

**SELF-CONCEPT, CONTENT AND EXPRESSION IN
A-T-P DRAWINGS**

by John A. Machry

Thesis presented to the Faculty of
Psychology and Education of the
University of Ottawa as partial
fulfillment of the requirements
for the degree of Doctor of
Philosophy in Clinical Psychology



Ottawa, Canada, 1966

UMI Number: DC53454

INFORMATION TO USERS

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleed-through, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.



UMI Microform DC53454
Copyright 2011 by ProQuest LLC
All rights reserved. This microform edition is protected against
unauthorized copying under Title 17, United States Code.

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

ACKNOWLEDGEMENTS

This thesis was prepared under the supervision of Associate Professor Gilles Chagnon, B.A., B.Ph., M.Ps., of the Faculty of Psychology and Education of the University of Ottawa.

The writer is deeply indebted to Professor Chagnon for his encouragement, contributions and many helpful criticisms. The writer acknowledges with appreciation the assistance of many fellow students in data collection and analysis notably, Michael Cavanagh, George Meaney, T. Lin, Michael McCarrey and Andrew Wozny. Appreciation is also extended to Father Zutterman and Father Leonard Murphy for their tireless efforts in acting as judges.

CURRICULUM STUDIORUM

John A. Machry was born August 1, 1932, in Toronto, Canada. He received his Bachelor of Arts (Psychology) from the University of Toronto in 1957. He received his Masters of Arts degree in Psychology from the University of Toronto, Canada, in 1960. The title of his thesis was Effect of Drive Level Upon the Rate of Learning for Two Initial Response Strengths.

TABLE OF CONTENTS

Chapter	Page
INTRODUCTION.	vii
I.- REVIEW OF THE LITERATURE.	1
1. Hypothetical Considerations	2
2. Levels Hypothesis	21
3. Stimulus Value and S.D.	28
4. Summary and Basic Hypotheses	42
5. Experimental Hypotheses, General Form	47
II.- EXPERIMENTAL DESIGN	49
1. Whole and Copied Drawing Definitions	49
2. Validity and Reliability of S.D.	53
3. Sample	69
4. Judges	71
5. Procedure	72
6. Analysis of Data	76
7. Hypotheses, Null Form	80
III.- PRESENTATION OF RESULTS	83
1. Reliability of Subjects' Ratings	83
2. Reliability of Judges' Ratings	89
3. Wilcoxon Z Scores - Subjects	94
4. Wilcoxon Z Scores - Judges	96
IV.- DISCUSSION OF RESULTS	99
1. Interpretation of Subjects' Results	99
2. Interpretation of Judges' Results	111
3. Evaluation of Results	115
4. Summary and Conclusions	120
BIBLIOGRAPHY.	124
Appendix	
1. DEFINITIONS	126
2. INSTRUCTIONS TO SUBJECTS FOR SEMANTIC DIFFERENTIAL RATINGS.	132
3. SAMPLE OF ONE OF THE 5-SCALE ORDERS OF THE 15 SCALES USED FOR SEMANTIC DIFFERENTIAL RATINGS . .	134
4. FACTOR LOADINGS OF THE 15 SCALES.	135
5. ABSTRACT OF <u>Self-Concept, Content and Expression in H-T-P Drawings.</u>	136

LIST OF TABLES

Table	Page
I.- Probability of Obtaining Given Deviations from Test-Retest on Subjects' Ratings of Self-Concept.	84
II.- Probability of Obtaining Given Deviations from Test-Retest on Subjects' Ratings of Whole H-T-P's	86
III.- Probability of Obtaining Given Deviations from Test-Retest on Subjects' Ratings of Copied Drawings	87
IV.- Probability of Obtaining Given Deviations from Test-Retest on Judge I's Ratings of Subjects' Self-Concept (Factor score reliability)	90
V.- Probability of Obtaining Given Deviations from Test-Retest on Judge II's Ratings of Subjects' Self-Concept (Factor score reliability)	91
VI.- Probability of Obtaining Given Deviations from Test-Retest on Judge III's Ratings of Subjects' Self-Concept (Factor score reliability)	92
VII.- Average Absolute Deviations and Related Probabilities of Occurrence for Factor Scores Based on Judges' Ratings.	93
VIII.- Source of T's, T's, M's, Means, S.D.'s, Wilcoxon Z Scores and Associated Probabilities Based on Pairs of D Scores Obtained from Comparisons Made on Subjects' Ratings	95
IX.- Source of T's, T's, M's, Means, S.D.'s, Wilcoxon Z Scores and Associated Probabilities Based on Pairs of D Scores Obtained from Comparisons Made on Judges' Ratings	97

LIST OF FIGURES

Figure	page
1.- Diagrammatical Representation of D Scores Obtained from Subjects' Ratings.	78
2.- Diagrammatical Representation of D Scores Obtained from Ratings made by Each Judge	79

INTRODUCTION

One of the least satisfying aspects of testing has been widely recognized as a lack of theory behind the use of individual tests. While this difficulty can be readily said of tests in general, it is perhaps more prevalent of the projective techniques in particular. These techniques, until quite recently, have been used with a zeal that transcends the knowledge that was available about them. Interpretations based on their data were seldom required to meet the criteria demanded of empirically based data. In the current psychological literature there is a growing number of research authorities, such as Zubin,¹ who are of the opinion that the time is long overdue for projective techniques to either live up to the claims made of them or for new tools to be devised.

Projective drawing techniques are in no better position than the majority of projective techniques. In fact, there may be more urgent need for research investigation of these techniques than for others. This point is supported by Levy² when he states that among psychological tests,

1 Josephy Zubin, Leonard D. Kron and Florence Schumer, An Experimental Approach to Projective Techniques, New York, Wiley, 1965, p. 610.

2 Sidney Levy, "Figure Drawing as a Projective Test", in Lawrence E. Abt and Leopold Bellak (eds.), Projective Psychology, New York, Grove Press, 1959, p. 258.

drawing techniques are particularly open to misuse. Frequency of use is another factor that accentuates this need. Sundberg³ compiled a list of thirteen tests most frequently used in clinical practice across the United States. Three drawing techniques, D.A.P., Goodenough Draw-a-Man Test and the H-T-P were among those tests most frequently used.

The need for research into the use of H-T-P has been pointed out by Hammer⁴ who states:

[...] the H-T-P is presently embarking on the thorny path of being transformed from an empirical and practical technique to a scientifically controlled and experimentally rooted method of personality analysis. The chief need at this time is for further validation studies.

The experiment reported here was conducted in an attempt to test the implications that two hypotheses--the one by Hammer made with specific reference to the H-T-P, the other by Allport made at a much more general level--have for H-T-P test behaviour. Chapter one considers the two hypotheses and the relevant research literature. In chapter two the experimental design is described. The results are presented in chapter three and the summary of results, the conclusions reached and suggestions for further research are indicated in chapter four.

³ Norman D. Sundberg, "The Practise of Psychological Testing in Clinical Services in the United States", in American Psychologist, Vol. 16, No. 2, February 1961, p. 79-83.

⁴ Emanuel F. Hammer, The H-T-P Clinical Research Manual, Beverly Hills, Cal., Western Psychological Services, 1955, p. 1.

CHAPTER I

REVIEW OF THE LITERATURE

The theoretical considerations and the research leading up to the experiment reported here are presented in this chapter in five parts. In part one, the hypothetical statements made by Hammer and Allport are considered. Part one also contains a discussion of the positions Wolff and Maslow held regarding the expressive aspects of behaviour. To conclude part one the implications that Osgood's theory of meaning has for certain aspects of the House-Tree-Person test¹ (hereafter referred to as the H-T-P) are presented. Part two presents the levels hypothesis applied to the H-T-P and contains a discussion of some relevant experimental findings. The third part outlines research relevant to using the semantic differential technique² (hereafter referred to as the S.D.) as an approach to the experimental investigation of the stimulus value of projective technique stimuli. The research using the S.D. to investigate aspects of the H-T-P is presented and analyzed to complete part three.

1 John N. Buck, "The H-T-P Test", in Journal of Clinical Psychology, Vol. 4, No. 2, April 1948, p. 151-159.

2 Charles E. Osgood, George J. Suci, Percy H. Tannenbaum, The Measurement of Meaning, Urbana, Illinois, University of Illinois Press, 1957, 342 p.

A summary of the theoretical implications and an outline of the reasoning followed in planning this experiment, are considered in part four. Part five contains a general statement of the hypotheses formulated and tested. Part five completes chapter one.

1. Hypothetical Considerations.

Allport and Hammer appear to hold contradictory views regarding expressive aspects of behaviour. In direct reference to the H-T-P Hammer³ states:

[...] it is the content of the drawings, more so than the expressive movements employed in the drawing, which express the more unconscious qualities that exist at the heart of personality.

In a general discussion of expressive behaviour Allport⁴ distinguishes coping and expressive as two basic aspects of behaviour. Coping behaviour is characterized as intentional, goal-directed, formally elicited, more readily controlled than expressive behaviour and typically conscious. Expressive behaviour is viewed as being spontaneously emitted, below the threshold of consciousness and reflective of the deeper personal structure.

³ Emanuel F. Hammer, The Clinical Application of Projective Drawings, Springfield, Illinois, Charles C. Thomas, 1958, p. 173.

⁴ Gordon W. Allport, Pattern and Growth in Personality, New York, Holt, Rinehart and Winston, 1961, p. 462-463.

Hammer's hypothetical statement is made with specific reference to the H-T-P and therefore is applicable in its present form to H-T-P testing drawing behaviour only. Allport's statements, however, are made on a much more general level but are at least potentially applicable to behaviour under varying conditions which could include H-T-P test drawing behaviour. Therefore, when Allport's statements are applied specifically to H-T-P test drawing behaviour, it seems that Allport and Hammer hold divergent views regarding the potency that expressive aspects of H-T-P test drawing behaviour have for the understanding of personality. Allport seems to imply that because the expressive aspects reflect the deeper structure of personality, they are more potent than coping aspects for the understanding of personality. Hammer appears to be implying the opposite. He states that the content aspects more than the expressive aspects reflect the unconscious qualities that are at the heart of personality. Since he sees the qualities that content aspects reflect as being at the heart of personality, Hammer seems to be implying that they are more potent than expressive aspects for understanding the basic dynamics of personality. In addition, he may be implying that since content reflects the more unconscious qualities, the expressive elements reflect the more conscious qualities. This latter possibility is not crucial for the experiment reported

here but it assists in specifying the precise point at which Allport and Hammer differ for Allport states that the expressive elements themselves are typically below the conscious level of awareness. Moreover, there are some aspects of their disparate views that seem somewhat similar. Before giving them consideration, the implications generated from Hammer's hypothesis that contributed to the development of the hypotheses tested in this experiment, will be presented.

To understand Hammer's position better, his distinction between the content and the expressive elements of H-T-P drawings will be stated. Hammer⁵ sees content as comprising: 1) postural tone of figures; 2) facial expression of figures; 3) emphasis placed on individual aspects such as chimney, windows, doors, branches, roots; 4) individual parts of the figure drawn including clothes and accessories. Hammer⁶ considers the expressive or structural elements as consisting of: 1) size of the drawings; 2) pressure of the lines; 3) quality of the lines; 4) placement of drawings on the pages; 5) exactness; 6) degree and areas of completion and detailing; 7) symmetry; 8) perspectives; 9) proportions; 10) shading; 11) reinforcement; 12) erasures.

5 Ibid., p. 60.

6 Hammer, Op. Cit., p. 59-60.

REVIEW OF THE LITERATURE

2

These considerations did present some difficulties for designing the experiment reported here. For example, the first two components of content, namely, postural tone of figures and facial expressions of figures drawn, could form part of the expressive aspects as readily as Hammer considers them to form part of content. The third component of content, the emphasis placed on individual aspects of the drawings, could also perhaps be more appropriately viewed as an element of the expressive aspects. It is unfortunate that the writer could not find among Hammer's numerous publications a clearer, more precise consideration by him of content and expressive elements. Such statements as the emphasis placed upon individual parts, for example, presents obstacles for operational defining of the content and expressive elements and therefore to testing experimentally, hypotheses related to them. That is, Hammer does not specify, for example, what type of emphasis he considers to be content as opposed to the emphasis that reinforcements, shading and erasures can give to individual parts of the drawings. These latter types of emphasis he includes among the expressive elements. In an attempt to overcome the confusion generated by Hammer's lack of precision in formulating his statements, the content aspects of the three H-T-P drawings were broadly conceived of as the concepts drawn, namely, house, tree and figure drawings. Similarly, expressive

elements were considered to be that particular combination of any of the twelve elements Hammer sees as comprising the expressive elements, by which each drawing is produced. A more complete account of all the definitions utilized in this experiment is contained in Appendix 1.

When Hammer states that the content of H-T-P drawings is expressive of the more unconscious qualities that exist at the heart of personality, several meanings seem implied. One meaning is that subjects tend to express the unconscious qualities of their personalities through the content aspects of their drawings rather than through the expressive movements they utilize in producing their drawings. Another possible meaning is that clinicians, when attempting to interpret H-T-P drawings, can ferret out clues to the unconscious qualities of personality more readily from the content of H-T-P drawings than from the expressive elements. Hammer's⁷ reference to content as content analysis lends support to this latter meaning. Nevertheless, Hammer may be referring to two kinds of analyses of H-T-P drawings, one that is based essentially on the expressive elements and one that is based mainly on the content aspects of the drawings. Moreover, Hammer could be implying that the clinician, when attempting to interpret H-T-P drawings,

7 Ibid., p. 60.

conceptually dichotomizes the drawings into two groups of elements, one group made up of the content aspects and the other group made up of the expressive elements, and that the clinician focuses on the content somewhat exclusively to find clues as to the unconscious qualities. In addition, Hammer may be implying that when so dichotomizing, the clinician is not hindered in his interpretative attempts by any effects from the content and expressive elements combining to produce an interaction influence on the process of analysis.

Hammer assumes that the unconscious qualities rather than the conscious qualities are more significant for the understanding of personality because of their proximity to the core of personality. That is, Hammer seems to imply that the unconscious qualities are more intimately involved in those aspects of personality that are more influential in determining the enduring, dominant characteristics of personality. It would follow therefore that the unconscious qualities of personality for Hammer reveal its uniqueness, its individuality. If he is so implying, then he needs to assume that the same qualities of personality are always unconscious and neither the conscious nor the unconscious qualities ever vary significantly along a hypothetical conscious-unconscious dimension. Conscious and unconscious qualities of personality then would belong to mutually

exclusive categories rather than different personality qualities becoming unconscious under certain conditions and conscious under other conditions. In summary then, Hammer appears to assume that certain given qualities of the personality are by nature unconscious and always unconscious and that these qualities more than any other qualities of the personality are revealing of its very essence.

Hammer's implication of a conceptual dichotomy between the expressive and content aspects of H-T-P drawings cannot be reified in H-T-P drawings. These two aspects of H-T-P drawings are inextricably bound together and it is inconceivable that the content, i.e., the roof of a drawing of a house for example, could be properly separated from the manner in which a person goes about drawing that roof! That is to say, the content aspects of H-T-P drawings must be either there or not there in a particular way, by some combination of expressive elements. Furthermore, the emphasis on each part of the content will vary considerably across subjects whereas all subjects to whom an H-T-P is administered produce drawings of the same concepts. However, the type of house, tree or figure drawn varies considerably among subjects as do the number, kind and relative emphasis of expressive elements with which each subject produces his drawings. This may be why Hammer synonymously refers to

expressive elements as structural elements in that expressive elements are those combined to structure the content.

Allport^{8,9} also seems to be commenting on this inextricability of content and expressive aspects of H-T-P test drawing behaviour when he states that every single act of behaviour has both its expressive and coping aspects and that no act of behaviour is ever purely expressive or purely adaptive. Contrary to what Hammer seems to imply, Allport suggests that examiners when analyzing H-T-P drawings cannot achieve such a cognitive dichotomy, for he defines expressive movement as individual differences in the way in which adaptive aspects of behaviour are performed. Incidentally, Allport uses the terms adaptive and coping interchangeably. Both terms refer to the conventionalized, intentional, conscious and goal-directed aspects of behaviour. In his earlier writings Allport tends to use the term adaptive when referring to this aspect of behaviour whereas in his latest writings he prefers using the term coping.

Returning to points of similarity between their disparate positions, Hammer and Allport seem to agree on the way they view expression. Allport¹⁰ is quite clear in

⁸ Allport, Op. Cit., p. 462.

⁹ Ibid., p. 466.

¹⁰ Ibid., p. 462.

defining expressive aspects as the "how" or style of behaving; the individual's manner of performing adaptive traits. Hammer¹¹ considers the expressive aspects of behavior to be the "how" of behaviour. Further clarifying his distinction between expressive elements and content aspects within H-T-P drawings, Hammer uses as an example the way in which the meaning of a spoken sentence is conveyed. He points out that the meaning of any spoken sentence depends both on the words used to make up the sentence as well as the manner in which a word (or words) is emphasized. By this analogy Hammer emphasizes the similarity between the expressive elements of H-T-P drawings and the particular emphasis given to a particular word (or words) in a spoken sentence and the similarity between the specific words used in such a sentence and the content aspects of the H-T-P drawings.

Another similarity may be seen in their concepts of coping and content. These concepts may not be synonymous but it appears that content could be subsumed under coping. Content can be said to be a type of coping behaviour. For example, Hammer includes in his definition of content individual parts of the drawings such as roof and chimney. Such parts of the drawings represent aspects of the environment which through common cultural learning become

¹¹ Hammer, Op. Cit., p. 60.

standardized and conventionalized as to the meanings they symbolize. It is reasonable to consider that this learning of their meanings and the learning of their functions, as parts of wholes, is included in a more general scheme, that of learning to cope with the demands of the environment. Such learning can be said to contribute to learning the most appropriate response that most adequately copes with a particular stimulus. Therefore both content and coping can be said to be learned as types of behaviour emitted in response to specific stimuli. The roof in the drawing of a house, for example, is emitted in response to having learned that as a necessary part of a house, the roof carries out a "vital" function. That is, a house cannot carry out its function adequately without a roof, and the complete concept of a house necessarily includes a roof. In addition, the drawing of a house is produced as the result of a set of emitted responses that are elicited in response to specific verbal stimuli, the instructions. Therefore it seems reasonable to view the content aspects of the H-T-P drawings as a result of learning to cope with the instructions of the H-T-P test. Moreover, how one is to produce the drawings is not specified in the instructions and therefore the manner in which subjects produce (the expressive elements they utilize) the content aspects of the H-T-P drawings is the prerogative of each individual subject. It is likely

that expressive elements would vary regardless of attempts to standardize them. That the particular kind of house, tree or figure drawn in response to the instructions and the particular combination of expressive elements used, along with their relative emphases, do vary considerably within the drawings of different subjects, is a well-documented facet of H-T-P test drawing behaviour. This seems to be the kind of variations Allport¹² is referring to when he states that expressive movement is the individuality in one's manner of performing adaptive behaviour.

In passing it can be pointed out that Maslow¹³ holds a view of expressive behaviour essentially the same as Allport's. He, like Allport, makes a distinction between the expressive and coping components of behaviour. Maslow compares coping and expression stating that expression is most often unlearned, more often uncontrollable, non-goal directed, and more often not conscious. On the other hand, coping he views as purposive, most often learned, more easily controlled and typically conscious.

Wolff also holds a view about the expressive movement aspects of behaviour that appears to be relevant to this

12 Allport, Op. Cit., p. 491.

13 A.H. Maslow, Motivation and Pure Reality, New York, Harper, 1954, p. 180.

discussion. Wolff¹⁴ and his colleagues have carried out research on the various forms of expression. One of the main tasks they set for themselves was to investigate the problem of consistency in the expressive movement aspects of behaviour. In studying this problem, Wolff and his colleagues compared judgments of one form of expression of a given subject with judgments of the subject's other forms of expression. Moreover, the relationship of judgments by others to self-judgments was investigated as was the relationship between self-judgments given when the judge was aware and when the judge was unaware that the form of expression was his own. The significance of the results obtained was tested by comparing the observed results with those that could be expected by chance. In spite of being faced with the problem that all forms of expression are not equally revealing of the same personality traits in all people, Wolff presents evidence that cogently supports his hypothesis of consistency among the various forms of expressive behaviour and similarities in what expressive aspects reveal about the personality. Therefore Wolff views personality as having an essential unity which is reflected in the consistency of the various forms of expressive movement and also sees expressive movements as reflecting the more

¹⁴ Werner Wolff, The Expression of Personality, New York, Harper Brothers, 1943, xiv-334 p.

unconscious qualities of the personality rather than the conscious and preconscious qualities.

Wolff and Precker¹⁵ make some interesting statements that also seem relevant to this discussion. They try to answer two questions. One question is directed at whether or not expressive behaviour is consistent and the second question concerns the relationship between expressive behaviour and the depth of personality. Wolff and Precker postulate that expressive movements are not chance occurrences but are consistent under different environmental situations and reflect the basic motivations of the individual. There is experimental evidence that supports their position in this matter.

They hold that expressive behaviour is generated out of the inner dynamics of personality and consists of two aspects which are called the intentional and the affective. They attribute the concrete form of expressive movement to the intentional aspect which they view as being basically adaptive by nature. Style of behaviour they attribute to the affective aspect. To further clarify their consideration of expressive behaviour, which they state is one's style of

¹⁵ Werner Wolff and Joseph A. Precher, "Expressive Movement and the Methods of Experimental Depth Psychology", in Harold H. Anderson and Gladys I. Anderson (eds.), An Introduction to Projective Techniques and Other Devices for Understanding the Dynamics of Human Behaviour, New York, Prentice Hall, Inc., 1951, p. 457-497.

response, Wolff and Precker differentiate adaptation, projection and expression. They see adaptive behaviour as that aspect of behaviour which is determined by the material with which one works. Projective behaviour they view as the ascribing of inner needs and qualities to others without conscious awareness of the process. Unlike adaptive behaviour projection is inversely proportional to the degree to which test stimuli are structured. Expressive behaviour, however, is said to be the individual's style of response. Moreover, these authors emphasize that each of these three types of behaviour seldom appear alone. There are varying relationships between the relative amounts that each contributes to the final effect in the behaviour. They emphasize this point by stating that none of the projective techniques currently used to study personality demand purely projective or purely expressive or purely adaptive behaviour. Rather, each technique calls for a combination of two or all three aspects of behaviour, each aspect being involved in a varying degree depending on the technique in question.

Wolff and Precker make a further distinction between expression, form and content. Expression they state is the style of performance whereas form is the end product of performance which is usually measurable. Content, for these authors, refers to the actual experiences of the behaving organism. In passing it can be noted that Wolff

and Precker's distinction between adaptive, projective and expressive aspects of behaviour is much like the distinction Allport¹⁶ makes between projective technique behaviour and expression. Moreover, Allport's position regarding the inextricability of behavioural aspects in behaving is supported by Wolff and Precker. In addition, Wolff and Precker appear to agree with Allport's contention that expressive movements are potentially very fruitful in understanding personality and that expressive movements tend to be more unconscious than conscious.

The views of Maslow, Wolff and Precker regarding expressive aspects of behaviour, indicate that Allport's position has support outside his own writings. These additional opinions and statements are useful ways of thinking about the distinctions held regarding the various aspects of behaviour but demonstrate that theoretical interest, particularly regarding the expressive aspects of behaviour has not developed beyond an elementary level. Their inclusion here was intended to help clarify the distinctions Allport has made between the expressive and coping aspects of behaviour as well as clarify the distinction made in this experiment between the expressive and content aspects of H-T-P test drawing behaviour. Both Allport and Wolff have

16 Allport, Op. Cit., p. 493.

summarized much of the research in expressive movement that has been done in North America. Their conclusions are generally, that there is a basic unity within personality, that expressive movements have consistency, that expressive aspects of behaviour reflect personality qualities that are potentially fruitful for the understanding of personality and that expressive aspects are typically below the threshold of conscious awareness. These conclusions are unfortunately not integrated into a systematic theory of expressive movement. Indeed, Allport¹⁷ has pointed out contemporary American psychology's lack of a comprehensive theory of expression. This deficiency has resulted for one thing in a general neglect of any systematic experimentation, notably into the relationship between expressive movement aspects of behaviour and personality. In addition, this lack of theory did present some difficulties for the experiment being reported here. For one thing it precluded the possibility of this experiment making a theoretical contribution by testing a hypothesis derived from the implication of such a theory. Nevertheless, it is possible that the results of this experiment could make a contribution to theory in two ways. First, it might suggest what role, for the understanding of personality, expressive aspects of behaviour could be given in a

17 *Ibid.*, p. 466.

general theory of expression. Secondly, it could indicate the role expressive elements could be assigned in a theory of H-T-P test drawing behaviour. Such a role might include if the qualities of personality that expressive elements reflect are more conscious than those reflected by content aspects. In addition, how potent the expressive elements are compared to the content aspects of the H-T-P for the understanding of personality might also be indicated. Finally, its contribution to a theory of H-T-P test drawing behaviour could be to a theory that eventually may be developed. Such a theory may already be conceptualized by such an authority as say Hammer, even though it may not be explicitly formulated in written form.

This current lack of such theories also presented difficulties for predicting the direction hypotheses in this experiment should take, as well as for interpreting its results. Without theory it is more difficult to interpret results as they cannot be integrated into a meaningful frame of reference. The decision that theory deficiency necessitated, regarding the formulation of the hypotheses tested in this experiment, is discussed in part five of this chapter. However, one possible source available for providing the hypotheses with direction is Osgood's¹⁸ theory of meaning, a consideration of which follows next.

18 Osgood, et al., Op. Cit., p. 5-10.

Turning to an outline of that theory it can be seen that to provide a theoretical framework for meaning, Osgood has adapted a somewhat Pavlovian learning theory model. Osgood views meanings as types of conditioned responses that are established as a result of a conditioning process. Meanings as conditioned responses, for Osgood, depend upon the occurrence of an intervening mediational reaction that is self stimulating. The mediational reaction can and does have partial identity with the unconditioned response and therefore is associated with it. It is this reactional, representational, mediation process that Osgood identifies as meaning.

The conditioning process Osgood refers to would occur as follows. Imagine a complex stimulus S_1 repeatedly eliciting a response R_1 in the presence of a sign I_1 . If this happens continuously sign I_1 will tend to become a conditioned stimulus to an implicit response E_1 which is some reduced covert portion of R_1 . The implicit response E_1 can be said to be the meaning of sign I_1 and to possess mediating stimulus properties which may elicit other overt responses appropriate to the meaning of sign I_1 . Therefore implicit response E_1 can be said to be both a mediating response and to possess stimulus properties. Osgood symbolizes this process as $rs \rightarrow sm$.

While this theory is not beyond criticism many psychologists would readily accept this theory as at least a useful way of thinking about meaning and how meanings become established. The theory is potentially useful in this experiment for interpreting the results as well as providing the hypotheses with direction. To this end it would be tempting to identify the content aspects of the H-T-P with what has been called the denotative aspects of meaning and the expressive elements with what has been called the connotative aspects of meaning. Then, if as Osgood¹⁹ and Kamano²⁰ maintain, the S.D. is measuring mainly the connotative aspects of meaning, when people rate concepts of the S.D. they are assigning to these concepts the connotative meaning these concepts have for them. Therefore, it could be predicted that subjects would rate their H-T-P's containing more expressive elements closer in semantic distance to their self-concepts than their H-T-P's containing fewer expressive elements. Such a prediction is, however, quite tenuous. For one thing, there

19 Charles E. Osgood (ed.), "Psycholinguistics", Supplement to the Journal of Abnormal and Social Psychology, Vol. 49, No. 4, Part 2, October 1954, p. 183.

20 Dennis K. Kamano, "Symbolic Significance of Rorschach Cards IV and VII", in Journal of Clinical Psychology, Vol. 16, No. 1, January 1960, p. 52.

is little agreement as to whether the S.D. measures mainly those aspects of meaning referred to as connotative or those called denotative. Both Osgood²¹ and Kamano²² maintain the S.D. measures essentially the former, whereas Carroll²³ holds the S.D. to measure one kind of the latter. Flavell²⁴ prefers to by-pass using these terms since there is no consensus of agreement as to their meanings and therefore their use tends to produce only confusion. Nevertheless, the implications for giving the hypotheses direction and for interpreting the results of this experiment that, this use of Osgood's theory can have, will be mentioned again in chapter four. The application of the levels hypothesis to the H-T-P and its relationship to this experiment needs comment. This follows in part two.

2. Levels Hypothesis.

The levels hypothesis,²⁵ which postulates that the component parts of personality are integrated at different

21 Osgood, Op. Cit., p. 183.

22 Kamano, Op. Cit., p. 12.

23 John B. Carroll, "Review of, Osgood et al., The Measurement of Meaning", in Language, Vol. 35, 1959, p. 76.

24 John H. Flavell, "Meaning and Meaning Similarity: A Theoretical Reassessment", in Journal of General Psychology, Vol. 64, 1961, p. 317.

25 Hewitt Sanford, "Surface and Depth in the Individual Personality", in Psychological Review, Vol. 63, No. 6, November 1956, p. 349-359.

levels within the personality, has been applied to the H-T-P. Buck,²⁶ in the original journal article in which he introduced the H-T-P, implies that it taps the personality at different levels. He states that subjects project their personality attributes much more readily into their drawing of the tree than they do into their person or house drawings since these latter two drawings tend to make subjects aware that they are projecting. He here implies that the tree reflects the more unconscious qualities of personality.

Other authors share this hypothesis. In an article comparing certain signs and patterns of the Rorschach with given signs and patterns on the H-T-P, Landisberg,²⁷ hypothesizes that the tree drawing reflects the more basic, long standing feelings and self-attitudes within the personality, much more so than the house and tree drawings. Landisberg supports this hypothesis by arguing that only significant changes produced by such dramatic conditions as deep therapy will produce changes in tree drawings. She adds that changes in the house and figure drawings are more sensitive to the more superficial or more temporary aspects of personality.

²⁶ Buck, Op. Cit., p. 151-159.

²⁷ Selma Landisberg, 'Relationship of the Rorschach to the H-T-P', in Journal of Clinical Psychology, Vol. 9, No. 2, April 1953, p. 132.

Hammer also applies the levels hypothesis to the H-T-P. Like Landisberg, Hammer²⁸ views the tree as tapping the more basic, the more enduring, the deeper feelings and attitudes within the personality. This is in spite of the fact that he considers the tree and person drawings as both representing a psycho-sexual self-concept. However, Hammer seems to imply that the tree drawing reflects the more unconscious aspects of self-concept whereas the figure drawing reflects the more conscious aspects.

There is some experimental evidence supporting the application of the levels hypothesis to the H-T-P. In one such study Cassel, Johnson and Burns²⁹ set out to investigate the possible influence the presence the examiner can have on the extent to which subjects project personality attributes into their H-T-P drawings. The subjects who took part in this investigation were selected from a sample of 130 employee applicants to a state school. The final sample involved were seventy-two subjects which comprised an examiner absent group and fifty-eight subjects who made up an examiner present

²⁸ Emanuel F. Hammer, "The Role of the H-T-P in the Prognostic Battery", in Journal of Clinical Psychology, Vol. 9, No. 2, October 1953, p. 371-372.

²⁹ Robert H. Cassel, Anna P. Johnson and William H. Burns, "Examiner, Ego Defense and the H-T-P Test", in Journal of Clinical Psychology, Vol. 15, No. 2, April 1958, p. 157-160.

group. Every applicant as a matter of routine was sent to the psychology department of the school and each was administered an H-T-P individually. Each examiner would alternate being absent and present during test administration. Each author rated each drawing for the presence or absence of interpretable features. In analyzing the data, the authors found the examiner present group had a significantly smaller number of interpretable features on both the person and the house drawings. In addition, they found that the means differed significantly on both the number of interpretable features and the size of the drawings of the person and the house but not on the tree drawings. These findings are interpreted by the authors as evidence in support of the hypothesis that the tree drawing represents a deeper level of personality integration than either the person or house drawings.

However, certain aspects of the Cassel, Johnson and Burns' study could limit the significance of its results as support for the levels hypothesis as applied to the H-T-P. First, an implication of the levels hypothesis was not the main hypothesis tested in this experiment. Evidence supporting the levels hypothesis is somewhat of an incidental finding. Secondly, since subjects were selected out of job applicants, the sample could have been biased in an untypical way. Even though the significant differences between the

control and the experimental group believe such an effect, the margin by which the differences were significant is small. Also, since the three authors of the study administered the H-T-P's to subjects as well as completed ratings on subjects' drawings, they may have unwittingly influenced their results. The influence that experimenters can unintentionally and unwittingly have on the results of their experiments has been adequately dealt with by Kintz et al.³⁰

Another relevant study was carried out by Orgel³¹ who investigated the possibility of a relationship between the degree of social acceptance as demonstrated on a sociogram of children's peer group relationships and their rated social adjustment as reflected in the H-T-P. The subjects involved were seventeen boys and fifteen girls in grade two in an elementary school from middle class socio-economic home environments. In the first phase of the experiment subjects were asked to make three choices among their peers concerning whom they would most like to sit with in first, second and third choice order. They were also asked to indicate the peer

³⁰ B.I. Kintz et al., "The Experimenter Effect", in Psychological Bulletin, Vol. 63, No. 4, April 1965, p. 223-232.

³¹ Rita G. Orgel, "The Relationship of the H-T-P to a Sociometric Evaluation of a Group of Primary School Children in Determining the Degree of Social Acceptance", in Journal of Clinical Psychology, Vol. 15, No. 2, April 1959, p. 222-223.

to whom they did not want to sit beside. These data were recorded on a sociogram. In the second phase H-T-P tests were administered to all subjects. A thirty-seven point checklist was devised and used by a judge to evaluate from H-T-P drawings each subject's popularity potential. While the correlation of the overall H-T-P drawing scores with sociometric evaluation was only .271 and not significant at the .05 level of confidence, nevertheless the author found that the tree drawings had a negative correlation with sociometric evaluation of $-.173$. Moreover, the drawings of a house were found to have a positive correlation of $+.378$ with sociometric evaluation and the figure drawings were shown to have a positive correlation with sociometric evaluation of $+.344$. It is argued by the author that since the sociometric choices involved in the sociometric evaluations were consciously made, the results provide evidence that tree drawings are a more subconscious projected self-portrait than those projected by either the figure or the house drawings.

However, an implication of the levels hypothesis was not the hypothesis under investigation in Orgel's study. Therefore, her results interpreted as supporting that hypothesis in the H-T-P can be seen as an incidental finding. In addition, using only one judge as the criterion for rating subjects' popularity potential may not have been an adequate way to test the hypothesis. The results seem to

support this criticism. Moreover, one of the difficulties in studies of this kind is establishing reliability estimates. Usually the estimate of the reliability of a given judge's use of a given instrument is compounded with whatever reliability the instrument may have when used with a given population. If reliability is either high or low it is very difficult, if not impossible, to determine what proportions of the total reliability can be attributed to each of its two possible sources.

When comparing Orgel's study with the Cassel, Johnson and Burns' investigation, the following can be noted. The subjects who took part in the former experiment were children of elementary school age while those involved in the latter experiment were adults of at least minimal legal working age. This difference, plus the differing conditions under which these experiments were carried out (for example, in Orgel's study H-T-P's were group administered whereas in the other study H-T-P's were individually administered), supports the generality of the levels hypothesis at least with respect to these two age groups. Furthermore, in spite of the deficiencies indicated previously, the findings of these studies do support the H-T-P as tapping different levels of personality. This evidence was considered since the hypotheses tested in the present experiment concern the H-T-P as tapping levels within personality. Two other

experiments relevant to this point will also be presented. However, before doing so, some research illustrative of current experimentation on the stimulus value of test stimuli, will be outlined. This seems necessary because this experiment also involved the stimulus value that aspects of H-T-P drawings have for subjects. This is dealt with in the next part of this chapter, part three.

3. Stimulus Value and S.D.

The stimulus value of projective technique stimuli has recently received considerable research attention and the S.D. has been a popular means by which stimulus value can be measured. One such study is an investigation carried out by Friedman, Johnson and Fade.³² In this study a thirty scale form of the S.D. was used to investigate the stimulus properties of ten I.A.T. cards. The subjects who took part in this investigation were 216 male and female students enrolled in introductory psychology courses at two universities. Subjects were informed that they were acting as judges in a study concerned with measuring the meaning of a series of pictures that were to be projected onto a screen in a dimly

³² Jack C. Friedman, Carleton A. Johnson and Kermit Fade, "Subjects' Descriptions of Selected I.A.T. Cards via the Semantic Differential", in Journal of Consulting Psychology, Vol. 28, No. 4, August 1964, p. 317-325.

lit auditorium. Since a pilot study indicated that scale order and polarity did not affect S.D. ratings these two variables were not controlled in the experiment proper. Subjects were given three minutes to complete their ratings on all thirty scales. A ten point scale was used in this study in an attempt to increase the precision of measurement.

The first analysis of data was intended to test the frequency of intragroup agreement among the subjects on their ratings of the cards. Each of the scales was scored for the direction of the ratings. Frequency was determined by counting which one of two bipolar adjectives was checked on each scale and this was tested for significance by means of a chi square analysis. The .05 level for a two-tailed test was chosen for rejection of the null hypothesis. It was found that 133 scales for male subjects and 181 scales for female subjects from the one university were significant while 141 scales for male subjects and 171 scales for female subjects from the other university were significant.

In another analysis of data, intercorrelations were computed among the four groups from the distributions of ratings. A count was made of the frequency of ratings favouring the left-hand side of the scales. This was done for each of the ten I.A.I. cards and Spearman rank order correlations were computed between like sex groups from different geographical regions and also for unlike sex groups from the

same geographic regions. The results showed that only two rhos, both for card 16, were not significant at the .01 level. These results were interpreted as indicating there is a lawful and orderly pattern to the manner in which people rate selected LAAT cards.

Some aspects of this study warrant further comment. It is surprising that the pilot study indicated no need to control for the effect of polarity and scale order on S.D. ratings. If the pilot study involved a very small N such effects could have been hidden just on the basis of that fact alone. Furthermore, Green and Goldfried³³ have demonstrated that polarity varies with the concept or concepts being rated as well as with the adjective scales used. At the present level of knowledge concerning the S.D., precisely how the rating of given concepts will influence the polarity of given scales used for the ratings, is essentially unknown. Until more precise knowledge is available some attempt to control for polarity, however indirect, seems warranted. In addition, this is one of few studies using the S.D. known to this writer in which an attempt to control for scale order was not made. Finally, the subjects involved in this experiment were university students and therefore the results cannot be safely

³³ Russell F. Green and Marvin R. Goldfried, "On the Bipolarity of Semantic Space", in Psychological Monographs, General and Applied, Vol. 79, No. 6, 1965, Whole No. 599, p. 29.

generalized beyond such a population. Since the experiment reported here selected its sample from a student population this latter consideration cannot be considered as a limitation of it.

Another example of research based on stimulus value is the study Zax and Loisel³⁴ carried out to assess the stimulus value that Rorschach's inkblots have for children. Forty male students with a mean age of 11.61 and forty female students with a mean age of 11.69 from a fifth grade elementary suburban school were matched for age and I.Q. The S.D. scales used were chosen for their appropriateness for rating inkblots as well as for their respective factor loadings on the three major factors. Seven scales were used to measure each of the three factors. Ratings for each card on each of the twenty-one scales used, were recorded separately for male and female subjects. From an analysis of directional trends of the ratings it was indicated that Rorschach's inkblots tend to convey somewhat consistent impressions since eighty chi squares were significant for male subjects and eighty-eight chi squares were significant for female subjects. In addition, sex and age differences among ratings were found. For example, female subjects tended to

³⁴ Melvin Zax and Robert H. Loisel, "Stimulus Value of Rorschach Inkblots as Measured by the Semantic Differential", in Journal of Clinical Psychology, Vol. 16, No. 2, April 1960, p. 160-163.

evaluate cards I and X more positively than male subjects. Moreover, younger subjects showed a greater tendency toward more positive evaluations than older subjects.

This experiment raises two issues of importance here which are the effects of age and sex on S.D. ratings. The question of age as a relevant variable requiring control seemed pertinent to the experiment reported here. The manner in which the effect of age on S.D. ratings was controlled for in this experiment will be indicated in chapter two. However, no attempt was made to control for the effect of sex on the ratings made in this experiment. In planning it was decided that if the results were insignificant after an analysis cutting across the sexes was completed, a further breakdown of the data in terms of like and unlike sex groups could be carried out. This point will also be dealt with in chapter four.

In another experiment on stimulus value Kamano³⁵ investigated the meaning of "Mother" and "Father" cards of the Rorschach using the S.D. Kamano set out to test the hypotheses that there should be greater semantic similarity (smaller D values) between card IV and the concept father than between card VII and the concept father and greater

³⁵ Kamano, Op. Cit., p. 50-52.

semantic similarity between card VII and the concept mother than between card IV and the concept mother. Kamano had forty male and forty female college students rate four concepts, Rorschach cards IV and VII, mother, and father on seven semantic scales. In an individual administration each subject rated each of the four concepts presented in a randomized order. The results indicated that the array of D scores between card IV and the concept father were smaller in magnitude than the array of D scores between card VII and the concept father. This difference was significant at the .01 level. The similar comparison of D scores between card IV and the concept mother were greater than the D scores between card VII and the concept mother. This difference was also significant beyond the .01 level.

Kamano recommends caution in interpreting these results as supporting the hypotheses he tested. In doing so he raised a question that has important implications for the experiment presented here. Kamano maintains that when two concepts are allocated to about the same point in semantic space, it does not necessarily follow that the two concepts have the same meaning. This inference is one potential weakness in many studies using the S.D. to assess the value of test stimuli.

Another study using the S.D. to investigate stimulus properties was conducted on the Bender-Gestalt. Taylor³⁶ investigated the stimulus value this test has for forty-one male and twenty-seven female undergraduates enrolled in two general psychology classes with a median age of nineteen. Each one of the nine Bender-Gestalt designs were exposed by projecting them on a screen and subjects were given three minutes to rate each design on the twenty scales selected for use in this experiment. Data were analyzed only for the direction of meaning and therefore all check marks were dichotomized into two groups, divided by the neutral or central position. These check marks that were placed in the neutral position were equally divided between the two groups. Data were analysed for the total sample as well as for like sex groups separately. Of the 180 chi square values for the total sample, seventy-nine were significant at the .05 level or beyond. Taylor concluded that at least some of the symbolic reactions the designs stimulated were reactions common to groups of individuals.

Taylor's subjects were undergraduate students in general psychology courses. If subjects could choose psychology as a course, this could result in a biased sample. That

³⁶ Alexander Taylor, "The Meaning of the Bender-Gestalt Test Designs: A Study in the Use of the Semantic Differential", in Journal of Projective Techniques, Vol. 24, No. 4, December 1960, p. 433-438.

is, the results could not justifiably be generalized beyond nineteen year old male and female undergraduates who elect to take courses in psychology. Many experiments in psychology involve students enrolled in various psychology courses and therefore their results are also best limited to such populations. Also, since such subjects could be considered as test sophisticated, it is not unreasonable to assume that they do respond to what they may anticipate to be the experimenter's expectations, and produce results that are biased in terms of some expectancy set. This could be more probable, particularly if the experimenter also teaches a course the students of whom are also the subjects for an experiment conducted by him. Another point can be raised about Taylor's experiment. The Bender-Gestalt is typically an individually administered test. Taylor, however, presented the Bender designs in his experiment to subjects in group sessions. Interpretations of data obtained under group administration may not be directly applicable to situations in which data is gathered under conditions in which tests are administered individually. That is, social and other possible variables operating in a group situation, to significantly influence results, likely do not operate in a two-person situation.

A limited number of experiments have been reported to use the S.D. for investigating hypotheses formulated from clinical experience with drawing techniques. The only three

such experiments known to the writer will be considered in the chronological order of their appearance in the research literature. The first of these experiments was carried out by Kamano³⁷ who tested the hypothesis that when a person responds to the D.A.P. instructions he draws a self portrait. Kamano assumed that if this hypothesis was valid the human figure drawn by a person should be similar in meaning to his self perception. Forty-five institutionalized women with a mean age of 29.11 labelled as schizophrenic and screened to assure sufficient contact with reality to complete the tasks were asked to rate the following concepts on fifteen scales of the S.D.: 1) figure drawing; 2) my ideal self; 3) my actual self; 4) my least-like self. Each subject was administered the tasks individually. Rank order correlations between the figure drawing and each of the other three concepts were determined. Kamano found that there was greater semantic similarity between the figure drawing and the actual self-concept than for the other comparisons made. The author concluded that these results support human figure drawings as representations of the drawer's self perception.

Some aspects of this experiment warrant further comment. The instructions specified subjects to draw the

³⁷ Dennis K. Kamano, "An Investigation on the Meaning of Human Figure Drawings", in Journal of Clinical Psychology, Vol. 16, No. 4, October 1960, p. 429-430.

picture of a woman. This innovation on the original instructions might have reduced the degree to which figure drawings of some subjects may have represented self perceptions. While the number of female subjects who draw male figures are not numerous and while this is not considered as significant as when it occurs with male subjects, there might have been significant differences between female subjects who produced male figure drawings and those who produce female figure drawings, when the instructions allow for such alternatives. Furthermore, Kamano suggests that the results cannot be generalized beyond women classified as schizophrenic. A sample so labelled does present difficulties for the generalization of research findings. The criteria for classifying schizophrenia has been said to vary so much that several groups so labelled may actually be from different populations. Indeed, schizophrenia as a group of people belonging to a specific psychotic population could be heterogeneous in more ways than they could be homogeneous. Nevertheless, Kamano's results can be taken as indicating why results in experiments of this kind vary as much as they do. One reason could be that samples considered to be from identical populations were actually from different populations. Finally, such variations in results could also be attributed to differences among subjects regarding the motivational conditions under which they take part in experiments. Subjects considered

abnormal, who go to a clinic for assistance, might be expected to have different motivational sets toward tests than college students. Moreover, subjects who are administered a test in group session, might perform in a different way than subjects to whom a test is administered individually.

The second investigation relevant here was carried out by Pickering³⁸ who investigated to what extent the figure drawings of the H-T-P drawings can be taken as representative of a person's self-concept. Thirty male and female nonprofessional employees of a general hospital with a mean age of 30.98 and a standard deviation of 9.4, were selected as subjects on the basis of I.Q. scores. These subjects were asked to rate their H-T-P drawings on a set of S.D. scales after they had been administered an H-T-P. Next, they were asked to rate their actual, ideal and least-like selves on the same scales. In another phase of the experiment, independent judges rated subjects' H-T-P drawings on the same S.D. scales. Both subjects' and judges' rating scores were converted to standard scores. Correlations were calculated between subjects' ratings of figure drawings and the ratings of their actual, ideal and least-like selves. Judges' ratings of each drawing were compared to those made by corresponding

³⁸ Fern Pickering, The Figure Drawings and the Phenomenon of Projection, doctoral thesis presented to the School of Psychology and Education of the University of Ottawa, Ontario, 1963, ix-83 p.

subjects so that correlations were calculated between judges' rating of figure drawings and subjects' ratings of their figure drawings, actual selves, ideal selves and least-like selves. The significance of all these correlations was determined by a t-test. Since all correlations were significant at the .01 level, Pickering interpreted these results as indicating that a person's figure drawing can be interpreted in terms of both his actual and ideal selves.

One aspect of this investigation relevant to the experiment presented here, is pointed out by the author. The judges rated subjects' drawings. Pickering indicates that it would have been more adequate to have judges rate some aspect of each subject from his drawings rather than rate his drawings per se. In passing, this investigation used a relatively small sample ($N=60$). The author attributes sample size as the main source of variation in results and supports her conclusion by the evidence of inexplicable significant differences found in the twenty member retest group (reliability) that were not demonstrated in the total sample.

The other relevant investigation referred to earlier was conducted by Dussault.³⁹ This study investigated Buck and Hammer's hypothesis that the three drawings of the H-T-P

³⁹ Claire Dussault, Concepts De Soi Et Niveaux De Conscience Reflétés Par les Dessins Sur H-T-P, thesis presented to the Faculty of Psychology and Education, University of Ottawa, Ontario, 1965, vi-98 p.

reflect different levels of consciousness within the personality. Ninety-two persons, 43 male and 49 female, following university extension courses, were asked to rate their H-T-P drawings on a set of S.D. scales as well as rate both their actual and ideal self-concepts. The self-concepts were rated first and followed by the administration of the H-T-P, after which subjects were asked to rate their H-T-P drawings. Dussault argued that subjects' ratings of actual and ideal self-concepts could serve as points from which the ratings of each of their drawings could be compared, inferring that the farther away from actual or ideal self-concepts any of the drawings might be rated, the more the drawings could be said to reflect unconscious aspects of personality. D scores were obtained from actual and ideal self ratings and ratings of drawings. The significance of the difference between these two sets of D scores was tested by using the Wilcoxon matched-pairs signed-ranks test.⁴⁰ The results indicated that all drawings were rated closer to the actual self as rated, than to ideal self as rated.

It can be noted that Dussault's results confirm those of Pickering with regard to tree drawings being rated closer to actual self than ideal self. They are also in agreement with those of Kamano's regarding figure drawings being rated

⁴⁰ Sidney Siegel, Nonparametric Statistics. For the Behavioural Sciences, New York, McGraw-Hill, 1956, p. 75-83.

closer to actual self than ideal self. Moreover, Dussault's results demonstrated that the house drawing was the one drawing of all rated farthest away from both self-concepts and the figure drawing was the one drawing of all three rated closest to both self-concepts. Tree drawings were rated between house and figure drawings. However, the distances involved were not significant at the .01 level. To interpret her results Dussault referred to the nature of the subjects who took part in her experiment. Her sample was a heterogeneous group with respect to age (ages ranged from sixteen to fifty-four years). Since they were not clients seeking service at a psychological clinic, Dussault considered them as relatively normal people who, because they could not actualize their potentialities, were dissatisfied with their present situations. Therefore, extension schooling likely represented for them a means of finding greater self-realization. Furthermore, Dussault points out that Buck and Hammer's levels of consciousness hypothesis may hold only when the H-T-P is administered to clinical groups and possibly for this reason her results did not agree with those presented by Cassel, Johnson and Burns.⁴¹

One variable not controlled in the Dussault experiment was scale order. This could have been one possible

⁴¹ Cassel, Johnson and Burns, Op. Cit., p. 157-160.

reason for her obtaining insignificant results. In addition, the particularly wide age range of her subjects could have been another relevant variable. That is, these subjects may have been heterogeneous with regards to other variables as well as age. In spite of their deficiencies these two experiments generally add support to the hypothesis that holds the various stimulus properties of the H-T-P to reflect aspects of personality that exist at different levels of conscious awareness. Of course, Dussault's results cannot be given the same weight given to Pickering's results. Before presenting the hypotheses formulated and tested in this experiment a brief summary of its hypothetical considerations will be considered. This follows in part four.

4. Summary and Basic Hypotheses.

In summary, it can be stated that this experiment was planned to investigate two seemingly contradictory positions regarding the potency for understanding personality, the qualities that expressive aspects of behaviour reflect. Neither Hammer's nor Allport's views were considered to provide an adequate basis for formulating the hypotheses in a given direction. Other possible sources, such as a systematic theory of expression and a systematic theory of H-T-P test drawing behaviour explicitly formulated, do not exist. One probable source, Osgood's theory of meaning, was

mentioned earlier. However, this theory is itself too elementary in its present incipient form for predictions based upon it to be particularly cogent. Since the conflicting hypotheses out of which this experiment grew are closely related to the levels hypothesis, applied to the H-T-P, this latter hypothesis was dealt with in addition to research supporting it. On the basis of this consideration it was deemed reasonable to expect that the H-T-P does tap different levels of conscious awareness within the personality. Several experiments, illustrative of the research aimed at investigating the stimulus value of test stimuli, were presented. In addition, those studies that have reported to use the S.D. to investigate the stimulus value of the H-T-P were evaluated. None of these related areas provided a sufficient basis to warrant formulating the hypotheses to predict directionality in the results. This presented a problem, for there is widespread opinion that research should test the implications of theory and therefore experimental hypotheses should be predictive in nature.

Marks⁴² represents one position that is taken with respect to the directional nature of hypotheses. He supports an early distinction made between inquiry and research.

⁴² Melvin R. Marks, "Two Kinds of Experiment Distinguished in Terms of Statistical Operations", in Psychological Review, Vol. 58, No. 3, May 1951, p. 179-184.

Marks states that the former is a type of experimentation that is carried out independent of theory whereas the latter is experimentation that tests theoretical implications. Marks champions the latter because he sees it as more productive since its appropriate statistical test of significance is the one-tailed model. However, there are those who hold positions polar to that of Marks, notably Hicks,⁴³ who maintains that it matters little if theory is conceived of before or after experiments. Accordingly, the important issue for Hicks is whether or not any statistical model of significance is an appropriate test of the hypotheses being investigated in a given experiment. He emphasizes that the role of statistical tests of significance is to indicate the degree to which the hypotheses in question likely fit the data obtained. Therefore, the issue of the directional nature of hypotheses seems to be quite controversial and seems to provide no unequivocal basis for deciding whether or not to formulate directional hypotheses. For this experiment then, there remained several research questions, open-ended in nature. One question was, which one of the two aspects of H-T-P test drawing behaviour, content or expressive, reflects qualities of personality more potent for the

⁴³ W.E. Hicks, "A Note on One-Tailed and Two-Tailed Tests", in Psychological Review, Vol. 59, No. 4, July 1952, p. 316-318.

understanding of personality? Another question was, do the content aspects of H-T-P test behaviour reflect qualities of personality more unconscious or do the expressive elements reflect the more unconscious qualities? Still another question was, when clinicians interpret H-T-P drawings which of the two aspects, content or expressive, is the more useful to them in understanding personality?

In formulating the hypotheses, it was decided to follow Hayes⁴⁴ who maintains that, by and large, the most significant tests in psychology are those that test non-directional hypotheses. In addition, it was considered that the demands of a two-tailed statistical model are generally greater than those of a one-tailed model and therefore a two-tailed test would place greater demand on the data in order for results to reach significance. As a result, the hypotheses formulated in this experiment were non-directional. Before presenting these hypotheses a brief statement of the reasoning followed in designing this experiment will be outlined.

One of the problems of an experiment of this kind is devising a way to adequately measure the unconscious qualities of personality. The manner in which unconsciousness was

⁴⁴ William L. Hayes, Statistics for Psychologists, New York, Holt, Rinehart and Winston, 1963, p. 285.

operationally defined, while not ideal, was deemed relatively appropriate. It was held that semantic distance was one way of operationally defining the conscious-unconscious dimension. Therefore, it was reasoned that the hypotheses could be tested by comparing subjects' ratings on the S.D. of their H-T-P drawings containing more expressive elements with their H-T-P drawings containing less expressive elements. It was argued that subjects' ratings of their self-concepts on the S.D. could act as an anchor point in consciousness so to speak. The self-concept was also considered as reflecting the core aspects of personality, the self. Therefore, the semantic distances from their self-concepts with which subjects rated each set of their H-T-P drawings, could be compared. The manner in which each subject's H-T-P drawings were treated to produce a set of these drawings containing less expressive elements will be dealt with in chapter two. Before turning to that chapter, the experimental hypotheses formulated for testing in this experiment will be stated in their general form.

It was argued that the closer to self-concept subjects rate either set of H-T-P drawings, the closer to self in meaning that set could be said to be. That is, if the set containing more expressive elements is rated farther from self-concept than the other set, it might be possible to infer that the content aspects of H-T-P drawings reflect

the more conscious qualities of personality. In addition, it might be possible to infer, that the qualities content aspects reflect, are more closely related to the core aspects of personality.

5. Experimental Hypotheses, General Form.

In order to make a test for evidence supporting either one of the two hypothetical statements regarding the qualities of personality expressive aspects of behaviour reflect, the following experimental hypothesis was formulated to be tested by subjects' ratings. There is no significant difference in the semantic distance from self-concept between subjects' ratings of their H-T-P drawings containing more expressive elements and subjects' ratings of their H-T-P drawings containing fewer expressive elements.

In order to test the implication of Hammer's hypothesis, that when interpreting H-T-P drawings clinicians find the content aspects more yielding of clues than expressive elements to the unconscious qualities of personality, the following hypothesis was formulated to be tested by judges' ratings. This experimental hypothesis in its general form is: There is no significant difference in semantic distance from subjects' self rated self-concepts between judges' (clinicians) ratings of the self-concepts of subjects on the

basis of their H-T-P drawings containing more expressive elements and judges' ratings of subjects' self-concepts on the basis of their H-T-P drawings containing fewer expressive elements.

CHAPTER II

EXPERIMENTAL DESIGN

The design of this experiment is presented in this chapter in seven parts. In part one the definitions, of copied and whole drawings, are presented. In part two, the validity and the reliability of the semantic differential are considered, followed by a brief comment on the House-Tree-Person Test. Part three contains a discussion of the sample and part four describes the role judges played in the experiment. The procedures followed first with the subjects and then with the judges, are presented in part five. An outline of data analysis is presented in part six and each of the hypotheses tested is presented, in their null form, in part seven. With part seven this chapter is completed.

1. Whole and Copied Drawing Definitions.

It has been pointed out earlier that the definitions followed in this experiment are detailed in Appendix 1. However, since it would be difficult to discuss the experimental design without them, two definitions are presented here.

Referring back to chapter one, the inextricability of the content and expressive aspects of the H-T-P are emphasized. This presented a problem since the views of Hammer and

Allport could not be directly tested as there seems to be no adequate means presently available for measuring separately either the content aspects in pure form or the expressive elements in pure form. If attempts were made to measure the aspects from H-T-P drawings as they stand (in which these aspects are interweven), there would be no way to prevent the measurement of one aspect from contaminating the measurement of the other. Therefore, a way had to be found whereby one aspect could be more or less controlled. Following, is presented the means whereby this was attempted.

A commercial artist with more than fifteen years experience in artistic reproductions was consulted with a sample of H-T-P drawings. The twelve expressive aspects of the drawings as seen by Hammer and presented in chapter one, were defined as detailed in Appendix 1. After three separate consultations with the artist, in which each definition of each expressive aspect was discussed, it was agreed that a set of H-T-P drawings could be reproduced to control for all parts of seven of the twelve expressive elements but only certain aspects of the five remaining aspects. Whenever the artist could use objective methods to control for each of the expressive elements, he did so. The following expressive elements were

considered to be controlled for completely by objective methods: 1) size of the drawings; 2) placement of drawings on pages; 3) shading; 4) reinforcement; 5) erasures. These latter three aspects were controlled by the artist intentionally not including them in the set of H-T-P drawings he reproduced. The following expressive elements were considered adequately controlled even though they were dependent upon the skill of the artist: 1) pressure of the lines; 2) qualities of the lines. Only five sets of drawings were rejected by the judges for varying too much from the majority of drawings in these aspects. Three expressive aspects: 1) proportions; 2) perspectives; 3) degree and areas of completion and detailing, were controlled for only in part. Approximately half of their component parts taken together were controlled for in an objective fashion because parts of their definitions also formed the definition of other expressive elements. For example, one component of the definition for perspective is height in the vertical plane. This aspect was controlled for when placement on the page was controlled for, by measurement. Finally, two expressive elements, symmetry and exactness, were considered as uncontrollable. Exactness was found to be quite undefinable in operational terms whereby it could be controlled. While the artist made an attempt to control for one aspect of symmetry, it was deemed essentially not subject to control.

Therefore, each set of H-T-P drawings produced by each subject was reproduced by the artist. Those sets of H-T-P drawings as they were reproduced (unmodified in any manner) by subjects were defined as whole drawings. Whole drawings were taken as containing more expressive elements than copied drawings, expressive elements being in combination with the content aspects. Content was defined broadly as the concepts house, tree and person that classified each drawing as a type of drawing. Therefore, the particular kind of house, tree or figure a subject drew was defined as the content aspects of H-T-P drawings. Expressive elements were defined as comprising the twelve elements stated by Hammer and presented in chapter one.

Copied drawings were defined as those H-I-P drawings that were reproduced by the artist to control for the expressive aspects as specified above. Copied drawings were considered as containing fewer expressive aspects than whole drawings and therefore of having a content emphasis. Therefore, subjects' S.D. ratings of their whole drawings could be compared to their copied drawing ratings. The comparison could be made in terms of the semantic distance from self-concept each set was placed. All ratings, including subjects' ratings of their self-concepts, were carried out with the S.D. The use of this technique in this experiment will now be considered in part two.

2. Validity and Reliability of S.D.

The S.D. was selected for use in this experiment for several reasons. First, semantic distance was considered an appropriate operational definition of the conscious-unconscious dimension of personality. Secondly, as it is said to measure meaning, the S.D. was considered the instrument of choice for measuring the stimulus value of both copied and whole H-T-P drawings. It was also chosen because, as Osgood¹ has indicated, it is appropriate for rating a wide variety of different types of concepts. In this experiment both verbal and drawing concepts were rated by the S.D. Yet another reason was that the S.D. has research supporting its general usefulness, its validity and its reliability. Some such research will be presented in the following paragraphs.

As a result of investigating the connotative meanings of Rorschach inkblots, Rosen² concludes that, because of its flexibility, the S.D. seems particularly promising as a means of carrying out a variety of research in the clinical area. In an investigation of the stimulus value

¹ Charles E. Osgood, George J. Suci, Percy H. Tannenbaum, The Measurement of Meaning, Urbana, Illinois, University of Illinois Press, 1957, p. 77.

² Ephraim Rosen, "Connotative Meanings of Rorschach Inkblots, Responses and Determinants", in Journal of Personality, Vol. 28, No. 4, December 1960, p. 413.

of T.A.I. cards Friedman, Johnson and Fade,³ found sufficient variations within the agreement of descriptions of the cards, to indicate that the S.D. can measure reliably fine distinctions when used to assess the stimulus value of projective technique stimuli. One of the major findings reported by Zax and Loisel, ⁴ in an experiment in which they used the S.D. to study the stimulus value of Rorschach inkblots, was evidence attesting to the usefulness of the S.D. as a device for developing knowledge of the stimulus value of projective technique stimuli. Rabin,⁵ as a result of investigating the global meaning of Rorschach inkblots, concluded that the S.D. has promise as a measuring instrument in psychological research. These studies illustrate research supporting the general usefulness of the S.D. There are, however, more fundamental issues for concern regarding the selection of the S.D. as a measuring device and it is to one of these, validity, the discussion now turns.

³ Jack C. Friedman, Carelton A. Johnson and Kermit Fade, "Subjects' Descriptions of Selected T.A.I. Cards via the Semantic Differential", in Journal of Clinical Psychology, Vol. 28, No. 4, August 1964, p. 317-325.

⁴ Melvin Zax and Robert H. Loisel, "Stimulus Value of the Rorschach Inkblots as Measured by the Semantic Differential", in Journal of Clinical Psychology, Vol. 16, No. 2, April 1960, p. 163.

⁵ A.I. Rabin, "A Contribution to the 'Meaning' of Rorschach's Inkblots via the Semantic Differential", in Journal of Consulting Psychology, Vol. 23, No. 4, August 1959, p. 372.

Osgood⁶ states that since there is no commonly agreed upon independent measure of meaning against which the S.D. can be validated, he bases the validity of the S.D. on face validity. When he says face validity he refers to a situation in which a group of concepts are arranged into three clusters, in terms of meaning, without using the S.D. and then when the S.D. is used essentially the same meaning clusters are produced. However, face validity would not be considered sufficient for most research purposes. Two studies that investigated the validity of the S.D. will now be presented.

It is indeed fortunate that there is tangible research evidence supporting the validity of the S.D. such as that reported by Kelly and Levy.⁷ The authors selected sixty verbal concepts randomly from an atlas of meaning profiles and divided them randomly into three groups of twenty each. Each concept in each group was paired with another concept which differed from it by a given D score. Each of the three groups represented a difference between paired concepts, one by a large D score, one by a medium D score and the third by a small D score. Mimeographed

6 Osgood, et al., Op. Cit., p. 140.

7 Jane A. Kelly and Leon H. Levy, "The Discriminability of Concepts Differentiated by Means of the Semantic Differential", in Educational and Psychological Measurement, Vol. 21, No. 1, Spring 1961, p. 53-58.

booklets were prepared containing a mean profile for one member of each pair of concepts depicted on each of the sixty pages and the pair of concepts appearing at the bottom of each page. Seventy-five undergraduates enrolled in introductory psychology were told how the profiles had been obtained and what the ratings on the various scales signified. These subjects were asked to indicate which member of each pair of concepts was represented by the profile. In this way the authors could measure discriminability in terms of the number of correct choices subjects made for pairs of concepts in identifying the member of the pair depicted by the profile. The authors predicted a decline in the mean number of correct choices from group I to group III (from larger to smaller D scores). The means of the groups (depicting mean correct choices) were as follows: group I, 19.19; group II, 16.07; group III, 11.27. Since these results were in the predicted direction and significant at the .01 level, Kelly and Levy conclude that their results support the validity of the S.D. Another study providing evidence supporting the validity of the S.D. was conducted by Grigg.⁸

In his experiment, Grigg tested whether the S.D. scores of normals would reflect greater semantic distance

⁸ Austin E. Grigg, "A Validity Study of the Semantic Differential Technique", in Journal of Clinical Psychology, Vol. 15, No. 2, April 1959, p. 179-181.

between ideal self and the concept "neurotic" than between self and "neurotic". Forty-two university undergraduates who had completed the psychoneurosis portion of a course in abnormal psychology rated their "self", "ideal self" and "neurotic" on a set of S.D. scales. Two days later all subjects read a selection on Miss X highly favourable to her after which they rated her on the same S.D. scales. Immediately following this, thirty of the subjects read a selection of Miss X branding her as an neurotic. Twelve (to make up the forty-two control group) were requested to wait quietly while the others read. These twelve did not read the selection on Miss X's neurosis. Next, all forty-two subjects rated Miss X on the same S.D. scales. A t-test of the difference between mean ratings of the "self" versus "neurotic" and "ideal self" versus "neurotic" was found to be significant at the .001 level for a one-tailed test. The subjects of this experiment rated their ideal self significantly farther from their rating of the concept "neurotic", than their self rated self concepts. Even though the experimental group rated Miss X closer to the concept "neurotic" after the second reading than after the first, the control group did not rate Miss X significantly closer to the concept "neurotic" at the second rating. Nevertheless, this latter group rated her somewhat slightly in the direction of the neurotic locus. Grigg interprets his results as

supporting the validity of the S.D. In addition to these supporting experimental findings, there are other experimental results and opinions supporting the validity of the S.D.

In a brief progress report Moss⁹ points out that researchers have been quick to adopt the S.D. as a tool because of its general applicability and usefulness. Moss does not deny that many methodological problems remain in addition to the fact that the problem of the "meaning" of meaning itself, has not been resolved. However, he concludes that the S.D. is a useful instrument and that the studies he reviews generally attest to the consistency with which the S.D. is measuring a yet to be completely specified "something" in a meaningful way.

Doubts regarding the metric properties of the S.D. have frequently been raised. Often, the assumption of equal intervals between the positions within a scale has been questioned. Moreover, since D is a measure usually involving several scales, the intervals between the positions over several scales are usually assumed to be equal. Messick¹⁰

⁹ C. Scott Moss, "Current and Projected Status of Semantic Differential Research", in The Psychological Record, Vol. 10, No. 1, January 1960, p. 47-54.

¹⁰ Samuel J. Messick, "Metric Properties of the Semantic Differential", in Educational and Psychological Measurement, Vol. 17, No. 2, Summer 1957, p. 200-206.

has investigated these assumptions in addition to the assumption involved in factor analytic studies on the S.D. that the zero point falls at the same place on each scale. Messick applied the psychometric method of successive intervals to nine of the most frequently used S.D. scales. He found that as a result of his study, the assumptions of a similar placement of origins across scales and the approximately equal interval length from scale to scale, have some basis. A small distortion within some scales revealed by his analysis Messick interprets as insufficient for abandoning the equal interval assumption. Messick's data, therefore, supports these assumptions that underlie the scaling properties of the S.D.

Another question raised about the S.D. has been directed to the assumption of scales being functionally antonymous. Mordkoff,¹¹ in investigating this assumption evaluated sixteen S.D. bipolar adjective scales by having each of the adjectives rated as a concept on a ten-scale S.D. He argued that if these scales were functionally antonymous in this context, they should give rise to profiles symmetrical around the neutral position, one profile for the left side

11 Arnold M. Mordkoff, "An Empirical Test of the Functional Autonomy of Semantic Differential Scales", in Journal of Verbal Learning and Verbal Behaviour, Vol. 2, 1963, p. 504-508.

and one for the right side. A test booklet was prepared for each subject, which contained instructions and the sets of ten scales, two sets per page. Above each set of scales was the adjective serving as the concept to be rated. The order of the adjectives was systematically rotated whereas the order of the scales was constant. Each subject made 320 ratings, that is, each subject rated each of the thirty-two adjective-concepts on each one of the ten scales involved. The subjects who took part in this experiment were forty-one undergraduates enrolled in an elementary course in psychology. The data were gathered in a single group session. A multivariate analysis revealed that only three of the sixteen scales were demonstrated not to be functionally antonymous. In discussing the results, the author proposes a hierarchy of validities for the S.D. and its associated theory of meaning. The first level he proposes is the validity of the specific measurement and scaling assumptions of the S.D. and related analytical techniques. The second level is the validity of the S.D. as a general measure of meaning. The third level is the validity of the S.D. as a measure of what Osgood refers to as the representational mediation process. Mordkoff accepts his results as supporting the first two levels of S.D. validity but not necessarily supporting the third level.

Another study, providing evidence in support of the measuring properties of the S.D., was carried out by Ross¹² who investigated the manner in which people use the S.D. Ross set out to test which one of two hypotheses appear to be the more tenable, Osgood's hypothesis or a counter-hypothesis. Osgood's hypothesis indicates that a concept provokes a response that is largely independent of the context in which it is situated. The counter-hypothesis indicates that a concept evokes a response that varies as a function of its context. In testing these hypotheses Ross asked 120 male and female first year university students in an English course to rate two concepts; half the subjects rated concept M and the other half rated concept B. Concept M was defined in terms of a set of concepts presented in a narrow or homogeneous context. Concept B was defined in terms of a set of concepts presented in a broad context. From a factor analysis carried out with the data, Ross found the results strongly supported Osgood's hypothesis rather than the counter-hypothesis. In addition to experimental evidence supporting its general usefulness, any measuring instrument must be able to demonstrate sufficient

¹² John Ross, "Change in the Use of the Semantic Differential with a Change in Context", in Journal of Verbal Learning and Verbal Behaviour, No. 4, 1965, p. 148-151.

reliability to be of value in research. Therefore, this aspect of the S.D. will now be considered.

Osgood¹³ refers to three types of reliability estimates. Factor-score reliability was established on part of the data obtained in this experiment, as it was considered quite sufficient for these purposes. Osgood establishes factor-score reliability in the following manner. He defines a scale as two bipolar adjectives in between which are seven check mark positions. An item he considers as the combination of one given scale and one concept. An item score is taken as a check mark at one of the seven positions between the two bipolar adjectives that make up a scale. When these item scores are averaged among the scales used to measure each one of the three factors, evaluative, potency and activity, that Osgood considers to measure the major portion of meaning, factor scores have been established. Factor-score reliability is established when these scores are reproduced under retest conditions. What is said about establishing factor-score reliability can be applied to the other two types of scores, for which reliability estimates are obtained.

Osgood's unique way of estimating reliability is based upon his contention that ordinary correlation techniques are inappropriate for establishing reliability estimates

13 Osgood, et al., Op. Cit., p. 126-140.

on separate ratings of the S.D. This is because, since he holds that the S.D. assigns concepts to points in semantic space, this semantic space is multidimensional. That is, since meaning is defined as some point or points in semantic space, two points could correlate highly along one dimension but be quite unrelated along the other dimensions that go to make up meaning, since meaning is multidimensionally determined. Therefore, Osgood estimates reliability by determining the absolute deviations between test, retest conditions for the scales used to measure each of the three factors that measure meaning. These deviations are ranked in absolute size in columns (the smallest deviation at the top--the largest deviation at the bottom). Next, the frequency of each deviation of a given size is determined and what percentage of the total sample that it represents is also determined. These percentages are accumulated from the bottom to the top of the percentage column. On the basis of one sample, Osgood¹⁴ shows that the probability of an average deviation of .50 or greater, of occurring on E factor is .024, on P factor, .009, and on A factor .017. Taken as a group, these deviations all have a probability of occurrence of less than .05 per cent.

¹⁴ Ibid., p. 139.

In his summary evaluation of the S.D., Osgood¹⁵ indicates that he and his colleagues have amassed considerable data on reliability. He indicates that for individuals, shifts in ratings between test-retest conditions of more than between 1.50 and 2.00 are significant at the five per cent level and for groups shifts as small as one-half a scale are also significant at the five per cent level. Therefore, variations in ratings from test to retest conditions of greater than those just indicated are highly unlikely events. There is some evidence that factor-score reliability estimates are more stable than the other types of estimates (particularly for work involving a single subject). In a personal communication to Norman,¹⁶ Osgood recommends factor-score reliability as the estimate of choice and Norman concludes for one of his experiments that some of his data partly justifies this recommendation. Therefore, it seems admissible to conclude that for the purposes of this experiment that the S.D. has sufficient reliability. Estimates of factor-score reliability of S.D. ratings carried out in this experiment on each of the concepts rated, are reported in chapter three. The next

15 *Ibid.*, p. 328.

16 Warren T. Norman, "Stability Characteristics of the Semantic Differential", in American Journal of Psychology, Vol. 72, 1959, p. 581-584.

point to be considered is the basis for the selection of the scales which were chosen for this experiment.

The fifteen scales chosen for this experiment were selected in terms of their judged appropriateness for the concepts to be rated. Five scales were selected to measure each of the three factors Osgood considers to measure meaning in the S.D. Each set of five scales was selected because it has been shown to have high factor loadings on the factor it represents and low factor loadings on the other two factors. The table in Appendix 4 shows a list of the fifteen scales that were used along with their factor loadings on each factor. All ratings carried out in this experiment were made on the same set of fifteen scales (a copy of which is presented in Appendix 3). Five scale orders were established to control for order effects of scales on successive ratings. These five orders were randomly distributed among the sets of booklets that were prepared for the subjects and booklets were assigned to each of the three administrations in random fashion. The polarity of scales was varied, some going from one to seven starting from the left and others going from one to seven starting from the right, and randomly assigned to the five different orders of scales. The five scales measuring each factor were ordered in a randomized order within each one of the five scale orders. These procedures were carried out to control for as many of the relevant variables as possible that have been known to

influence S.D. ratings. Wylie¹⁷ lists these as: 1) response sets; 2) halo effect; 3) the tendency to check one end of scales; 4) touchy content areas that encourage denial and scoring in the neutral position of the scale; 5) saturation causing fatigue and less attention toward the end of the test; 6) a preceding scale affecting both the integration of the next scale and the way it is scored; 7) the same scales being scored differently by the two sexes. Pickering¹⁸ followed similar randomizing procedures to control for these variables.

A subject's self-concept in this study was operationally defined as the manner in which the subject rated himself or herself on the fifteen semantic scales used in the study. The appropriateness of using self descriptive techniques for such measures has been questioned. Lochlin¹⁹ has pointed out that differences in self descriptions could represent differences in meaning applied to the words used in self descriptive techniques, rather than represent differences in the way people perceive of themselves. While Lochlin presents some evidence in support of this view, some restrictions of this conclusion seem warranted. When such a homogeneous group as

17 Ruth C. Wylie, The Self Concept, A Critical Survey of Pertinent Research Literature, Lincoln, Nebraska, University of Nebraska Press, 1961, p. 32-33.

18 Pickering, Op. Cit., p. 36.

19 John E. Lochlin, "Word Meanings and Self Descriptions", in Journal of Abnormal and Social Psychology, Vol. 62, No. 1, 1961, p. 28.

college students are the subjects involved, the consensus on word meaning would likely be quite high and greater agreement as to word meaning would be more probable in such a sample than in a more heterogeneous sample.

Madden²⁰ investigated a problem related to similarity in concept meaning. He attempted to investigate the extent to which there is justification in interpreting descriptive concepts about the person as equal to the "self" of that person. Thirty college students rated fifty-nine personal characteristics derived from the MF scale of the M.M.P.I. For these ratings items were put in the third person singular. Only characteristics derived from items marked true by subjects during an earlier administration of the M.M.P.I. were used. Subjects rated "I, myself" on fifteen scales of the S.D. and each one of the other fifty-nine items. The range of distances for each subject was divided into tenths. In each subject's data the percentage of true items (those marked true on the earlier administration of the M.M.P.I.) was computed for each of the ten areas. A fifty per cent value for an area meant that true items had no greater or lesser tendency than false items of falling in that area. The percentage of

²⁰ James E. Madden, "Semantic Differential Ratings of Self and Self Reported Personal Characteristics", in Journal of Consulting Psychology, Vol. 25, No. 2, April 1961, p. 183.

true items in the areas was then averaged across all subjects. Madden found that the mean percentage of true items in the first five areas was significantly larger than that of the last five areas, beyond the .01 level of significance. These results support the use of descriptive statements for reflecting aspects of self.

The problem of a self-descriptive technique used as an operational definition of self-concept and, therefore, being open to distortion by subjects, has also been raised. After careful consideration, the writer shared the opinion of Sperber and Spanner²¹ that a self-descriptive method is not necessarily a distorted view of a person's self-concept. To support their view, these authors predicted that self-descriptive responses to a personality inventory made by subjects suffering some pathology, would be significantly less related to the social desirability of the items than the self-descriptive responses of normal subjects. This prediction was made in spite of the likelihood that both groups have common views of what constitutes social desirability. Testing this prediction with seventy-one subjects from an out-patient psychiatric clinic against a control group of college students, the authors' findings confirmed their prediction.

²¹ Zarwill Sperber and Marvin Spanner, "Social Desirability, Psychopathology and Item Endorsement", in The Journal of General Psychology, Vol. 67, First Half, July 1962, p. 111.

To increase the precision of measurement in this experiment, each position of the seven steps in each scale was defined by adverbial qualifiers. When investigating this aspect of the S.D., Wells and Smith²² found that so defining scale steps resulted in a significant increase in the precision of measurement when one concept is rated on all scales rather than when all concepts are rated simultaneously on one scale at a time. The former method was therefore utilized in this experiment.

The H-T-P was the other technique involved in this experiment. Since only one administration of it was necessary, no estimate of either any aspect of its validity or its reliability was established. However, estimates of the reliability of subjects' ratings of whole H-T-P drawings and of copied H-T-P drawings are reported in chapter three. In addition, estimates of the reliability of judges' ratings of both whole and copied sets of H-T-P drawings are also given in chapter three.

3. Sample.

The subjects who took part in this experiment were two classes of male and female first year college students enrolled

²² William D. Wells and Georgianna Smith, "Four Semantic Rating Scales Compared", in Journal of Applied Psychology, Vol. 44, No. 6, December 1960, p. 397.

in an introductory course in psychology. These subjects were selected to form the sample on the basis of their availability as well as other factors. These other factors included the fact they were deemed sufficiently sophisticated to understand the instructions of the S.D. and sufficiently sophisticated to become involved in the rating and drawing tasks as well as homogeneous enough in age for this variable not to influence ratings. To maximize involvement tendencies, subjects were promised personal anonymity and therefore each subject was provided a research number that was recorded on each one of his or her productions. The original total sample consisted of 133 subjects. One set of whole H-T-P drawings unwittingly was lost in the artist's reproduction process. Of the remaining 132 H-T-P's, some subjects who were present for the first administration were not present at the second administration thereby reducing the final number of H-T-P ratings from subjects available for analysis to 118.

The retest reliability estimates for subjects' ratings were limited to fifty ratings. This was considered sufficient since it could be expected that an increase in the number of retest ratings would likely demonstrate at least the same level of reliability and likely even higher reliability estimates. The total sample was considered sufficiently homogeneous in age for this variable not to influence significantly

B.D. ratings. However, age may not be relevant as Neuringer²³ has demonstrated. Due to the limited time judges had available for their ratings, each of the three judges rated the same one hundred H-T-P drawings, 50 copied, 50 whole, and unfortunately, only twenty of these were available for providing retest reliability estimates of judges' ratings.

4. Judges.

The three judges were graduate students in psychology, all but one of whom had completed a course in the analysis of drawing techniques. Of the original three who volunteered to act as judges, one, in his second M.A. year, had to withdraw due to the pressure of other academic commitments. Therefore, the writer was forced to act as the third judge. While this meant that all three judges were Ph.D. candidates in clinical psychology, it may have introduced a relevant variable (or set of variables) that may have biased part of the results in a manner difficult to ascertain. The possible effect an experimenter can have on the results of his experiments, even

²³ Charles Neuringer, "Effect of Intellectual Level and Neuropsychiatric Status on the Diversity of Intensity of Semantic Differential Ratings", in Journal of Consulting Psychology, Vol. 27, No. 3, 1963, p. 280.

unwittingly, is amply dealt with by Kintz et al.²⁴ However, under the practical exigencies in which this study was carried out this innovation was unavoidable.

5. Procedure.

a) Subjects.- This experiment was carried out in three phases. In phase one, each one of the two classes of subjects was administered the tests as a group. First, booklets containing a face sheet, two and one half pages of instructions and one page containing the fifteen scales and the concept "I, myself" at the top of the page, were handed out to each subject. The S.D. instructions were read over carefully, aloud and subjects were asked to follow along with the reading of them in their booklets. Next, subjects were given the opportunity to ask questions to clarify the nature of the rating task, after which subjects were asked to rate their self-concepts. Then, each subject was given the standard H-I-P drawing technique form. The instructions of the H-I-P were read aloud. They are as follows: "Using the pencil given you, I want you to draw as good freehand drawings as you can of a house, tree and person in that order. On the form given you draw first a house on that page of the form that has

²⁴ B.L. Kintz, et al., "The Experimenter Effect", in Psychological Bulletin, Vol. 63, No. 4, April 1965, p. 223-232.

the word house printed on the top and after that a tree on the page that has the word tree printed on the top and finally draw a whole person on the page of the form that has the word person printed on the top." After the subjects had completed the H-T-P drawings, another rating booklet was provided differing from the self-concept rating booklet only in that it had three pages of scales, each page containing one set of the fifteen rating scales. Subjects were then requested to read over the instructions before starting and next rate each one of the three H-T-P drawings. One page of fifteen scales was used in the second booklet for rating each drawing. At the top of each one of the pages containing the scales was written the following concepts in this order: "my drawing of a house", "my drawing of a tree", "my drawing of a person".

During the first and second administrations, a commercial artist was hired to take each one of the 133 sets of H-T-P drawings and reproduce it, controlling for as many of the expressive elements as possible. Three weeks after the first administration, the second administration was carried out. First, each subject was asked to rate each one of their copied H-T-P drawings. Instructions for the S.D. ratings were read aloud again and a period was provided for subjects to ask questions related to the rating task. The instructions given for rating the copied H-T-P drawings were as follows: "Each

of you will be given a set of three drawings and two rating booklets. You will be asked to make two sets of ratings, one in each booklet. These ratings are the same type of ratings you completed three weeks ago. First, take the booklet which contains the three pages of rating scales. Please follow along as I read the instructions again. (Instructions read aloud). Using one page for each drawing, rate the three drawings given you. As you rate each drawing have only that drawing you are rating beside you for reference. Before you place your check mark, consult the drawing you are rating. You may or may not feel that the set of drawings given you are familiar. Try not to worry about it. Just rate each drawing as best you can. After you have completed a rating for each drawing take the other booklet given to you first and rate the concept at the top of the page containing the scales (this was to provide retest reliability data for subject's self rated self-concept)".

Four weeks intervened between the second and third administrations. The third administration was necessary to provide retest reliability data on the ratings of both copied and whole H-T-P drawings. Subjects in the one class consisting of fifty-seven students from the original sixty-four in that class who had rated their copied H-T-P drawings, were asked to rate their copied H-T-P drawings again. Subjects in the other class consisting of sixty-three students from the

original sixty-nine who had rated their whole H-T-P drawings, were asked to rate their whole H-T-P drawings again. In the final analysis retest reliability of ratings on both whole and copied H-T-P ratings were based on fifty ratings each.

b) Judges.- Two sets of H-T-P drawings, one consisting of ten sets of H-T-P drawings, and the other consisting of twenty sets of H-T-P drawings, were given to each of the three judges to be rated. Each set contained half copied and half whole H-T-P's. Judges were provided with three pages of scales for rating each type of H-T-P. These three pages were collated into a booklet containing only a front sheet but no instructions. Judges were made familiar with the S.D. at three different one-hour sessions. At the top of each page of scales a concept similar to the following was printed: "The Self-Concept of this subject from the (copied) or (whole) drawing of his house". Before rating the first ten H-T-P drawings, judges did not agree upon what aspects within the drawings would be used for the basis of their ratings. Indeed, at no time during this experiment was any effort made for judges to reach such an agreement. Once completed, these two sets of ratings were each analyzed for inter-judge reliability so that the entire sample of H-T-P drawings could be divided equally among the judges. However, since time did not allow for an extended period during which high inter-judge rating reliability could be developed, each judge was asked to rate

the same one hundred sets of H-T-P's. Twenty H-T-P's were rated twice by each judge to provide retest reliability (factor-score) data. These estimates are presented in chapter three.

6. Analysis of Data.

The seven points on each scale were numbered from one to seven, the neutral position being assigned the score of four. Each page of fifteen scales yielded five scores on each factor. These scores were averaged to provide a mean score for each factor.

D scores were calculated from subjects' self-concept ratings and their ratings of their whole drawings, and also from subjects' self-concept ratings and their ratings of their copied drawings. In calculating these D scores, means for each of the factors were determined by taking the mean scores from each page of rating scales and averaging them for the three drawings. D scores were also calculated from subjects' self-concept ratings to each judge's ratings of subjects' self concepts on the basis of their whole drawings, and also from subjects' self-concept ratings to each judge's ratings of subject self concepts on the basis of their copied drawings. In addition, D scores were calculated from subjects' self-concept ratings to the ratings of each of their whole drawings and from subjects' self concept to the ratings of each of

their copied drawings. D scores were also calculated for each of the three judges in a similar manner. Therefore, four pairs of D scores were obtained from subjects' ratings and four pairs of D scores were calculated from the ratings carried out by each of the three judges.

The arrays of D scores obtained from subjects' ratings are presented in summary form in Figure 1. Those obtained from judges' ratings are summarized in Figure 2.

The significance of differences between each set of D scores were tested with the Wilcoxon matched-pairs signed-ranks test.²⁵ This test was chosen because Osgood²⁶ has indicated that it is an appropriate test for testing the significance of differences between pairs of D scores. Furthermore, its merit lies in the fact that it accounts for both the direction and the magnitude of differences. Moreover, assumptions concerning the distribution of the population of D scores, from which the obtained D scores can be said to be a sample, are unnecessary.

All D scores were obtained in the manner presented by Osgood.²⁷ According to Osgood, D stands for semantic

²⁵ Sidney Siegel, Nonparametric Statistics. For the Behavioural Sciences, New York, McGraw-Hill, 1956, p. 75-83.

²⁶ Osgood, et al., Op. Cit., p. 102.

²⁷ Ibid., p. 91.

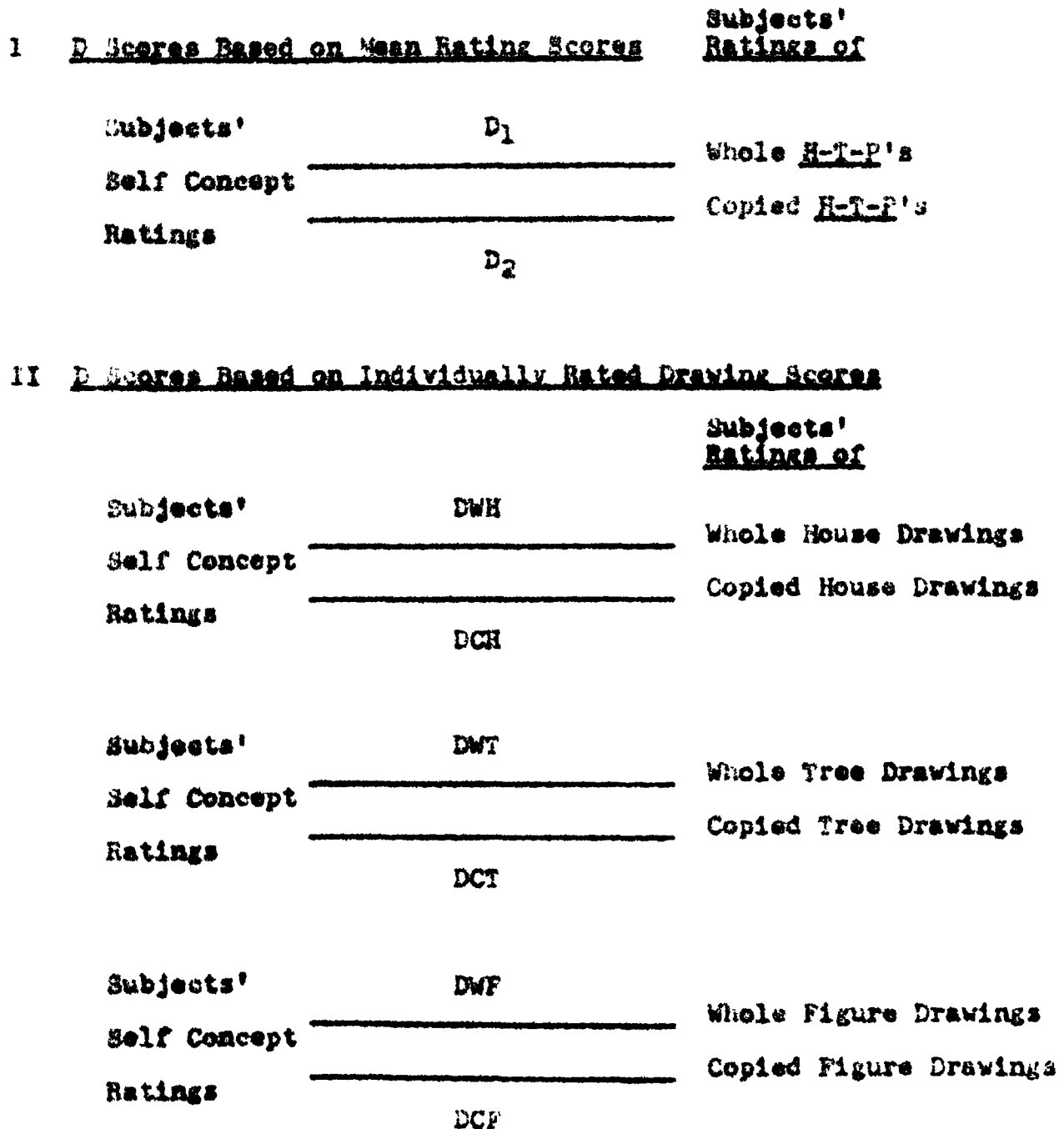


Figure 1.- Diagrammatical Representation of D scores Obtained from Subjects' Ratings.

I D Scores Based on Mean Rating Scores Judge's Ratings of Subjects' Self Concepts Based on Subjects'

Subjects'	D_3	Whole H-T-P's
Self Concept	_____	
Ratings	_____	Copied H-T-P's
	D_4	

II D Scores Based on Individually Rated Drawing Scores

Judge's Ratings of Subjects' Self Concepts Based on Subjects'

Subjects'	DJWH	Whole House Drawing
Self Concept	_____	
Ratings	_____	Copied House Drawing
	DJCH	

Subjects'	DJWT	Whole Tree Drawing
Self Concept	_____	
Ratings	_____	Copied Tree Drawing
	DJCT	

Subjects'	DJWF	Whole Figure Drawing
Self Concept	_____	
Ratings	_____	Copied Figure Drawing
	DJCF	

Figure 2.- Diagrammatical Representation of D Scores Obtained from Ratings made by Each Judge.

distance and is based upon the formula:

$$D_{1L} = \sqrt{\sum_j d_{1L}^2}$$

where D_{1L} is the linear distance between the points in the semantic space representing the concepts 1 and L and d_{1L} is the algebraic difference between the coordinates of 1 and L on the same dimension or factor, J. Each D score is found by taking the difference between the scores of two concepts on each factor, squaring each of the three differences, summing them and taking the square root of this sum.

7. Hypotheses, Null Form.

The hypotheses formulated and tested in this experiment are now presented and stated in the null form. Those formulated to test the general experimental hypotheses relating to subjects' ratings are stated first.

1. There is no significant difference between the sum of the positive ranks and the sum of the negative ranks obtained from difference scores resulting from matched pairs of D scores represented by D_1 and D_2 .
2. There is no significant difference between the sum of the positive ranks and the sum of the negative ranks obtained from difference scores resulting from matched pairs of I scores represented by D_{w1} and D_{ch} .

3. There is no significant difference between the sum of the positive ranks and the sum of the negative ranks obtained from difference scores resulting from matched pairs of D scores represented by D_{wt} and D_{ct} .
4. There is no significant difference between the sum of the positive ranks and the sum of the negative ranks obtained from difference scores resulting from matched pairs of D scores represented by D_{wf} and D_{cf} .

Those hypotheses formulated to test the general experimental hypotheses relating to each one of the three judges' ratings are stated as follows:

1. There is no significant difference between the sum of the positive ranks and the sum of the negative ranks obtained from difference scores resulting from matched pairs of D scores represented by D_3 and D_4 .
2. There is no significant difference between the sum of the positive ranks and the sum of the negative ranks obtained from difference scores resulting from matched pairs of D scores represented by D_{jwh} and D_{jcl} .
3. There is no significant difference between the sum of the positive ranks and the sum of the negative ranks obtained from difference scores resulting from matched pairs of D scores represented by D_{jwt} and D_{jct} .

4. There is no significant difference between the sum of the positive ranks and the sum of the negative ranks obtained from difference scores resulting from matched pairs of D scores represented by D_{jw} and D_{jf} .

CHAPTER III

PRESENTATION OF RESULTS

The results of this experiment are presented in this chapter in the following order. In part one, reliability estimates obtained from subjects' ratings of their self-concepts are presented. These are followed first, by reliability estimates obtained from subjects' ratings of their whole H-T-P drawings and second, by those obtained from subjects' ratings of their copied H-T-P drawings. Part two contains reliability estimates of judges' ratings of subjects' self-concepts based on both subjects' whole and copied H-T-P drawings. In part three, the Wilcoxon formulas and resulting Z scores are presented for testing the significance of differences resulting from comparisons made between subjects' ratings of their whole drawings (comparisons made on complete H-T-P's as well as on individual drawings) and their copied drawings. The fourth part outlines these formulas and Z scores for comparisons made of each judge's ratings presented in the order, Judge I, Judge II, and Judge III.

1. Reliability of Subjects' Ratings.

Table I shows reliability estimates for subjects' self-concept ratings. From Table I it can be seen that the chances are one in one hundred that the scales used to measure

PRESENTATION OF RESULTS

4

Table I.-

Probability of Obtaining Given Deviations from Test-Retest on
Subjects' Ratings of Self-Concept.

(N=100)

Absolute Deviation	Evaluative Factor		Potency Factor		Activity Factor	
	Per Cent	p	Per Cent	p	Per Cent	p
0	17	1.00	14	1.00	9	1.00
.20	26	.83	33	.86	31	.91
.40	21	.57	18	.53	22	.60
.60	17	.36	11	.35	14	.38
.80	7	.19	10	.24	5	.24
1.00	4	.12	6	.14	3	.19
1.20	1	.03	4	.08	4	.11
1.40	2	.07	2	.04	2	.07
1.60	1	.05	1	.02	3	.05
1.80	2	.04			1	.02
2.00					1	.01
2.20	1	.02				
2.40			1	.01		
2.80	1	.01				
3.00						
3.20						
3.40						
5.00						
5.80						

the E, P and A factors will produce deviations as large as or greater than 2.80, 2.40 and 2.00 respectively. Therefore, at least as far as the scales used in this experiment and for the one hundred self-concept ratings involved, the scales used to measure the A factor for these subjects appear to be the most reliable since, of all the scales, they produced the smallest deviation that has a probability of occurrence at the one per cent level.

Table II presents reliability estimates of subjects' ratings of their whole H-T-P drawings. (Table II shows factor scores based upon means obtained from averaging subjects' ratings for all three drawings.) From this table it can be seen that the probability of occurrence of deviations as large as or greater than 1.39 for E factor scales, 1.59 for P factor scales and 1.69 for A factor scales is at the 4 per cent level. As far as subjects' ratings of their whole H-T-P's are concerned the scales measuring the P and A factors have almost equal reliability but seem less reliable than E factor scales since these latter scales, as they were used by the subjects involved, produced the smallest absolute deviation having a probability of occurrence of four chances in one hundred.

The reliability of subjects' ratings of their copied drawings is shown in Table III. At the 2 per cent level of chance occurrence, deviations as large as or greater than 3.19

PRESENTATION OF RESULTS

86

Table II.-

Probability of Obtaining Given Deviations from Test Retest on
Subjects' Ratings of Whole H-T-P's.

(N=50)

Absolute Deviation	Evaluative Factor		Potency Factor		Activity Factor	
	Per Cent	p	Per Cent	p	Per Cent	p
0- .09	26	1.00	10	1.00	12	1.00
.10- .19	6	.74	20	.90	10	.88
.20- .29	18	.68	6	.70	18	.78
.30- .39	4	.50	8	.64	4	.60
.40- .49	14	.46	12	.56	16	.56
.50- .59	0		6	.44	2	.40
.60- .69	12	.32	12	.38	14	.38
.70- .79	0		2	.26	0	
.80- .89	6	.20	6	.24	14	.24
.90- .99	0		0		2	.10
1.00-1.09	6	.14	10	.18	2	.08
1.10-1.19	0		0		2	.06
1.20-1.29	4	.08	4	.08	0	
1.30-1.39	4	.04	0			
1.40-1.49	0		0			
1.50-1.59	0		2	.04		
1.60-1.69	0		0		2	.04
1.70-1.79	0		0		0	
1.80-1.89	0		2	.02	2	.02

Table III.-

Probability of Obtaining Given Deviations from Test-Retest on
Subjects' Ratings of Copied Drawings.

(N=50)

Absolute Deviation	Evaluative Factor		Potency Factor		Activity Factor	
	Per Cent	p	Per Cent	p	Per Cent	p
0- .19	32	1.00	26	1.00	30	1.00
.20- .39	22	.68	22	.72	24	.70
.40- .59	18	.46	24	.50	18	.46
.60- .79	8	.28	8	.26	20	.28
.80- .99	10	.20	8	.18	4	.08
1.00-1.19	2	.10	6	.10	2	.04
1.20-1.39					2	.02
1.40-1.59	2	.06				
1.60-1.79	2	.06	2	.04		
1.80-1.99						
2.00-2.19						
2.20-2.39	2	.04	2	.02		
2.40-2.59						
2.60-2.79						
2.80-2.99						
3.00-3.19	2	.02				

for E factor scales, 2.39 for P factor scales and 1.39 for A factor scales are highly improbable. The scales measuring the A factor appear to be more reliable than the other factor scales, as far as subjects' ratings of their copied drawings are concerned, since they produced the smallest deviation at the 2 per cent level of chance occurrence.

In comparing the reliability estimates from Tables I, II and III, it can be seen that in general those scales measuring the A factor were used by subjects more reliably than the other factor scales. From these tables it can be shown that the average A factor scale deviation is 1.76 which is the smallest average deviation for all the factor scales for which the average probability of occurrence is between one and 2 per cent. In addition, it can be seen that scales measuring the E factor were the least reliable, with P factor scales falling between the other two factor scales in terms of reliability. This finding varies from the reliability data Osgood¹ presents which indicate repeatedly the E factor scales are generally more reliable than the other factor scales. Furthermore, it can be demonstrated that subjects were more reliable in their ratings of their whole

¹ Charles E. Osgood, George J. Suci, Percy H. Tannenbaum, The Measurement of Meaning, Urbana, Illinois, University of Illinois Press, 1957, p. 131, 135.

H-T-P's compared to their ratings of copied H-T-P's and self-concepts. Among these ratings, the average absolute deviation of whole H-T-P ratings was .65 at the .01 level of chance occurrence which is the smallest deviation at the .01 level of the three types of ratings made. That is, subjects were least reliable in their self-concept ratings with their copied drawing ratings falling between the other two types of ratings in terms of reliability.

2. Reliability of Judges' Ratings.

Table IV illustrates the reliability of Judge I's ratings. It indicates that this judge was most reliable with scales that measured the A factor and least reliable on the A factor scales. Table V shows that Judge II was most reliable on the A factor scales and least reliable on the P factor scales. Table VI indicates that Judge III was most reliable on those scales that measured the P factor and least reliable on those that measured the A factor. In Table VII a reliability comparison is made among the three judges.

This comparison indicates that Judge II was most reliable on all the scales. Moreover, Judge III was the least reliable of all the judges on all scales, with Judge I falling between the other two judges in terms of reliability. Furthermore, there appears to be a relatively high degree of reliability indicated among the judges' ratings.

Table IV.-

Probability of Obtaining Given Deviations from Test-Retest on
Judge I's Ratings of Subjects' Self-Concept.
(Factor score reliability)

(N=20)

Absolute Deviation	Evaluative Factor		Potency Factor		Activity Factor	
	Per Cent	p	Per Cent	p	Per Cent	p
0	25	1.00	5	1.00	10	1.00
.20	10	.75	30	.95	15	.90
.40	10	.65	25	.65	5	.75
.60	25	.55	10	.40	10	.70
.80	10	.30	15	.30	25	.60
1.00			5	.15	10	.35
1.20	15	.20	10	.10		
1.40	5	.05			15	.25
1.60					5	.10
2.20					5	.05

PRESENTATION OF RESULTS

51

Table V.-

Probability of Obtaining Given Deviations from Test-Retest on
Judge II's Ratings of Subjects' Self-Concept.
(Factor score reliability)

(N=20)

Absolute Deviation	Evaluative Factor		Potency Factor		Activity Factor	
	Per Cent	p	Per Cent	p	Per Cent	p
0	20	1.00	25	1.00	25	1.00
.20	25	.80	25	.75	15	.75
.30			5	.50	10	.60
.40	10	.55	10	.45	5	.50
.50	10	.45			10	.45
.60	15	.35	10	.35	25	.35
.70			5	.25		
.80	10	.20	10	.20		
.90					5	.10
1.00					5	.05
1.10	5	.10				
1.20	5	.05	10	.10		

Table VI.-

Probability of Obtaining Given Deviations from Test-Retest on
Judge III's Ratings of Subjects' Self-Concept.
(Factor score reliability)

(N=20)

Absolute Deviation	Evaluative Factor		Potency Factor		Activity Factor	
	Per Cent	p	Per Cent	p	Per Cent	p
0			5	1.00		
.20	20	1.00	15	.95	20	1.00
.30			5	.80	30	.80
.40	15	.80	20	.75	15	.50
.60	10	.65	5	.55	15	.35
.80	15	.55	20	.50	5	.20
1.00	5	.40	15	.30	5	.15
1.10			5	.15	5	.10
1.20	25	.35				
1.40			5	.10		
1.50			5	.05		
1.60	5	.10				
1.80	5	.05				
2.00					5	.05

Table VII.-

Average Absolute Deviations and Related Probabilities of Occurrences for Factor Scores Based on Judges' Ratings.

	E		P		A		All	
	Factor	P	Factor	P	Factor	P	Factors	P
Judge I	1.40	.05	1.20	.10	2.20	.05	1.67	.06
Judge II	1.20	.05	1.20	.10	1.00	.05	1.13	.06
Judge III	1.00	.05	1.50	.05	2.00	.05	2.10	.05

3. Wilcoxon Z Scores - Subjects.

Table VIII shows Wilcoxon data and Z scores obtained from the comparisons of subjects' ratings. All Wilcoxon Z scores were calculated from the following formulas, as given by Siegel.²

$$\begin{aligned} \text{Mean (ut)} &= \frac{N(N+1)}{4} \\ \text{Standard deviation (S.D.)} &= \frac{N(N+1)(2N+1)}{24} \\ T &= \text{sum of the ranks with the less frequent sign} \\ Z &= \frac{T - NT}{T} \end{aligned}$$

Wilcoxon Z scores, based on samples with N's greater than twenty-five, are compared to scores from a normal distribution table such as Siegel³ presents. For a two-tailed test the P given in the table for each level of confidence is doubled. From the "Source of T" column in Table VIII it is seen that for all pairs of D scores, other than those from tree drawing ratings, more difference scores come from whole drawing ratings than from copied drawing ratings. Since all Z scores are significant at the .01 level of confidence for

² Sidney Siegel, Nonparametric Statistics for the Behavioral Sciences, New York, McGraw-Hill, 1956, p. 79.

³ Ibid., p. 247.

Table VIII.-

Source of T's, T's, N's, Means, S.D.'s, Wilcoxon Z Scores
and Associated Probabilities Based on Pairs of D Scores
Obtained from Comparisons Made on Subjects' Ratings.

	Source of T		T	N	Mean	S.D.	Z	p ^a
	D ₁	D ₂						
Complete H-T-P's	+64	-94	28260	118	3510.5	3723.77	+6.65	.0018-
House Drawing	+69	-47	21920	116	3393	3621.22	+5.12	.0018-
Tree Drawing	+53	-66	31965	119	3570	3770.97	+7.53	.0018-
Figure Drawing	+63	-55	32430	118	3510.5	2759.02	+10.48	.0018-

^a Probabilities associated with values as extreme as
observed values of Z in the normal distribution for a two-tailed
test.

a two-tailed test, it is indicated that whole drawings were rated by subjects significantly farther away (on the average, larger D scores) from self-concepts than copied drawings. Therefore, the null hypotheses stating that, there is no significant difference in the semantic distance from self-concept between subjects' ratings of their whole drawings and copied drawings, can be rejected. The majority of the deviations for pairs of D scores that compared ratings on whole tree and copied tree drawings, comes from the ratings on copied tree drawings. This indicates that subjects rated their whole tree drawings significantly closer than their copied tree drawings to their self-concepts.

4. Wilcoxon Z Scores - Judges.

Table IX presents Wilcoxon data and Z scores for each of the judges' ratings. Since all but two Z scores (Judge II's figure drawing ratings and Judge III's tree drawing ratings) are significant, the null hypotheses stating that, there is no significant difference in the semantic distance from subjects' self-concepts, between subjects' self-concepts as rated by judges on the basis of subjects' whole drawings and subjects' self-concepts rated by judges on the basis of subjects' copied drawings, can be rejected. The two null hypotheses pertaining to Judge II's figure drawing ratings and Judge III's tree drawing ratings, of course, cannot be

Table IX.-

Source of T's, T's, N's, Means, S.D.'s, Wilcoxon Z Scores
and Associated Probabilities Based on Pairs of D Scores
Obtained from Comparisons Made on Judges' Ratings.

	Source of T		T	N	Mean	S.D.	Z	P ^a
	D ₃	D ₄						
Judge I								
All Drawings	+32	- 6	55.0	38	370.5	21.81	-14.46	.0018-
House	+25	-13	145.5	38	370.5	21.81	-10.31	.0018-
Tree	+23	-15	223.0	38	370.5	21.81	- 6.76	.0018-
Figure	+27	-11	176.5	38	370.5	21.81	- 8.98	.0018-
Judge II								
Full Drawing	+26	-12	281.5	38	370.5	21.81	- 4.08	.0018-
House	+25	-13	246.5	38	370.5	21.81	- 5.68	.0018-
Tree	+23	-15	286.0	38	370.5	21.81	- 3.87	.0018-
Figure	+21	-15	251.5	36	333.0	636.5	- .128	.2262
Judge III								
Full Drawing	+28	-10	175.5	38	370.5	21.81	- 8.90	.0018-
House	+29	- 9	119.5	37	351.5	66.28	- 3.50	.0018-
Tree	+20	-18	351.5	38	370.5	21.81	- .87	.4180
Figure	+27	-11	183.0	38	370.5	21.81	- 6.60	.0018-

^a Probabilities associated with values as extreme as observed values of Z in the normal distribution for a two-tailed test.

rejected. A description of these two latter Z scores will be outlined after a description of the significant Z scores for judges, is carried out.

Ten out of twelve Wilcoxon Z scores for judges are significant which indicates that for the pairs of D scores they represent, most difference scores come from D scores based on judges' ratings of subjects' self-concepts on the basis of subjects' whole drawings. This indicates that judges rated subjects' self-concepts closer to their self-concepts on the basis of subjects' copied drawings, than on the basis of their whole drawings.

The Wilcoxon Z score for Judge II's figure drawing ratings is not significant even though more difference scores come from whole figure drawing ratings than from copied drawing ratings. The Wilcoxon Z score for Judge III's tree drawing ratings is also not significant. Nevertheless, the data indicate that more difference scores come from Judge III's ratings of subjects' self-concepts on the basis of subjects' whole tree drawing ratings, than from ratings based on subjects' copied tree drawing ratings. The interpretation of the data and the conclusions reached as a result of this experiment are dealt with in chapter four which follows.

CHAPTER IV

DISCUSSION OF RESULTS

The discussion of results and conclusions are presented in this chapter in three parts. Part one deals with the interpretation of the results based on subjects' ratings. Part two contains an interpretation of the results produced by judges' ratings. In part three, criticisms of this experiment are made and suggestions for further research generated from its results, are indicated. Part four contains summary and conclusions.

1. Interpretation of Subjects' Results.

The four Z scores derived from subjects' ratings are all significant well beyond the .01 level of confidence. Therefore, the four null hypotheses formulated to test the general experimental hypothesis regarding the subjects can all be rejected. Moreover, all of these Z scores with the exception of one, are in the same direction. That is, subjects rated their whole drawings significantly farther from their self-concept than their copied drawings, with the exception of their tree drawings. These results could be interpreted as indicating that the expressive elements are either more unconscious than content aspects in themselves, or that they reflect the more unconscious qualities of the

personality. This interpretation requires defining consciousness-unconsciousness in terms of semantic distance. It would therefore follow that the greater the semantic distance (the larger the D scores) the more unconscious the level of personality involved. Such an interpretation would support Allport's hypothesis that expressive elements reflect the deeper structure, the more unconscious qualities of personality. This interpretation would also support Wolff's and Maslow's views regarding expressive aspects of behaviour. Moreover, the data obtained from subjects supports Allport's position rather than Hammer's, regarding the expressive aspects of behaviour, when this interpretation of the results is applied.

This interpretation, that expressive elements of W-F test drawing behaviour are either more unconscious than content aspects, or reflect the more unconscious qualities of personality, is not particularly satisfying for two reasons. First, while the operational definition of the consciousness-unconsciousness dimension of personality defined in terms of semantic distance is quite admissible, it is based upon an inference that is difficult to validate. That is, there appears to be no way of substantiating the equivalence of semantic distance and the conscious-unconscious dimension. One conceivable difficulty is that two such concepts may be at opposite poles of one dimension of semantic space yet not

be at opposite poles of other related dimensions since semantic space is multidimensional. Another difficulty lies in subjects rating their whole tree drawings closer to their self-concepts than copied tree drawings. This finding does not fit the interpretation of subjects' results as supporting the hypothesis that expressive elements are more unconscious than content aspects.

Another problem this interpretation presents is related to self-concept as reflecting both core aspects and conscious qualities of personality. First, the self-concept as reflecting the "self" and thereby reflecting core aspects of personality is an explanation that does not appear at variance with theoretical issues. McQuitty¹ points out that those theorists who postulate a self consider it as more or less the central core of personality. Also, doubts could be raised as to the adequacy of considering the self-concept as reflecting the "self", the central core of personality. These doubts do not seem completely justifiable. For example, Moustakas² has indicated that most research studies on the self take self-concept, variously measured, as an

1 Louis L. McQuitty, "A Measure of Personality Integration in Relation to the Concept of Self", in Journal of Personality, Vol. 18, 1950, p. 467.

2 Clark E. Moustakas, "True Experience and the Self", in Donald E. Hamachek (ed.), The Self in Growth, Teaching and Learning: Selected Readings, Englewood Cliffs, N.J., Prentice-Hall, Inc., 1965, p. 46.

operational definition of self and by so doing such studies are made possible, however narrow a concept of self this procedure generates. Moustakas, in addition, shares the view that self, while not limited to its definitions, is the central aspect of the individual personality.

Accordingly, this interpretation holds that the expressive aspects of the H-T-P are more unconscious than content aspects. Since content aspects were rated closer than expressive elements to self-concept, it could be said that content aspects reflect qualities of personality closer to its core. This could mean that expressive elements reflect the more peripheral, superficial or temporal aspects of personality. The main point to be made here is that, in spite of the problems presented, this interpretation is still a possibility. Two qualifications of its acceptance, however, should be made.

One qualification is that it seems more appropriate to use the concept, level of awareness, in place of the phrase, conscious-unconscious dimension. This innovation places less emphasis on certain qualities of personality as being beyond the individual's conscious awareness unless dramatic intervention occurs such as psychoanalysis. It also implies shifting levels in terms of degrees of awareness and that the focus of complete awareness can potentially range anywhere from total unawareness to full awareness. This

innovation's appropriateness is supported when it is considered that expressive aspects of behaviour can be said to be brought into awareness when intentional effort is made to do so, but as Allport³ states, are aspects of personality that usually and most typically are not in the focus awareness. The second qualification is that this interpretation may be an acceptable one but is perhaps not the best of the alternative interpretations available. Another interpretation of the results, introduced in chapter one, was based upon Osgood's theory of meaning. An evaluation of that interpretation will now be taken up.

In chapter one, the cogency of identifying the expressive elements of the H-I-P with what has been called connotative meaning, was pointed out. It was indicated that if the S.D. measures essentially what has been termed the connotate aspects of meaning, it could be expected that subjects would rate their whole drawings closer to their self-concepts than copied drawings. With the exception of tree drawings, subjects rated their copied drawings significantly closer than their whole drawings to their self-concepts. These results are in the opposite direction to what could be predicted from Osgood's theory. Because of the problems

³ Gordon W. Allport, Pattern and Growth in Personality, New York, Holt, Rinehart and Winston, 1961, p. 403.

related to this prediction that were outlined in chapter one, this attempt to interpret the results cannot be taken too seriously. Nevertheless, if it were considered, the results of this experiment would suggest that either the S.D. is not measuring what is referred to as the connotative aspects of meaning or that it is not appropriate to attempt identifying the expressive aspects of the H-T-P with what has been called connotative meaning. Carroll⁴ suggests that the S.D. may be measuring essentially what has been described as the denotative aspects of meaning. If Osgood's theory were used, the results, both the subjects' ratings and the judges' ratings, generally would support Carroll's opinion. Returning again to the hypothesis that the S.D. measures essentially what has been referred to as connotative meaning, it is seen that the overall results of this experiment do not support it. Subjects rating their whole tree drawings significantly closer to their self-concepts than their copied drawings, is its only support. Moreover, when complete sets of copied and whole H-T-P ratings are scrutinized it is seen that the copied drawings taken together are rated significantly closer to self-concept than whole drawings taken together. However, in the final analysis, the results of this experiment suggest

⁴ John B. Carroll, "Review of Osgood, et al., The Measurement of Meaning", in Language, Vol. 35, 1959, p. 76.

that it is premature to look to Osgood's theory of meaning for assistance in formulating directional hypotheses and for interpreting results. That is to say, that this theory in its present form would make any such interpretations much too tenuous to assist in clarifying the issues involved. Also, much more needs to be known about what aspects of "meaning" the S.D. in fact measures as well as about the 'meaning' of "meaning" itself.

* * *. The most appropriate interpretation of the subjects' results would seem to be one that has been suggested by Marstein.⁵ In a consideration of research and theory on the stimulus properties of the I.A.A.I. Marstein summarizes experimentation that has attempted to throw light on the relationship between the physical properties (structure), the ambiguity (uncertainty of meaning) of the I.A.A.I. stimuli and I.A.A.I. responses. Contrary to former hypothesizing, Marstein⁶ concludes that for studies investigating the projection of sex, hostility and fear, low-relevant cards almost invariably elicit the given drive, whereas high-relevant cards seem to evoke a variety of defenses and therefore are much less likely to result in projection of a given drive. High-relevant cards to sex, for example, would be cards in which

⁵ Bernard I. Marstein, Theory and Research in Projective Techniques (emphasizing the I.A.A.I.), New York, Wiley, 1963, p. 167-235.

⁶ Ibid., p. 190.

some aspect of sexuality such as partial nudity, is actually depicted. One complicating variable in these studies is the degree of ambiguity of the cards. Yet, in spite of this difficulty, Murstein⁷ still concludes that all ambiguity levels have significance for reflecting some aspect of behaviour and that experimentation indicates that cards of high structure are most diagnostic because subjects mainly tend to avoid their stimulus-pull. He adds that cards of low and medium structuring are most conducive to the expression of a given drive.

Murstein⁸ finds further support for his conclusion in reviewing research into variations in the central characters of the T.A.I. particularly those investigating physical similarity between the central character and the subject. He points out that such similarity not only vitiates against projection but may actually be an obstacle to it. That is, Murstein⁹ emphatically concludes that similarity between the central character and the subject even to the point of identification, encourages defensiveness and reduces the extent to which a subject will project. This is also indicated by research on the Negro T.A.I. which suggests

7 Ibid., p. 193.

8 Ibid., p. 208.

9 Ibid., p. 209.

that the effect of T.A.T. stimuli is not a simple function of the degree of its similarity to a given subject.¹⁰

Furthermore, it was once the consensus of psychological opinion that projective techniques were beyond attempts of subjects to present themselves as other than what they are as personalities. Murstein¹¹ points out, contrary to this opinion that there is considerable evidence in a number of experiments indicating subjects often could and in fact did control the meaning of their responses. These conclusions provide some guidelines for the meaning that can be given the results of this experiment. The appropriateness of these conclusions will now be considered.

It will be remembered that subjects rated their copied drawings significantly closer to their self-concepts than their whole drawings, with the exception of their copied tree drawings, which they rated significantly farther away from their self-concepts than their whole drawings. One possible explanation for this variance in subjects' results lies in considering the rating of drawings on the S.D. as a type of projective test behaviour itself. This consideration

¹⁰ Ibid., p. 206.

¹¹ Bernard I. Murstein (ed.), Handbook of Projective Techniques, New York, Basic Books, 1965, p. xv.

is supported by Pickering¹² who interpreted some part of her subjects' behaviour also in this way. Therefore, the subjects in this experiment may have become more defensive in their whole drawing ratings than in their copied drawing ratings. This could be explained in terms of the whole drawings, because they contain a combination of content and expressive aspects have greater stimulus-pull than copied drawings. It would follow then that subjects would be less ready to identify whole drawings than copied drawings as belonging to them, by rating their whole drawings on the average farther from their self-concepts. This explanation is made more compelling when it is pointed out that both Allport¹³ and Wolff¹⁴ take expressive elements as reflecting the individuality of the personality. Subjects then would be less ready to reveal themselves, or project their individuality into their rating behaviour, by failing to admit the whole drawings were personally more meaningful to them than their copied drawings. Therefore, subjects would more readily acknowledge their copied drawings as being closer in meaning to the meaning they place on themselves than their whole drawings.

12 Fern Pickering, The Figure Drawing and the Phenomenon of Projection, doctoral thesis presented to the School of Psychology and Education, University of Ottawa, Ontario, 1963, p. 50.

13 Allport, Op. Cit., p. 473.

14 Werner Wolff, Diagrams of the Unconscious (Handwriting and Personality in Measurement, Experiment and Analysis), New York, Grune and Stratton, 1943, p. 185.

Two factors suggest this explanation is valid. First, the subjects who took part in this experiment were college students enrolled in an introductory psychology course of which they had completed one-half. There is good reason to take them as being sufficiently test sophisticated to be aware of the implications their rating behaviour could have for revealing personality. This explanation is further supported by pointing out that many of the studies Murstein¹⁵ refers to involved college students in psychology courses as subjects and other somewhat test sophisticated subjects. Secondly, the motivational set partly determined by the conditions under which this experiment was carried out could have been conducive to subjects rating their copied drawings closer to their self-concepts.

The major obstacle to accepting this explanation lies in the fact that subjects rated their whole tree drawings significantly closer to their self-concepts than their copied tree drawings. One possible answer for this singular reversal in the data may be in the hypothesis that the tree drawing represents a level of personality integration of greater unawareness than either the house or person drawing. If this explanation for these subjects is valid, they may have found

¹⁵ Murstein, Theory and Research in Projective Techniques, p. 167-235.

it possible to project their feelings and personality dynamics into their rating behaviour by rating the whole tree drawings closer to self-concepts than copied tree drawings even though whole drawings contain an emphasis of expressive elements that caused them to rate the other two whole drawings in the opposite direction. Therefore, the whole drawings of person and house may have had the strongest stimulus-pull for the projection of subjects' drives, feelings and attitudes. However, since these can be considered as highly relevant drawings they can be said to stimulate a defensive type of mechanism which is perhaps a denial that caused the drawings to be rated as they were.

This explanation finds some support when the age of subjects is focused on. The age range of these subjects is between eighteen and twenty-one years. As a whole they could be taken as either young adults or late adolescents or a mixture of both groups. In these groups the concepts house and person can readily be taken as calling out in them approach-avoidance conflicts, particularly pressing at these periods in their lives. That is, it is a well-accepted consideration that one of the major conflicts with which these groups are faced, is the struggle between their felt need of continued psychological dependence on their families and their strivings for personal autonomy. They are said to have a need to be independent at a period when they also need

support to face the conflicting demands society places upon them. It is not too difficult to consider their rating the copied house drawings closer to their self-concepts than whole house drawings as a result of both the content and expressive elements of those drawings combining together to reactivate the dependence-independence conflict (bring it into the focus of awareness) which generates a tension that is avoided, to some degree, by a denial of whole house and whole person drawings as being close to their self-concepts in meaning. Again, it is not difficult to see whole figure drawings being rated farther from self-concepts than copied figure drawings, as an attempt to defend against revealing individuality as well as a need to deny the whole figure drawing as having a meaning close to the meaning placed on the self-concept. This means considering figure drawings as giving rise to conflicts over the self. A difficulty with being introspective and conflicts over self, such as self worthiness-unworthiness, has been identified as also prominent among the problems typical of adolescents and to some degree young adults.

2. Interpretation of Judges' Results.

Turning now to consider explanations of judges' ratings, it can be restated that judges on the whole rated subjects' self-concepts closer to their self-concepts on the basis of

their copied drawings than on the basis of their whole drawings. Generally the judges' ratings were in the same direction as were subjects' ratings. The two exceptions were Judge II's ratings based on figure drawings and Judge III's ratings based on tree drawings. These varied mainly in that they were the only two types of ratings (based on copied and whole drawings) whose differences from self-concepts were not statistically significant even though the direction of these ratings was in keeping with the other ratings. One explanation of these results is that judges find copied drawings more useful in determining how subjects rated themselves because there were fewer (expressive elements) variables in these drawings to be considered as compared to whole drawings. That is to say, with more variables available, a greater number of hypotheses about a subject is possible and this increases the number of errors possible. Also, with more variables contained in drawings the greater the probability that one element could indicate one hypothesis but another element present could contraindicate it and therefore whole drawings may present more difficulties for explaining personality aspects of a subject than copied drawings.

Two findings from judges' data that need explanation are that Judge II's figure drawing ratings and Judge III's tree drawing ratings were not significant at the .05 level. There does not appear to be any explanation resulting from

hypotheses related to the H-T-P that seems to fit these findings. Since both these sets of data were in the same direction as the other ten, it seems reasonable to conclude that these represent temporary idiosyncracities in the rating behaviour of the two particular judges involved and are not representative of trends capable of being generalised beyond these judges using the drawings from this sample.

Another aspect of the results that requires interpretation is found in the fact that except for subjects' ratings of their tree drawings, both the ratings of subjects and judges were closer to subjects' self-concepts for copied drawings rather than for whole drawings. One possible interpretation is that these results suggest the need for two differing sets of principles to explain behaviour on the H-T-P. There would seem to be a need for one set to explain the H-T-P test taking behaviour of subjects and the need for another set to explain clinicians' interpretative behaviour. Even though judges' and subjects' data are in the same direction, the similarity in their behaviour is explained theoretically in a different manner. The direction of subjects' ratings was interpreted as a need to control against any possible projection of their individuality that might take place through their rating behaviour. The direction of judges' ratings was interpreted as indicating that the task of interpretation is made more efficient when the number of (expressive) variables in H-T-P

drawings is reduced thereby decreasing the opportunities for some variables to nullify the interpretation given other variables. Therefore, since the reasons given for similarity in subjects' and judges' ratings are different, it seems that a basis is found for the principles governing subjects' and clinicians' behaviours to be different. This argument is supported when the hypothetical aspects of this experiment are again considered. The subjects' results can be said to indicate that Allport is correct in considering expressive aspects as being those aspects of personality usually outside the focus of awareness. It can be also said that the expressive elements can be seen as reflecting aspects of self not readily available to the central focus of awareness. However, judges' results seem to support Hammer's hypothesis that the content aspects of the H-T-P more than the expressive elements, reflect the qualities of personality closer to its central core. It might also be added that these results do not necessarily oppose that part of Hammer's hypothesis that states content aspects rather than the expressive elements reflect the more unconscious qualities of personality. This point is still a possibility as far as the results from this experiment are considered for there was no attempt at any stage in this experiment to establish agreement among the judges as to what specific aspects within the H-T-P drawings would be used to guide judges on rating subjects' self-concepts.

This matter was purposely left to the judgment of each one of the judges. Therefore, the data provided by this experiment seem to indicate that Hammer and Allport do not hold disparate views regarding the expressive aspects of behaviour but actually view the expressive aspects of behaviour from differing vantage points. Hammer's hypothesis seems to be generated from the interpretive behaviour point of view of the clinician whereas Allport's hypothesis seems to arise from subjects' test behaviour on the H-T-P. Before turning to deficiencies of this experiment, which limits the extent to which the interpretation made of its results can be accepted, it can be mentioned that since subjects rated copied and whole drawings at differing semantic distances from their self-concepts, the results here can be taken as some support for the H-T-P tapping different levels of personality integration. This interpretation can be given further support particularly from subjects rating their copied tree drawings at a greater semantic distance from their self-concepts than their whole tree drawings.

3. Evaluation of Results.

The following criticisms can be directed at this experiment. First, the precise effect that providing subjects with personal anonymity had on the degree of their involvement is too difficult to assess. A pilot study to investigate how

ratings might vary in either direction or intensity or both and how H-T-P drawings might vary under different motivational sets provided for in the instructions and procedure should have been carried out. One variation would be to avoid providing subjects with personal anonymity and emphasize that they could discuss their H-T-P results and ratings with a counselor if they so desire. The drawings and ratings of subjects who might take advantage of this offer could be scrutinized for differences with those subjects who do not seek to discuss their ratings and H-T-P drawings. These groups might vary in significant ways to suggest what motivational set indicated the highest degree of involvement. The importance of motivational set on the experimental behaviour of subjects cannot be ignored. Davids and Pildner¹⁶ concluded from an experiment they carried out involving male college students that motivational set has significant effects on personality test performance and is among the most important set of relevant variables to be controlled for in personality research. Hutt¹⁷ also emphasizes the influence this type of variable has on projective technique data.

16 Anthony Davids and Henry Pildner, Jr., "Comparisons of Direct and Projective Methods of Personality Assessment under Different Conditions of Motivation", in Psychological Monographs, General and Applied, Vol. 72, No. 11, 1958, p. 464.

17 Max L. Hutt, "The Assessment of Individual Personality by Projective Tests: Current Problems", in Journal of Projective Techniques, Vol. 15, No. 3, September 1951, p. 391.

Another criticism of this study can be made with regard to rating both drawn and verbal concepts. For example, there is no way of knowing whether subjects rated their self-concepts presented verbally and their drawings, in terms of the value they placed on them (the meaning they assigned to them) or rated both types of concepts in terms of their impressions. This is particularly pertinent to judges' ratings since there seems to be no way to control for judges rating their impressions of subjects' drawings rather than the self-concept of subjects from their drawings.

The fact that both subjects' and judges' ratings were in the same direction suggests that these ratings may have shared direction more as a function of the measuring properties of the S.D. than of the explanations presented earlier. It could be, for example, that the copied drawings caused both subjects and judges to become more extreme in their judgments so to speak, and that since the S.D. sets limits on the number of directions (only two directions are provided for by bipolar adjective scales) and the intensity (three positions on either side of a neutral position provides only three degrees of intensity for each of the two directions) that ratings can take, the S.D. may have some tendency with some raters to weight the ratings in a given direction. This may be similar to what Kelly and Levy¹⁸ refer to when they deduct from experimental results that there may be an upper

¹⁸ Jane A. Kelly and Leon H. Levy, "The Discriminability of Concepts Differentiated by Means of the Semantic Differential", in Educational and Psychological Measurement, Vol. 21, No. 1, Spring 1961, p. 57.

limit to the S.D.'s ability to measure meaning. They conclude that differences in nuances of meaning can give rise, for example, to two verbal concepts, both of which could be members of a group that expresses differences in nuances of meaning and lead to some consensus of use as to which one of the two concepts is more appropriately used in a given context. This, they suggest, is an aspect of meaning to which the S.D. does not seem sensitive. This would indicate that as far as verbal concepts are involved there would be less variation in ratings as raters would not vary their ratings in response to nuances of meaning that are generated from specific concepts being imbedded in given contexts. The more variations in direction there are in S.D. ratings the greater the possibilities that differences between copied and whole ratings might not differ significantly. However, since both judges' and subjects' ratings were on both verbal and drawing concepts it would appear that rating other than verbal concepts would tend to de-emphasize failure to respond to nuances in the meanings of verbal concepts. The fact that subjects were less reliable on their self-concept ratings, since these were completely verbal concepts, than on their drawing ratings vitiates against placing much weight on the degree of significance and on the directionality of the results as being in some significant portion a result of the S.D. failing to account for differences in nuances of meanings.

Another deficiency arises from a difficulty brought out after completion of this experiment. Judges reported the following difficulty. If, for example, judges considered insecurity of a subject was indicated from his drawings it was difficult to determine whether to rate the subject as seeing himself as insecure or whether to rate him as seeing himself as particularly secure which could be seen as a reaction formation type of defense against insecurity. There appears to be no satisfying way for overcoming this difficulty. One possible solution might be to ask judges to rate the "self" defined in terms of what the core or basic personality of a subject is on the basis of his drawings and compare these ratings to those made by independent judges on the "self" in terms of some other test of self. For example, independent judges might be asked to rate the "self" of subjects on the basis of their M.M.P.I. profiles.

Suggestions for further research are implied in these criticisms. In addition to testing the influence of variations in motivational set and other conditions such as the effect of individual as opposed to group administration of the H-T-F on S.D. ratings, it would be interesting to repeat this experiment with samples from populations other than college students. Cassel, Johnson and Burns¹⁹ found some

¹⁹ Robert H. Cassell, Anna P. Johnson, and William H. Burns, "(E) Examiner, Ego Defense and the H-T-F Test", in Journal of Clinical Psychology, Vol. 14, No. 2, April 1958, p. 159.

evidence to indicate that some aspects of H-T-P's (notably size) vary depending on the presence or absence of an examiner. Examiner absence (in individual administration) might be a condition similar to group administration. This latter condition could be seen at least as de-emphasizing the presence of the examiner. Moreover, one aspect (for example, quality of the lines) of H-T-P's could be rated one at a time by exposing, in individual session, to each subject his drawings with one aspect at a time vividly brought out and all other aspects reduced in vividness by using a projector and other illumination techniques. The data for such a study might be best analyzed in a factor-analytic manner to provide such basic information as, for example, the extent to which each expressive element is independent of all other expressive elements. In addition, aspects exposed tachistoscopically, might provide information regarding which aspects appear less conscious; consciousness defined in terms of threshold for recognition.

4. Summary and Conclusions.

In final summary, it can be emphasized that generally the subjects' results of this experiment support Allport's view that expressive aspects of (H-T-P test) behaviour are typically less available to the focus of awareness. Also, the results obtained from judges' ratings generally support

Hammer's contention that content aspects are more useful in providing clues as to the central, core aspects of personality than are expressive aspects. However, the best interpretation of subjects' data was considered to be that subjects controlled against the possibility of projecting any of their individuality into their ratings by rating whole drawings farther from their self-concepts than copied drawings. A possible explanation was made in terms of a need to control fears or conflicts called out in subjects by their whole drawings since these drawings were highly relevant and stronger than copied drawings in the stimulus-pull for such conflicts. Since tree drawings (tree as a concept) are less relevant in this way than the house and figure drawings, subjects can be said to have rated whole tree drawings closer to self-concepts than copied tree drawings and therefore they assigned whole tree drawings a meaning similar to the meaning they assigned their self-concepts than the meaning they assigned their copied tree drawings.

It was indicated that these explanations of results are in essential agreement with current research on projective techniques. Moreover, these results were taken as indicating that Osgood's theory of meaning is presently too elementary in form to provide useful predictions. Repeating this experiment with the innovation on the S.D. suggested by

Flavell²⁰ on the S.D. and comparing such results with results obtained here would do much to indicate the extent to which the S.D. measures that aspect of meaning referred to as connotative as well as provide some criterion against which to further evaluate the present results.

Flavell²¹ points out that the meaning of a sign includes two categories of responses. One class is composed of those responses that are to the properties which the subject discriminates in the object to which a sign refers. The other class contains responses that are to the attributes discriminated in non-referent objects that are present in the milieu that surrounds the referent object. Flavell believes that the S.D. measures the first class of responses but not the second class and therefore he describes two methods whereby the second class of responses can be measured on the S.D. in addition to the first class of responses. Also, by suggesting that subjects' H-T-P test drawing behaviour requires a different set of theoretical principles than clinicians' H-T-P interpretive behaviour, these results can also be taken as making a contribution to theory by indicating this need for two theories. Indeed, this experiment, as well as considered

²⁰ John H. Flavell, "Meaning and Meaning Similarity: I A Theoretical Reassessment", in Journal of General Psychology, Vol. 64, 1961, p. 307-319.

²¹ Ibid., p. 318.

as indicating this theoretical need, could perhaps be considered as premature in that there are no theoretical systems of the kind indicated above with which these results can be integrated so that, against a theoretical background, the adequacy of their explanation can be more properly evaluated. Perhaps time will give to these results a perspective that will indicate the full meaning they have as well as allow for the development of the urgently needed theory, the lack of which so obviously obstructs the search for meaning of much current research. It is more than a truism to note that contemporary psychology is a discipline in which much fragmented knowledge lies fallow, awaiting the birth and development of theoretical systems that can give knowledge meaning.

BIBLIOGRAPHY

Allport, Gordon W., Pattern and Growth in Personality, New York, Holt, Rinehart and Winston, 1961, xiv-593 p.

This book contains one of the latest, up-to-date discussions on the expressive aspects of behaviour. It contains many broad, sweeping statements that have summary value. In some respects it is an up-to-date reworking of Allport's 1937 book on Personality. His summary statements, while valuable, fail to consider all relevant research and tend to use research as illustrative of given points. This selection of research studies is highly selective and not representative. Allport fails to formulate any theory. This seems particularly acute regarding expression.

Buck, John N., Administration and Interpretation of the H-T-P Test, Beverly Hills, California, Western Psychological Services, 1950, 67 p.

This manual is a valuable contribution and contains the main hypotheses generated by the author of the H-T-P. The many hypotheses Buck developed require considerable research to indicate the populations, testing conditions and similar variables under which they are likely valid.

Flavell, John H., "Meaning and Meaning Similarity: I A Theoretical Reassessment", in The Journal of General Psychology, Vol. 64, 1961, p. 307-319.

This publication is of considerable importance since it points to a potential improvement that may be possible on the measuring properties of the semantic differential. The innovations the author suggests have important implications for research completed on the H-T-P and further experimentation. These innovations will require considerable research work likely of many researchers.

Hammer, Emanuel P., The Clinical Application of Projective Techniques, Springfield, Illinois, Charles C. Thomas, 1958, xxii-663 p.

This book provided one hypothetical statement from which this experiment was developed. It is replete with many hypotheses, most of which have not been tested experimentally. Its greatest deficiency is a lack of any beginnings at formulating theory for either interpretation of subjects' H-T-P test behaviour or for the interpretative process of examiners.

Hammer, Emanuel P., The H-T-P Clinical Research Manual, Beverly Hills, California, Western Psychological Services, 1961, 58 p.

This manual contains Hammer's latest thinking on the H-T-P. References and a bibliography while selective contain many of the most significant studies on the H-T-P. It is unthinkable that research on the H-T-P could be carried out without consulting this manual. Hints about but no theory formulation is also a deficiency of this manual.

Homachek, Don E. (ed.), The Self, in Growth, Teaching and Learning, Englewood Cliffs, N.J., Prentice-Hall, Inc., 1965, xi-576 p.

This is one of the most valuable selections of readings in self theory and research to be combined under one cover. It was valuable in planning this experiment for both theoretical and empirical considerations of self, its relation to personality and self-concept measurement.

Kamano, Dennis K., 'An Investigation on the Meaning of Human Figure Drawings', in Journal of Clinical Psychology, Vol. 16, No. 4, October 1960, p. 429-436.

This study is important since it is one of the first to use the semantic differential to test experimentally hypotheses related to the H-T-P. While its deficiencies are many some excellent criticism of it, notably by Pickering, enabled this experiment to improve the precision of measurement.

Osgood, Charles C., George J. Suci and Percy H. Tannenbaum, The Measurement of Meaning, Urbana, Illinois, University of Illinois Press, 1957, 342 p.

No one could use the semantic differential without knowing this book completely. It is difficult to understand only because semantics and measurement are demanding areas within psychology that touch on a prodigious amount of research. This book is likely unique in providing no subject or author index. It is a worthy example of how the work of many researchers over many years have converged in one confluent stream to provide one of the most promising measuring techniques available in current psychological measurement.

APPENDIX 1

DEFINITIONS

APPENDIX 1

Structural Components or Expressive Phase Aspects of the R-T-P Drawings

Hammer considers twelve such aspects as follows:

1. Size
2. Pressure of the line
3. Quality of the line
4. Placement on the page
5. Erasures
6. Shading
7. Reinforcement
8. Proportions
9. Degree and area of completion and detailing
10. Symmetry
11. Perspective
12. Exactness

A brief definition of each aspect will be made along with an outline of the procedures used to standardize each aspect in copied drawings.

1. Size

Size was defined as the height and width of a drawing measured in inches. Both dimensions were determined from that part of each drawing which was extended the most. For example, in most tree drawings the foliage was wider than the trunk. In these drawings therefore width was measured from the widest point of the foliage.

Each copied drawing of a kind (of a kind refers to all the drawings of one concept, i.e., all tree drawings) was made a standard size. Standard size was determined by measuring all drawings for width and finding the average width as specified below. The height of all copied drawings was increased or decreased in proportion to the amount by which they were changed in width to reach the average width. This proportion in each case was determined by an apparatus on tracks that projected drawings onto a screen and determined height change in proportion to width change within .5 centimeters of accuracy.

	<u>Average Width</u>		
House	4.25	inches	
Tree	.50	"	trunk
	4.25	"	foliage
Person	1.25	"	

2. Pressure of Line

There is some evidence that people tend to be constant in the pressure they apply when drawing.

Pressure of the line was defined as the weight used or exerted in producing the lines of the drawings. (If a pentograph can be used it will be -pressure of the line that is defined as the weight placed on the pentograph). Thus pressure of the line was standardized in this manner.

(If a pentograph cannot be used) Judges were asked to rate drawings for pressure of the lines in a manner outlined by Caligor. In this case pressure of the line will be defined in terms of darkness of the line ranging from very dark to very light. Copied drawings judged to vary in darkness or lightness outside the determined range were not used as part of the sample.

3. Quality of the line

All copied drawings were produced with one continuous, straight line of constant width and pressure. With a pentograph line quality could be considered constant for all copied drawings as constant pressure would partially control for quality of the line. The pencil placed in the pentograph was re-sharpened or replaced (as conditions indicated) before each copied drawing was completed in terms of lines. Without a pentograph judges would be asked to rate drawings for line quality and reject any copied drawing considered to vary sufficiently from a continuous, straight single line.

4. Placement on the page

Each drawing had an 8 $\frac{1}{2}$ by 7 inch area on the H-I-P form of available space in which it could be placed. The geometric center of each available area for each drawing was determined as the intersection of two lines one 4 $\frac{1}{4}$ inches from the 8 $\frac{1}{2}$ inch side of the form and the other 3 $\frac{1}{4}$ inches from the 7 inch side of the form. Hammer places the average mid-point of drawing to be approximately $\frac{1}{2}$ inch to the left and above the geometric center. Since this discrepancy between the usual mid-point for placement of a drawing by subjects and the geometric center is small and the important factor was to standardize all drawings all drawings were centered at the geometric center.

5. Erasures

Erasures were defined as any attempt to change all or any part of a drawing by using the erasure on the end of the No. 2 pencil supplied to each subject. No erasures were necessary for the artist to produce any of the copied drawings. There were no copied drawings in which erasures were present.

6. Shading

Shading was defined as any light or heavy, prominent or subtle and uncertain lines which accented any part of any drawing. This definition included shading effect produced by holding the pencil somewhat more parallel to the surface of the paper than in the more vertical position a pencil is usually held for writing, creating an effect similar to that created by charcoal. This expressive aspect was standardized throughout copied drawings by purposely leaving it out of copied drawings. Therefore all copied drawings contained no shading as defined.

7. Reinforcement

Reinforcement was defined as going back over a line more than once producing a heavy dark line, making a line that was once broken, solid or adding any pencil markings to a line already sufficient complete. Every line in all copied drawings will be continuous, unbroken, straight, of constant width and pressure. By standardizing pressure of the line and quality of the line in the manner described above reinforcement was also standardized by purposely not including it as defined in any copied drawing. No copied drawings contained reinforcement as defined.

Erasures, shading and reinforcement were adequately standardized by excluding them in copied drawings. Size and placement on the page were objectively standardized as described above. If a pentograph can be used pressure of the line can be objectively defined in terms of the weight used. The use of a pentograph would do much to also standardize the quality of the line. If a pentograph cannot be used pressure of the line and quality of the line can be rated by judges as an additional check for more precise standardization.

8. Proportion

Proportion was defined as size relationship in three ways. One way was the proportional size relationship of a whole drawing to the form page. This aspect of proportion was standardized when all copied drawings were standardized for size. Another way was the proportional size relationship of a detail (part) to the whole drawing within which that detail lies. The third way is the proportional size relationship of a detail to a detail (a detail was defined as a part identifiable as a complete entity in itself yet part of a greater whole). Examples of details are door, window, chimney, roots, head, arm, hand and leg.

- (a) Detail size in proportion to a whole drawing
- (b) Detail size in proportion to other details in a whole drawing.

9. Degree and Area of Completion and Detailing

This expressive aspect was defined as excessive detailing of a part in relation to the remaining portion of a drawing or less detailing in a part of a drawing in relation to the other parts of the drawing. This definition included a part of a whole drawing left incomplete when all other parts of that drawing are complete or one part of a whole drawing completed when all other parts of that drawing are incomplete. This was one of the expressive aspects considered unstandardized because of its very nature. This aspect as defined was not standardized in copied drawings.

10. Symmetry

Symmetry was defined as size and placement of details and component parts of whole being balanced both horizontally and vertically. This definition included lines of component parts of whole drawings and details being equal in length, shape, quality and pressure. For example, a window on one side of the house drawn being balanced by a window of the same size and type placed in the same place on the opposite side of the drawing. Another example of this would be a chimney on the roof of a drawn house being balanced by steps at the base of the drawn house. This expressive aspect was considered as unstandardized due to its complex nature. Symmetry was not standardized in copied drawings.

11. Perspective

Perspective was defined as spatial relationship in four ways. The first way was placement of the whole drawing on the form page. This was standardized when each copied drawing was standardized for placement on the page and for size. The second way was the relationship of the whole drawing to the observer. This is based on size, height in the vertical plane (distance from the bottom of the page) relative size and position of additional elements (other than whole drawings called for) linear perspective, superimposition of parts of a drawing or additional elements imposed on the drawing, shading, clearness parts of a whole drawing, relative position of the whole drawing (angle of whole drawing to observer) conventional perspective rendering (rendering a three dimensional object on a two dimensional surface-angle of vision) and transparencies. This was one aspect considered standardized only in part. Those parts of perspective dependent on

size, placement on the page, shading and transparencies were standardized. All other parts of perspective most of which have been mentioned above were considered unstandardized in copied drawings.

12. Exactness

Exactness was defined as the precision with which details were produced in size and placement relationship to each other and to the whole drawing in which they were placed. Exactness was also defined as the degree of precision achieved by the drawer in depicting the whole drawing of a particular type. This was another aspect considered impossible to standardize because of its very nature.

Of the twelve expressive aspects a total of eight (if proportion can be standardized) were considered as totally standardized. These were, size, pressure of the line, quality of the line, placement on the page, erasures, shading, reinforcement and proportion. Three other aspects were considered standardized only in part, the greater part of all three were considered as impossible to standardize. These were symmetry, perspective, degree of and area of completion and detailing. The twelfth aspect was considered as impossible to standardize even in part. This was exactness.

Content

Content was defined as the concept of a drawing that indicated to what class of objects it belongs.

House Drawing

This indicates that this is a drawing of that class of objects referred to as houses.

Tree Drawing

This indicates that this is a drawing of that class of objects referred to as trees.

Figure or Person Drawing

This indicates that this is a drawing of that class of objects referred to as persons.

Expressive Elements

Expressive elements are those twelve aspects of H-T-P drawings defined by Buck and Hammer as expressive elements.

Reproduced or Copied Drawings

Those set of H-T-P's reproduced by the artist to control for the expressive aspects as specified. These were considered to have a content emphasis.

Whole or Original Drawings

Those set of H-T-P's as reproduced by subjects and unmodified in any manner. These were considered to have an expressive element emphasis yet combined with content.

The essential difference between whole and copied drawing was considered to be that copied drawings had fewer expressive elements than whole drawings.

APPENDIX 2

INSTRUCTIONS TO SUBJECTS FOR SEMANTIC DIFFERENTIAL RATINGS

INSTRUCTIONS

The purpose of this study is to measure the meaning of certain things by having you judge them against a series of descriptive scales. In taking this test please make your judgments on the basis of what these things mean to you. On each page of this booklet you will find a different concept to be judged and beneath it a set of scales. You are to rate the concept on each of these scales in order.

but not at both ends of a scale.

If the concept seems only slightly related to one end as opposed to the other end (but is not really neutral then you should place your check mark as follows:

active : : : X : : **passive**

The direction toward which you check, of course, depends upon which of the two ends of the scale seems most characteristic of the thing you are judging.

If you consider the concept to be neutral on the scale, both ends of the scale equally associated with the concept or if the scale is completely irrelevant, unrelated to the concept then you should place your check mark in the middle space.

safe _____!_____!_____! X !_____!_____!_____ dangerous

IMPORTANT:

1. Place your check mark in the middle of the spaces not the boundaries.
2. Remember, never put more than one check mark on a single scale.

Sometimes you may feel as though you have had the same item before on the test. This will not be the case, so do not look back and forth through the items.

Make each judgment independent of all other judgments. Work at fairly high speed through the test. Do not worry or puzzle over individual items. It is your first impressions, the immediate "feelings" about the items that we want. On the other hand, please make each judgment with care. We want your true impressions.

Please do not turn the page until instructed to do so.

APPENDIX 3

**SAMPLE OF ONE OF THE 5-SCALE ORDERS OF THE 15 SCALES
USED FOR SEMANTIC DIFFERENTIAL RATINGS**

APPENDIX 3

BY DRAWING OF A HOUSE

sour	very	quite	slightly	neutral	slightly	quite	very	sweet
thin	very	quite	slightly	neutral	slightly	quite	very	thick
slow	very	quite	slightly	neutral	slightly	quite	very	fast
hot	very	quite	slightly	neutral	slightly	quite	very	cold
cruel	very	quite	slightly	neutral	slightly	quite	very	kind
large	very	quite	slightly	neutral	slightly	quite	very	small
passive	very	quite	slightly	neutral	slightly	quite	very	active
sharp	very	quite	slightly	neutral	slightly	quite	very	dull
heavy	very	quite	slightly	neutral	slightly	quite	very	light
rounded	very	quite	slightly	neutral	slightly	quite	very	angular
lstatateful	very	quite	slightly	neutral	slightly	quite	very	tasty
good	very	quite	slightly	neutral	slightly	quite	very	bad
awful	very	quite	slightly	neutral	slightly	quite	very	nice
shallow	very	quite	slightly	neutral	slightly	quite	very	deep
strong	very	quite	slightly	neutral	slightly	quite	very	weak

APPENDIX 4

FACTOR LOADINGS OF THE 15 SCALES

APPENDIX 4

Factor Loadings of Scales Based on Two Factor Analyses.

Scales		Analysis I ^a			Analysis II ^b		
		Factors			Factors		
		E	P	A	E	P	A
E	1. good-bad	.88	.05	.09	2.29	.84	.07
	2. cruel-kind	.82	-.10	-.18	2.40	.49	-.18
	3. nice-awful	.87	-.08	.19	2.39	1.27	-.02
	4. sour-sweet	.83	-.14	-.09	2.29	.71	.14
	5. tasty-distasteful	.77	.05	-.11	2.11	1.05	.21
P	6. weak-strong	.19	.62	.20	.33	1.81	.67
	7. thin-thick	-.06	.44	-.06	-.35	1.48	-.37
	8. small-large	.06	.62	.34	.12	1.76	-.02
	9. heavy-light	-.36	.62	-.11	-1.60	1.68	-.92
	10. deep-shallow	.27	.46	.14	.30	1.46	-.65
A	11. slow-fast	.01	.00	.70	.42	1.10	1.50
	12. active-passive	.14	.04	.59	.30	1.64	1.39
	13. cold-hot	-.04	-.06	.46	.42	.83	.61
	14. rounded-angular	-.17	.08	.43	-1.31	.37	.77
	15. dull-sharp	.23	.07	.52	.51	1.31	1.08

^a Charles C. Osgood, George J. Suci, Percy H. Tannenbaum, The measurement of meaning, Urbana, Illinois, University of Illinois Press, 1957, p. 37

^b Ibid., p. 43.

APPENDIX 5

ABSTRACT OF

Self-Concept, Content and expression in
H-T-P Drawings

APPENDIX 5

ABSTRACT OF

Self-Concept, Content and Expression in H-T-P Drawings.¹

This experiment was carried out to investigate two hypotheses regarding the expressive aspects of behaviour. Allport's hypothesis, that the expressive aspects of behaviour are typically unconscious and reflect the deeper personality structure, was tested by applying it to H-T-P test drawing behaviour. Hammer's hypothesis, that content aspects more than expressive aspects reflect the unconscious qualities that exist at the heart of personality, made with specific reference to the H-T-P, was also investigated. To study these hypotheses experimentally the following procedures were devised and implemented.

One hundred and thirty-three college students between eighteen and twenty-one years of age, rated on fifteen semantic differential scales their self-concepts, each one of their H-T-P drawings as they had completed them and each one of a set of their H-T-P drawings reproduced by a commercial artist to control for as many of the twelve expressive elements of

¹ John A. Machry, doctoral thesis presented to the Faculty of Psychology and Education of the University of Ottawa, Ontario, 1966, viii-139 p.

the drawings as possible. Appropriate estimates of subjects' rating reliability for each of the different rating tasks they completed, were obtained at retesting sessions. Only 118 complete sets of subjects' ratings were available at the end of the data gathering phase due to uncontrollable circumstances. To test one implication Hammer's hypothesis generates, that clinicians may find content aspects more potent for understanding personality than expressive aspects, three judges carried out the following tasks.

One hundred sets of H-T-P drawings were randomly selected and each judge rated the self-concept of these one hundred subjects on the basis of both sets of H-T-P drawings. Appropriate estimates of each judge's rating reliability were obtained at retest sessions. Comparisons were made between D scores obtained from ratings of subjects' self-concepts and original H-T-P drawings, and subjects' self-concepts and reproduced H-T-P drawings. A similar comparison was made for each judge's ratings. The results obtained were interpreted in the following manner.

Subjects rated reproduced drawings significantly closer to their self-concepts than their original set of drawings with the exception of tree drawings. Original tree drawings were rated significantly closer to self-concept than reproduced tree drawings. The first set of results was

interpreted as indicating a need on the part of subjects to deny that original drawings were similar to their self-concepts in meaning. The second finding was interpreted as indicating the tree drawing to be less conflict stimulating than house and figure drawings. Therefore, original tree drawings could be rated closer to self-concepts than reproduced tree drawings in spite of having greater stimulus-pull for underlying conflicts. Original drawings were considered to have the stronger stimulus-pull since they contained a combination of content and expressive elements. When these aspects were combined in the house and figure drawings, the combination of these aspects and these concepts produced sufficient stimulus-pull in order for these original drawings to be highly relevant to revealing individuality and, therefore, subjects rated them farther from their self-concepts than reproduced house and figure drawings. Tree drawing ratings were explained as subjects being able to express their more spontaneous feelings and meanings. House and figure drawing ratings were interpreted as subjects exerting control over their tendencies to potentially project themselves in their rating behaviour. In addition, results produced by subjects were taken as supporting Allport's hypothesis.

Judges rated subjects' self-concepts closer to their ratings of them on the basis of reproduced drawings rather than original drawings. These results were interpreted as

suggesting that H-T-P's containing fewer expressive elements and/or expressive elements controlled, facilitate the task of interpretation as there are fewer chances for certain elements to nullify the interpretation given to other elements. Moreover, judges' results were taken as supporting Hammer's hypothesis.

The results of this experiment, particularly those provided by subjects, were seen as in agreement with other current research which seems to indicate that subjects control their responses to a greater extent than was formerly thought possible. Furthermore, these results were also considered as providing some evidence for the adequacy of applying the levels hypothesis to the H-T-P.

The limitations placed on accepting the results of this experiment due to inadequacies in design were indicated. Criticisms of it were made and the suggestions for further research generated by its results were indicated.

APPENDIX 6

A List of Errors and their Corrections of the Manuscript

APPENDIX 7

ABSTRACT OF

Self-Concept, Content and Expression in
H-T-P Drawings

APPENDIX 7

ABSTRACT OF

Self-Concept, Content and Expression in H-T-P Drawings.¹

This experiment was conducted to test two hypotheses. Allport hypothesis, that the expressive aspects rather than the coping aspects of behaviour, reflect the deeper personality structure, was applied to the H-T-P. Hammer, with specific reference to the H-T-P, states that content aspects more than expressive elements reflect the unconscious qualities that exist at the heart of personality.

College students, using fifteen semantic scales, rated their self-concepts, their H-T-P drawings and these drawings reproduced by an artist to control for expressive elements. Judges rated the self-concept of subjects from their original and reproduced H-T-P's.

Subjects rated their reproduced drawings significantly closer to their self-concepts than their original drawings. This finding was taken as a defensive action by subjects who thereby controlled the extent to which they revealed themselves. This required considering the act of rating itself being seen as potentially revealing of personality. The finding that subjects rated their reproduced tree drawings farther from their

¹ John A. Machry, doctoral thesis presented to the Faculty of Psychology and Education of the University of Ottawa, Ontario, 1966, viii-139 p.

APPENDIX 7

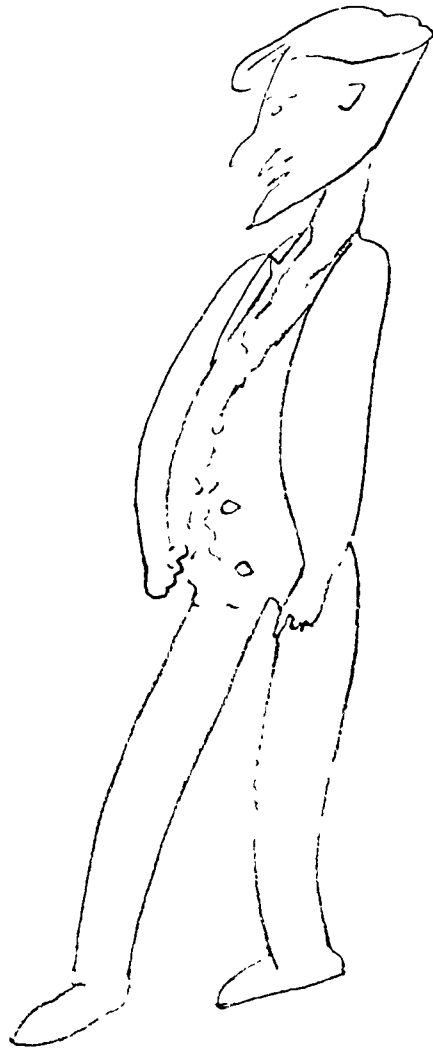
self-contents than their original tree drawings was considered as demonstrating that tree drawings, even containing an expressive element emphasis, are less conflict stimulating than either house or figure drawings and thus subjects did not need to control the possibility of revealing themselves. Judges rated subjects' self-concepts significantly closer to their self-concepts when using reproduced drawings rather than when they used original drawings. This was interpreted as suggesting that reproduced H-T-P's facilitate interpretation by reducing the chances that alternative interpretations will nullify each other.

Both sets of results were said to support the need for two theories, one to account for the drawing behaviour of subjects and another to explain the cognitive processes involved in H-T-P interpretation. Limitations set on the generalization of these results and suggestions for further research were made.

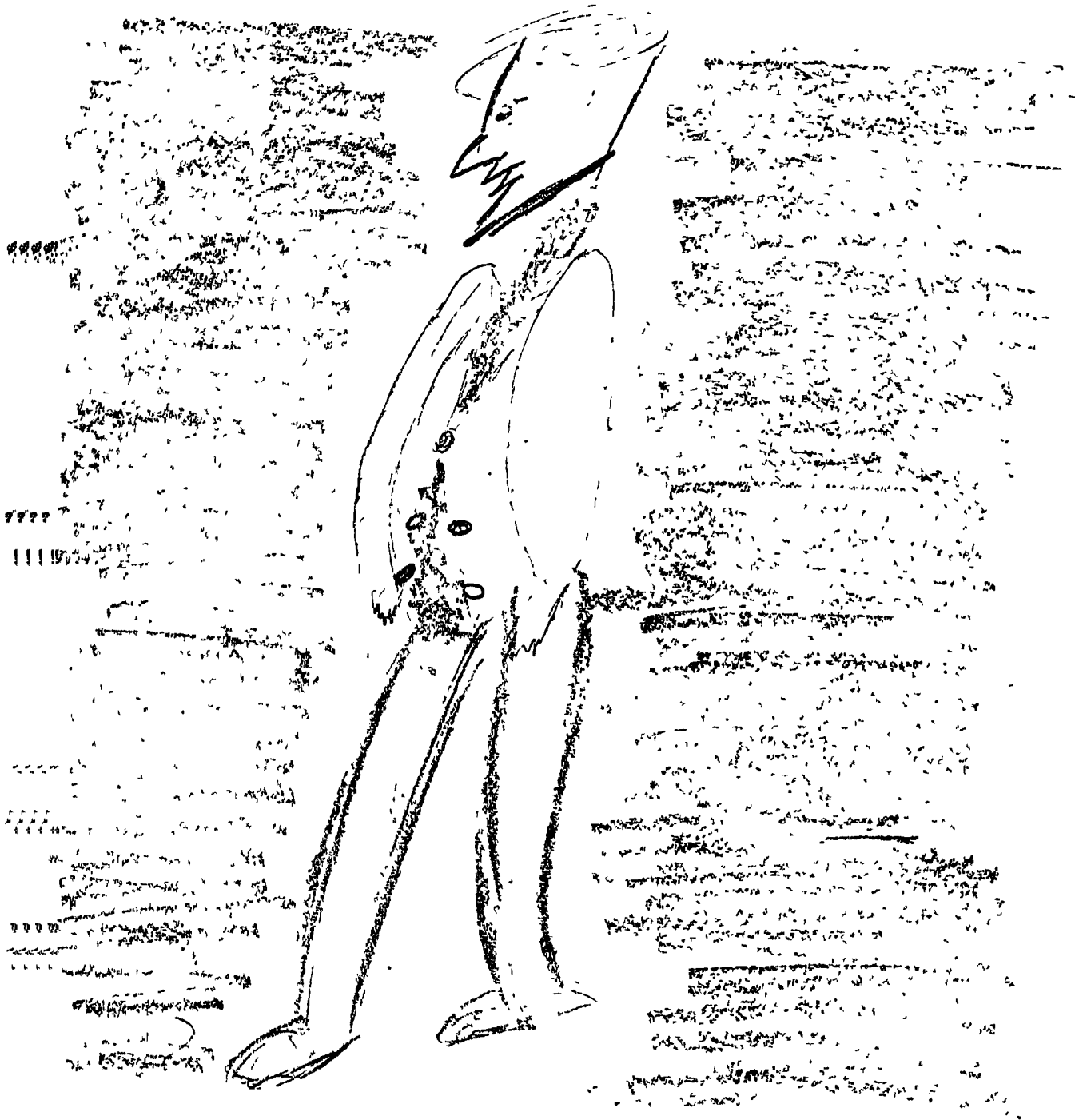
APPENDIX 8

ILLUSTRATIVE REPRODUCTIONS OF WHOLE AND COPIED DRAWINGS

PERSON



PERSON



ERRATA

- PAGE iv Hypothetical Considerations should read Problems of Theory.
- PAGE iv Stimulus Value and S.D. should read Stimulus Value and the Semantic Differential.
- PAGE iv Whole and Copied Drawing Definitions should read Definitions of Whole and Copied Drawings.
- PAGE iv Validity and Reliability of S.D. should read Validity and Reliability of the Semantic Differential.
- PAGE iv Sample should read Subjects.
- PAGE iv Abstract of Self-Concept, Content and Expression in H-T-P Drawings should read Abstract of Self-Concept, content and Expression in House-Tree-Person Drawings.
- PAGE v Probability of Obtaining Given Deviations from Test-Retest on Subjects' Ratings of Whole H-T-P's should read Probability of Obtaining Given Deviations from Test-Retest on Subjects' Ratings of Whole House-Tree-Person Drawings.
- PAGE v Probability of Obtaining Given Deviations from Test-Retest on Subjects' Ratings of Copied Drawings should read Probability of Obtaining Given Deviations from Test-Retest on subjects' Ratings of Copied House-Tree-Person Drawings.
- PAGE vii Footnote one Josephy should read Joseph.
- PAGE viii Line 6 H-T-P should read House-Tree-Person test.
- PAGE viii Line 7 H-T-P should read House-Tree-Person test.
- PAGE viii Line 9 H-T-P should read House-Tree-Person test.
- PAGE viii Line 17 H-T-P should read House-Tree-Person test.
- PAGE viii Line 19 H-T-P should read House-Tree-Person test.
- PAGE 19 Last line rm-----sm should read rm —————> sm.
- PAGE 26 Line 8 nevertheless omit.
- PAGE 26 Line 17 those should read than the self-portraits.

PAGE 31 Footnote 34 Lioselle should read Loiselle.

PAGE 33 Line 17 raised should read raises.

PAGE 37 Line 5 are should read is.

PAGE 37 Line 14 schizophrenia should read schizophrenics.

PAGE 41 Line 7 tancesinvolved should read tances involved.

PAGE 45 Line 9 Hayes⁴⁴ should read Hays⁴⁴.

PAGE 45 Footnote 44 Hayes should read Hays.

PAGE 51 Line 7 even though they were dependent should read even though control of them was dependent.

PAGE 54 Line 6 Loisell⁴ should read Loiselle⁴.

PAGE 57 Line 9 an should read a.

PAGE 63 Line 6 he should read be.

PAGE 64 Line 14 for should read from.

PAGE 66 Line 6 integration should read interpretation.

PAGE 76 Line 16 ratings to each should read ratings and from each.

PAGE 76 Line 18 ratings to each should read ratings and from each.

PAGE 103 Line 16 connotate should read connotative.

PAGE 110 Line 9 However omit.

PAGE 115 Line 10 limits should read limit.

PAGE 138 Line 5 Therefore, original tree drawings could be rated closer to self-concepts than reproduced tree drawings in spite of having greater stimulus-pull for underlying conflicts. omit.