionable skill, if it exists, is concealed by these elaborate fronts. It is kindness to suppose that the pretentious facades are perhaps only camouflage and that somewhere behind them real work is going on.

When a stimulus energy (S) impinges on a suitable receptor organ, an afferent neural impulse (s) is generated and is propagated along connected fibrous branches of nerve cells in the general direction of the effector organs, via the brain. During the continued action of the stimulus energy (S), this afferent impulse (s), after a short latency, rises quickly to a maximum of intensity, following which it gradually falls to a relatively low value as a simple decay function of the maximum. After the termination of the action of the stimulus energy (S) on the receptor, the afferent impulse (s) continues its activity in the central nervous tissue for some seconds, gradually diminishing to zero as a simple decay function of its value at the time the stimulus energy (S) ceases to act.

Repetition of a stimulus-response sequence sometimes results in increased sensitivity--an effect exactly the opposite of that just considered. A Nostoc hormogonium, which at first stimulation failed to react to illumination lasting 340 seconds, on the nineteenth trial responded to a stimulus duration of only 45 seconds. Amebas, moving into a lighted area, continue to send out pseudopods in the same general direction until, after a number of repetitions of the stimulus, they reverse the direction of locomotion. On subsequent trials the number of repetitive responses in the original direction decreased to zero (in 6 to 27 trials). Effects such as these, which are concerned with successive stimulus-response sequences, are not to be confused with the summation of stimuli to produce a single response.

Chi square can be used to test the significance of the difference between 2 observed frequency distributions, but this simply becomes a 2 by k table with expected values computed from the marginal totals. It is incorrect to treat either set of frequencies as those expected, against which the other is compared as a set of observed values. Such a procedure doesn't allow for the fact that both sets of frequencies are subject to sampling fluctuations. If one set of frequencies is for the universe, and the second set is based on a sample from the universe, then the universe frequencies can be used to determine expected frequencies, against which the sample values may be checked to test whether the sample represents the universe within the limits of chance sampling error.

APPENDIX C

The Test Paragraphs: Numbers 1 - 6 were used in Experiments I, II, and III; the two with asterisks in Experiment V; all items in Experiment IV. Titles for the types of discourse follow those of Experiment I. The scoring is E's, and additional scoring choices (order of preference) are given along with complete identifying information.

No. 1

Scoring Key: Rhetorical Other Choices: 2nd - Political
3rd - Grammatical

I want to attempt to answer this question--are there functions of the nervous system which are at the same time both physiological and psychological? I don't mean by this occurrences like the electrical changes which can be detected in the back of the brain when we see, for since we do not know the connection between the electrical changes and seeing, they take us no further. But, are there events in the brain which we can understand as being at the same time both physiological and mental? Now I believe that there are, and it is in the realm of speech and thought that we shall find them, even though at the moment they are little more than dim outlines seen through the mist of our ignorance.

Brain, Russel. "Speech and Thought" in The Physical Basis of Mind, pub. by Macmillan, 1950. Pgs. 46 and 47, paragraph No. 2.

* * *

No. 2

Scoring Key: Mythical Other Choices: 2nd - Poetic

When I utter a statement a "part" of me, a characterizable sub-personality of mine, listens. My "listener" can use my eyes and ears, and adjust "himself" to signs from the 'real' audience, immediately present. Not so, my "reader," who has to represent a mediate public. "He" is so troublesome that I have had to make a study of the patterns characterising "his" activity. I conclude that "he" is a queer combination of a high-grade imbecile and a bitterly paranoid critic. Every sentence has to be tortured into such shape that it cannot be misunderstood by the stupid or savagely misconstrued by an intelligent but bitterly hostile critic. My dealings with "him" are simply exhausting, and my writings reflect this and the fact that one writes books for a small audience.

Sullivan, Harry Stack. "Multidisciplined Coordination of Interpersonal Data" in Personality and Culture, pub. by the Viking Fund, 1949. One sentence from paragraph 3 on page 188 and most of paragraph 1 on page 189, plus changes by me, e.g., "in medias res" to "immediately present."

No. 3*

Scoring Key: Poetic Other Choices: 2nd - Fictional

The successful sciences are of humble birth; each had its lowly origin in a simple skill. Some centuries of hard and unremitting labor enabled chemistry and physics to achieve structures of knowledge that are most imposing. In doing this, they have not strayed into other paths, no matter how entrancing the prospect. The social sciences are impressed by this achievement, there is no doubt of that; but the unfortunate effect has been to encourage too much jerry-building of imposing facades in the social area. The pedestrian step-by-step development of a simple unquestionable skill, if it exists, is concealed by these elaborate fronts. It is kindness to suppose that the pretentious facades are perhaps only camouflage and that somewhere behind them real work is going on.

Mayo, Elton. The Social Problems of an Industrial Civilization. Published by Div. of Research, Grad. School of Bus. Adm., Harvard University, 1945. Pg. 21, paragraph No. 1.

* * *

No. 4

Scoring Key: Logical-Mathematical Other Choices: 2nd - Scientific

When a stimulus energy (S) impinges on a suitable receptor organ, an afferent neural impulse (s) is generated and is propagated along connected fibrous branches of nerve cells in the general direction of the effector organs, via the brain. During the continued action of the stimulus energy (S), this afferent impulse (s), after a short latency, rises quickly to a maximum of intensity, following which it gradually falls to a relatively low value as a simple decay function of the maximum. After the termination of the action of the stimulus energy (S) on the receptor, the afferent impulse (s) continues its activity in the central nervous tissue for some seconds, gradually diminishing to zero as a simple decay function of its value at the time the stimulus energy (S) ceases to act.

Hull, Clark L. Principles of Behavior. Appleton-Century-Crofts, Inc. Page 47, "Postulate 1."

No. 5

Scoring Key: Scientific Other Choices: 2nd- Legal

Repetition of a stimulus-response sequence sometimes results in increased sensitivity--an effect exactly the opposite of that just considered. A Nostoc hormogonium, which at first stimulation failed to react to illumination lasting 340 seconds, on the nineteenth trial responded to a stimulus duration of only 45 seconds. Amebas, moving into a lighted area, continue to send out pseudopods in the same general direction until, after a number of repetitions of the stimulus, they reverse the direction of locomotion. On subsequent trials the number of repetitive responses in the original direction decreased to zero (in 6 to 27 trials). Effects such as these, which are concerned with successive stimulus-response sequences, are not to be confused with the summation of stimuli to produce a single response.

Hissen, Henry W. "Phylogenetic Comparison." Ch. 11 in Handbook of Experimental Psychology. Ed. by S. S. Stevens. Pub. by John Wiley and Sons, Inc., 1951. Pg. 366, paragraph No. 4. Omitting: "Sensitization."

* * *

No. 6

Scoring Key: Technological Other Choices: 2nd- Grammatical

Chi square can be used to test the significance of the difference between 2 observed frequency distributions, but this simply becomes a 2 by k table with expected values computed from the marginal totals. It is incorrect to treat either set of frequencies as those expected, against which the other is compared as a set of observed values. Such a procedure doesn't allow for the fact that both sets of frequencies are subject to sampling fluctuations. If one set of frequencies is for the universe, and the second set is based on a sample from the universe, then the universe frequencies can be used to determine expected frequencies, against which the sample values may be checked to test whether the sample represents the universe within the limits of chance sampling error.

McNemar, Quinn. Psychological Statistics. John Wiley and Sons, Inc., 1949. Pg. 215, paragraph 1. Omitting: "as previously indicated," "In such a situation," "(or proportions)," and "in order." And changing: "does not" to "doesn't" and "set up" to "determine."

No. 7

Scoring Key: Poetic Other Choices: 2nd - Rhetorical
3rd - Critical
4th - Metaphysical

The human individual is endlessly simplifying and generalising his own view of his environment, and constantly imposing on this environment and on the events which occur in it his own constructions and meanings, but these constructions and meanings are largely group products of men and women who lived long ago. In other words, they are, in part at least, characteristic of one culture as opposed to another. It is true, of course, that no two individuals within a society have an identical set of basic concepts, explicit or implicit, conscious or unconscious; each person adds a little here, detracts a little there, makes this or that emphasis stronger than do most of his neighbors. Indeed, the cultural system of premises and categories is an abstraction, a statement of central tendencies and a range of dispersion.

Kluckhohn, Clyde K. "The Nature of Concepts Their Inter-Relation and Role in Social Structure." Fifth Session: "The Special Character of Integration in an Individual Culture." Pg. 80, paragraph No. 2. Deleted: "It is true, of course, that....." and "It is a meaningful and highly useful abstraction, but we must never lose sight of the fact that it is an abstraction none the less."

* * *

No. 8

Scoring Key: Scientific Other Choices: None

During the first period of 5 min. the animals required an average of 100 contacts. This number fell off abruptly in the second and third periods, then increased somewhat; but after the seventh five-min. period there was a steady decrease in the number of required contacts. Toward the end of the series it was seldom that an animal would require the tactual stimulus, and successions of two or three consecutive motionless periods were the rule. A 10-hr. rest in darkness was then given, after which an average of only 20 contacts was required to keep the worms motionless during a 5-min. test. In the next 5-min. period one or two contacts sufficed, and the former low point was maintained in the following test periods.

Maier, N. R. F., and Schneirla, T. C. Principles of Animal Psychology. McGraw-Hill Book Co., Inc., 1935. Pg. 85, paragraph No. 1.

No. 9

Scoring Key: Moral Other Choices: 2nd - Poetic
3rd - Mythical

The method contributes to the field of treatment the concept of limitations, which is conspicuously lacking in other modes of analyzing or diagnosing behavior. It helps to divide more clearly those aspects which cannot be changed from those which we may be able to alter, and thus promotes a healthy realism which has not always been an accompaniment to clinical endeavor. It also assists in overcoming individual bias in regard to treatment. One worker may see for every child the virtue of group activity, another the need for sex education or the possibilities of therapy based on a deep relationship to the worker. The method here described tends to reduce such individual predilections by drawing attention to the variety of individual needs and of possible points of attack.

Rogers, Carl R. The Clinical Treatment of the Problem Child. Houghton Mifflin Co., 1939. Pg. 57, paragraph No. 2.

* * *

No. 10

Scoring Key: Religious Other Choices: 2nd - Legal

Two persons, however, after they have begun work, may discover that they are mutually restimulative--which is to say each is a pseudo-person in the other's engrams or one is restimulated (voice tone, incidents) by the other. This should be no bar to therapy. It has been overcome and therapy has gone forward despite the most severe restimulative circumstances. A common avoidance technique on the part of a subject is to claim the auditor restimulates him: it is not sufficiently important to stop therapy. It may be, however, that two people can enter a third into the chain and by one clearing the next considerably ease the tension. The triangular work plan, where no person is working on the person who is working him, is quite successful.

Hubbard, L. Ron. Dianetics. The Modern Science of Mental Health. New York: Hermitage House, 1950. Pg. 389, paragraph No. 2.

No. 11

Scoring Key: Metaphysical Other Choices: 2nd - Rhetorical
3rd - Fictional

Another primitive reaction to the first objects appears simpler and more comprehensible: the infant wants to put them into its mouth. It was hunger, repeatedly disturbing the peacefulness of sleep, which compelled the recognition of the outside world. The experience of satiation, which first banished this tension, then became the model for the mastery of external stimuli in general. The first reality is what one can swallow. Recognising reality originally means to judge whether something helps to gain satisfaction or whether it raises tensions, whether one should swallow it or spit it out. Taking-into-the-mouth or spitting-out is the basis for all perception, and in conditions of regression one can observe that in the unconscious all sense organs are conceived as mouth-like.

Fenichel, Otto. The Psychoanalytic Theory of Neurosis. W. W. Norton and Co., Inc., 1945. Pg. 37, paragraph No. 2.

* * *

No. 12

Scoring Key: Moral Other Choices: 2nd - Critical

The bases for choice among factorial methods which have just been enumerated are all intrinsic to factor analysis. Yet one can conceive of a system of factor analysis which is efficient, mathematically elegant, self-consistent from experiment to experiment, and entirely solipsistic in ultimate outcome. The last word on the usefulness of factor analysis, one method against another or factor analysis as a whole, must come from outside factor analysis. If factor analyses enable us to predict nothing but future factor analyses, then the world will be no poorer for forgetting all about it. If factor analysis facilitates some predictions extrinsic to itself, then we have a right to know what predictions and to ask how much these predictions are improved by knowledge derived from factor analysis.

Loevinger, Jane. "Intelligence" in Theoretical Foundations of Psychology. Ed. by Harry Helson, pub. by D. Van Nostrand Co., Inc., 1951. Ch. 12, page 583, paragraph No. 3.

PART II

RAW DATA FOR EXPERIMENTS I, II, III AND V. CHARLES MORRIS SCORING KEY
EXPERIMENT I.

TABLE XVIII

EXPERIMENT I: CHANGE SCORES WITH 1-1-4 SCORING

EXPERIMENTAL GROUP (MORRIS) N = 10				CONTROL GROUP (FEIGL) N = 18				
S[a]	Post-Test (O)	Pre-Test (I)	Change[b] Score (c)	S	Post-Test (O)	Pre-Test (I)	Change Score (c)	
HB	3	6	-3	MB	8	8	0	Feigl N = 9
JC	4	5	-1	JJ	3	3	0	
MC	7	4	3	MJ	7	1	6	
RD	11	9	2	KL	7	4	3	
RH	4	4	0	GS	3	8	-5	
CJ	7	6	1	ES	2	3	-1	
AM	9	6	3	LW	7	2	5	
FR	3	3	0	BWa	5	7	-2	
RR	3	4	-1	BWn	7	7	0	
CS	8	1	7					
				VB	6	9	-3	Osgood I N = 5
				JC	6	5	1	
				MC	3	2	1	
				AG	3	4	-1	
				CM	7	3	4	
				LC	3	7	-4	Osgood II N = 4
				WJ	5	4	1	
				AM	5	3	2	
				JS	8	8	0	

a. S = Subjects' Initials

b. c = Change Score = Post-test - Pre-test

TABLE XIX

EXPERIMENT I: 1-4-16 SCORING

EXPERIMENTAL GROUP (MERRIS) N = 10				CONTROL GROUP (FEIGL) N = 18				
S	Post-test (0)	Pre-test (1)	Time" (2)	S	Post-test (0)	Pre-test (1)	Time" (2)	
*[a]HB	12	18	19	*MB	26	26	19	Feigl N = 9
JC	10	14	13	JJ	9	9	10	
*MC	22	13	20	MJ	25	4	15	
RD	41	36	15	*KL	22	13	20	
RH	10	10	10	GS	12	32	7	
CJ	25	18	15	*HS	5	9	25	
*AM	33	18	17	LW	25	8	12	
PR	12	12	10	*BWa	17	25	18	
*RR	9	10	23	BWh	22	19	12	
*GS	23	4	17					
				VB	18	33	12	Osgood I N = 5
				JC	15	11	13	
				MC	9	2	10	
				AG	6	10	15	
				CM	25	9	15	
				*LC	6	22	17	Osgood II N = 4
				WJ	17	10	5	
				*AM	17	6	16	
				JS	26	26	10	

a. * = a worker (16" or more spent on post-test)

TABLE XX

EXPERIMENT II: 1-4-16 SCORING

EXPERIMENTAL GROUP (MORRIS) N = 6				CONTROL GROUPS N = 17				
S	Post-test (0)	Pre-test (1)	Time" (2)	S	Post-test (0)	Pre-test (1)	Time" (2)	
BB	6	21	9	*AA	12	20	18	Feigl N = 6
MB	21	5	12	JB	4	4	15	
BR	16	33	12	DC	12	18	12	
BS	0	21	10	*MC	1	18	20	
*FS	37	17	16	FK	9	6	5	
*MS	34	5	22	FS	21	6	10	
				BS	8	8	15	
				PS	10	5	15	
				*EB	21	9	20	Osgood I N = 5
				JC	24	17	12	
				PE	16	18	15	
				GH	9	20	15	
				KS	9	8	7	
				JB	13	5	15	Osgood II N = 4
				SL	19	5	15	
				FM	13	11	10	
				AS	5	1	10	

a. * = a worker (16" or more spent on the post-test)

TABLE XXI

EXPERIMENT III: 1-4-16 SCORING

EXPERIMENTAL GROUP (MORRIS) N = 9				CONTROL GROUP (FEIGL) N = 9			
S	Post-test (0)	Pre-test (1)	Time" (2)	S	Post-test (0)	Pre-test (1)	Time" (2)
[a]JD	18	24	14	[a]LA	13	25	13
JH	18	52	15	VB	32	32	12
NK	26	5	12	SH	22	22	10
[c]*WW	35	34	16	MT	26	23	5
[b]*DH	37	21	22	[b]*[illegible]	34	7	19
*BH	68	50	27	RM	51	29	9
*RM	18	38	22	RS	49	48	10
*JH	42	17	21	*JB	25	36	13
*GM	22	25	17	JS	35	34	10

a. - Psychology students

b. - Physics students

c. - Philosophy students

d. * - a worker (16" or more spent on post-test)

TABLE XXII

EXPERIMENT I: 1-4-16 SCORING AND THE CHARLES MORRIS KEY

EXPERIMENTAL GROUP (MORRIS) N = 10				CONTROL GROUP (FEIGL) N = 18				
S	Post-test (0)	Pre-test (1)	Time" (2)	S	Post-test (0)	Pre-test (1)	Time" (2)	
*HB	21	22	19	*MB	7	7	19	Feigl N = 9
JC	10	10	13	JJ	18	5	10	
*MC	17	10	20	MJ	6	20	15	
RD	7	3	15	*KL	4	23	20	
RH	5	5	10	GS	8	4	7	
CJ	7	3	15	*HS	0	2	25	
*AM	22	22	17	LW	9	8	12	
PR	24	20	10	*BWa	25	17	18	
*RR	5	5	23	BWh	11	6	12	
*CS	25	21	17					
				VB	37	22	12	Osgood I N = 5
				JC	19	19	13	
				MC	0	1	10	
				AG	5	9	15	
				CM	22	6	15	
				*LC	17	8	17	Osgood II N = 4
				WJ	5	2	5	
				*AM	10	1	16	
				JS	11	11	10	

a. * - a worker (16" or more spent on post-test)

TABLE XXIII

EXPERIMENT V: 1-4-16 SCORING ("MORRIS TREATMENT" ONLY)

Pre-Test Group				Non-Pre-Test Group		
S	Post-test (0)	Pre-test (1)	Time" (2)	S	Post-test (0)	Time" (2)
EA	20	8	13	JB	16	10
JA	32	5	6	VB	4	11
MB	16	1	7	ABu	17	7
NB	17	1	12	DC	20	8
CD	20	8	9	SC	4	6
ME	32	0	7	FG	32	7
RF	4	8	6	GH	20	9
*WG	4	4	5	JM	5	17
*RHa	17	17	5	JP	20	10
LHo	32	8	8	CS	5	13
HJ	17	1	8	JS	8	9
JM1	32	16	8	ER	1	7
JMu	8	17	8	JT	5	7
WP	32	17	7	JW	5	6
AW	32	0	7			

* - Non-worker (5" or less spent on post-test)

APPENDIX D

THE MULTIPLE REGRESSION ANALYSIS

The technique employed was a multiple regression covariance analysis. The procedure was outlined for the writer by Dr. E. E. Cureton. However, attention should be called to the paper by Shen (72) where the generalized formula is fully developed and to the somewhat different approach of Snedecor (78, pp. 340-373). The equations can be derived from the defining formulae for the multiple correlation; i.e., this technique approaches covariance not in terms of adjusted sums of squares but rather in terms of multiple correlation, or--more specifically--*total* and *within group* multiple correlations. (The *between group* multiple correlation is not *computed*.) Here t^2 is equal to F. However, additional information can be more readily obtained than is possible with the usual F test alone; i.e., the nature of the control variables can be investigated and their differential predictive effects (inter- and intragroup) can be analyzed.

The computing formula is as follows:

$$t^2 = \frac{N_a + N_b - (n + k)}{1 - r^2} \left[\frac{(\bar{X}_a - \bar{X}_b)(1 - R^2)}{(\Sigma x_a^2 + \Sigma x_b^2)\left(\frac{1}{N_a} + \frac{1}{N_b}\right)} + r^2 - R^2 \right] \quad (1)$$

where "a" refers to one group and "b" to the other; $\bar{X}_a$, $\bar{X}_b$, Σx_a^2, and Σx_b^2 refer to the means and squared deviations of the two groups on the criterion (O) or posttest; the number of groups is "n"; and "k" is the number of predictors--in this case two, or pretest (1) and time spent on posttest (2); $N_a + N_b - (n + k)$ equals the degrees of freedom. "R" is the *total* multiple correlation, and "r" is the *within group* multiple

correlation ($R_{0.12}$ and $r_{0.12}$). To compute $R_{0.12}$ the groups are combined ($N = N_a + N_b$); the appropriate zero-order correlations (r_{01}, r_{02}, and r_{12}) are computed and substituted in the following formula:

$$R_{0.12}^2 = \frac{r_{01}^2 + r_{02}^2 - 2r_{01}\, r_{02}\, r_{12}}{1 - r_{12}^2} \qquad (2)$$

The within-group zero-order correlations are computed in the following manner and substituted in formula (2) to give $r_{0.12}^2$, the within-group multiple correlation:

$$r_{01} = \frac{\overset{M}{\Sigma x_0 x_1} + \overset{O}{\Sigma x_0 x_1}}{\sqrt{\left(\overset{M}{\Sigma x_0^2} + \overset{O}{\Sigma x_0^2}\right)\left(\overset{M}{\Sigma x_1^2} + \overset{O}{\Sigma x_1^2}\right)}} \qquad (3)$$

where "M" represents the group which had the "Morris Treatment," and "O" stands for "Others" or the groups which had the "Feigl Treatment" and the "Osgood Treatment."

The procedure outlined above involves determining the correlations of the predictors (pretest score and amount of time spent on the post-test) with the criterion of posttest score. The nature of the technique is to correct for these differences, with a t value ultimately being produced.

A breakdown of the coefficients of correlation and significance tests for the three experiments of Chapter V will be found in Tables XXIV, XXV, and XXVI.

Since the differences between the criterion means are so small, the last factor in formula (1) often determines whether the multiple regression t is negative or not; the t is positive only for

Experiment III, and it is only here that $r_{0.12}^2$ is larger than $R_{0.12}^2$. (Recall that with the classical t tests of the mean differences between obtained change scores, the only negative t is given by Experiment III; it is only here that the mean change score is larger for the control (Others) group.

For reasons advanced in Chapter V, the implied curvilinearity of regression was not tested or corrected in any way. But the intragroup correlations and means give us some idea of how the separate regression lines for the two predictors (pretest score and time spent on posttest) would look. The pretest predictor would not be linear for either group--those Ss who started high would regress downward if their score was the result of chance factors primarily; but if not primarily the result of chance factors, they could partially overcome the effects of regression--particularly those in the Morris group as the "worker" analysis shows. The regression line for time spent on the posttest would be almost linear for both experimental and control groups, but would have a fairly steep upward direction for the experimental group and a moderate downward slope for the control group. The multiple regression covariance tests of significance have already demonstrated what happens when the above factors are combined into a common (assumed to be linear) regression line.

Doubtless other interpretations of the multiple covariance analyses are possible. At one time I debated the possibility of omitting these analyses altogether since they only cloud the issue. However, they did tell us something of the nature of the predictor variables and the group differences, and--perhaps more important--they give a dramatic

demonstration of the pitfalls of covariance with the attendant likelihood of gross misinterpretation of the statistic and thus of the experimental results.

TABLE XXIV

EXPERIMENT I: INTERGROUP COEFFICIENTS OF CORRELATION[a]

		1	2	
Within Group:	0	.273	-.033	$r_{0.12}^2 = .084$
	1		-.441	$r_{0.12} = .290$
Total:	0	.337	.021	$R_{0.12}^2 = .140$
	1		-.383	$R_{0.12} = .374$

N Morris = 10
N Others = 18
N = 28 df = N-4 = 24

$t = -1.122$

$P = .275$ (two tails)

$F_o = 2.186$

$n_1 = 9$

$n_2 = 17$

$P > .05$

Means and Variances

	Morris	Others
$X_0 =$	19.70	16.78
$s_0^2 =$	121.79	55.71
$X_1 =$	15.30	15.22
$X_2 =$	15.90	13.94

[a] 0 - Post-test (criterion)

1 - Pre-test (predictor)

2 - Time spent on post-test (predictor)

TABLE XXV

EXPERIMENT II: INTERGROUP COEFFICIENTS OF CORRELATION[a]

		1	2	
Within Group:	0	-.078	.177	$r_{0.12}^2 = .033$
	1		-.609	$r_{0.12} = .181$
Total	0	-.073	.338	$R_{0.12}^2 = .119$
	1		.014	$R_{0.12} = .345$

N Morris = 6

N Others = 17

N = 23

t = -1.179

df = 19

P = .253

$F_0 = 5.345$

$n_1 = 5$

$n_2 = 16$

$P < .01$

Means and Variances

	Morris	Others
$\bar{X}_0 =$	19.00	12.12
$s_0^2 =$	218.40	40.86
$\bar{X}_1 =$	19.00	12.12
$\bar{X}_2 =$	13.50	13.47

[a] 0 - Post-test (criterion)

1 - Pre-test (predictor)

2 - Time spent on post-test (predictor)

TABLE XXVI

EXPERIMENT III: INTERGROUP COEFFICIENTS OF CORRELATION[a]

		1	2	
Within Group:	0	.270	.407	$r_{0.12}^2 = .218$
				$r_{0.12} = .467$
Total:	0	.269	.297	$R_{0.12}^2 = .145$
	1		.106	$R_{0.12} = .381$

N Morris = 9
N Others = 9
N = 18
t = 1.111
df = N-4 = 14

P = .286 (two tails)
$F_0 = 1.780$
$n_1 = 8$
$n_2 = 8$
$P > .05$

Means and Variances

	Morris	Others
$\bar{X}_0$ =	31.56	31.89
s_0^2 =	269.03	151.11
$\bar{X}_1$ =	29.56	28.44
$\bar{X}_2$ =	18.44	11.22

[a] 0 - Post-test (criterion)
1 - Pre-test (predictor)
2 - Time spent on post-test (predictor)