University of Tennessee, Knoxville
TRACE: Tennessee Research and Creative Exchange

Doctoral Dissertations | Graduate School

12-1959

The Cue Value of Certain Attributes of Faces

Joseph William Openshaw
University of Tennessee - Knoxville

Follow this and additional works at: https://trace.tennessee.edu/utk_graddiss

 Part of the Psychology Commons

Recommended Citation

Openshaw, Joseph William, "The Cue Value of Certain Attributes of Faces. " PhD diss., University of Tennessee, 1959.
https://trace.tennessee.edu/utk_graddiss/2950

This Dissertation is brought to you for free and open access by the Graduate School at TRACE: Tennessee Research and Creative Exchange. It has been accepted for inclusion in Doctoral Dissertations by an authorized administrator of TRACE: Tennessee Research and Creative Exchange. For more information, please contact trace@utk.edu.

To the Graduate Council:

I am submitting herewith a dissertation written by Joseph William Openshaw entitled "The Cue Value of Certain Attributes of Faces." I have examined the final electronic copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Psychology.

W. O. Jenkins, Major Professor

We have read this dissertation and recommend its acceptance:

O. M. Pascal, R. R. Shrader, Kenneth H. Moore, Donald D. Holloway

Accepted for the Council:

Carolyn R. Hodges

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

December 1, 1959

To the Graduate Council:

I am submitting herewith a thesis written by Joseph William Openshaw entitled "The Cue Value of Certain Attributes of Faces." I recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Psychology.

W. O. Jenkins
Major Professor

We have read this thesis and recommend its acceptance:

R. R. Shrader

Harold D. Holloway

Accepted for the Council:

Dean of the Graduate School

THE CUE VALUE OF CERTAIN ATTRIBUTES OF FACES

A THESIS

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

---

by

Joseph William Openshaw

December 1959

ACKNOWLEDGEMENT

The author is indebted to a number of individuals whose help and encouragement made this thesis possible. A debt of gratitude is hereby acknowledged to those among my colleagues who gave aid at times when it was most needed. Thanks are also extended to the members of the committee, and especially to Dr. William O. Jenkins and Dr. Raymond R. Shrader whose assistance and guidance were extremely valuable. Thanks also go to Dr. James M. Porter, Jr. for help in securing subjects for the experiment. Finally, a great deal of credit belongs to my wife Lucille who contributed many hours to typing and data processing, and whose help enabled the author to weather several crises during the accomplishment of this task.

J. W. Openshaw

457809

TABLE OF CONTENTS

CHAPTER I

INTRODUCTION

Of all the objects considered to be members of the broad class of things called "stimuli," undoubtedly the most important subset for the understanding of human behavior is other humans. The present research investigates some aspects of the complex stimulus provided by the appearance of another person and the function of selected dimensions of variation of such stimuli in a learning situation. It is an attempt to produce information regarding the relative effectiveness of certain kinds of cues provided by a person's face, when one is required to learn to respond differentially to faces. The importance of research in this general area is attested to by the simple fact that the bulk of mankind's activity occurs in a social situation wherein the important determinants of his behavior are his fellow beings. This is of course the raison d'etre of social psychology, in which the fundamental unit of behavior has been described as one person interacting with another (Kretch & Crutchfield, 1948; Newcomb, 1950). Of course the social psychologists can hardly lay sole claim to the phenomena of persons interacting as a proper subject matter. The clinical psychologists also have an interest in the topic: The term "interpersonal relations" has become quite commonplace in the literature pertaining to psychotherapy and personality theory. It plays an especially prominent role in the theoretical contributions of Harry Stack Sullivan (1948). Sullivan contends, for example, that

neurotic anxiety is primarily social at its inception and this "interpersonal induction and the exclusively interpersonal origin of every instance of its manifestations is the unique characteristic of anxiety." Sullivan is obviously stressing the social nature of behavior problems and because he places so much importance on interpersonal relations, Ruth Munroe calls it the key phrase in Sullivan's system (1955, p. 354). But Munroe further points out that all theorists of the psychoanalytic school have been concerned with our dealings with one another and consider them matters of great importance. In addition to the social psychologists' concern and the clinicians' proper interest in such matters once the "other person" is identified as a "stimulus" the process becomes a suitable problem for general-experimental psychology, since stimuli can be quite readily assigned a place in the S-R paradigm which characterizes this division of psychological activity.

It seems unlikely that any present day psychologist would deny the importance of "other persons" as determining factors in human behavior, but nevertheless it appears worth pointing out that even Clark Hull whose primary interests are usually thought of as far removed from this particular kind of problem expressed a similar conviction and concern. Relatively early in his theorizing Hull recommended conceiving of the external environment as composed of two parts, the inanimate and animate. The behavior of organisms with respect to the animate portions of the environment was declared to be the principal subject matter of the social sciences, and Hull

describes his own goal as "the elaboration of the basic molar behavioral laws underlying the 'social' sciences" (Hull, 1943, p. 16-17).

Additional argument supporting the need for information concerning the process whereby one person serves as the stimulus for the response of another seems almost gratuitous, since it is obviously a process in which we are all involved everyday. However the general process of persons interacting is much too gross a topic to be subjected to direct experimental investigation. A more specific formulation is needed. To achieve this it may be noted that two somewhat different problems are involved here: One concerns the process wherein the action of another person, especially his verbal behavior, serves as a stimulus, the second concerns the process wherein the appearance of another person serves as a stimulus. In the former case, certainly the "action" cannot be considered entirely apart from the person performing the act, to do so is to commit what Osgood has called the "abstraction fallacy" (Osgood, 1953, p. 268). In short, the stimulus provided by the action of another person is more likely to depend upon who the other person is than upon the action alone. An especially important example of a situation in which there is a tendency to overlook the possible significance of the person and consider his action, alone, to be the stimulus occurs in the conduct of certain psychological experiments. It is often assumed that the acts of the experimenter are altogether impersonal. Wells (1956) has drawn attention to the need to consider the experimenter as a variable in certain experimental procedures and he has investigated one aspect of the problem. The results of Wells' research suggest that the experimenter-

variable may not be ignored without risking error, at least in experiments involving the galvanic skin response.

The foregoing is to some extent a digression from the central thesis presented here, but it does serve to indicate the importance of the person as a part of the stimulus complex and the desirability of more information concerning how persons function as stimuli. The fact is that both social psychologists and clinical psychologists have tended to focus attention on actions of persons as stimuli, rather than on the stimulus properties of the person as a relatively static stimulus object. This emphasis is most apparent in the social psychologists' definition of the fundamental unit of subject matter as persons interacting.

Turning to the relatively less complicated aspect of the social stimulus which was designated the appearance of another person, by this is meant the relatively permanent properties of a particular human being as a stimulus object, or to put it yet another way, those characteristics which remain constant across a variety of different actions, from this point of regard of the person as a stimulus there are many questions worthy of systematic investigation, and these lend themselves somewhat more readily to experimentation. Indeed there has been a large amount of research in this area, however nearly all of it has been devoted to one particular kind of problem: the accuracy with which one person judges another. The method most frequently employed involves obtaining impressions or judgments of persons on the basis of their appearance, concentrating usually on the faces, and testing the "validity" of these impressions by comparing them with measures obtained by a different operation.

For example, a number of researchers have attempted to discover if there is a relationship between judges' impressions of another person's personality traits, and measures of the other person's personality traits made by other methods, usually by standard personality tests or appraisal by experts. The evidence from investigations conducted according to this general formulation is inconclusive and somewhat confusing, as might be expected in view of the obvious difficulty of selecting a wholly satisfactory criterion. A related problem that has been subjected to experimental attack is the matter of judging emotions from facial expressions. There is a special distinction between this and the process of judging personality traits in that emotions are a more transitory phenomena and their expression is primarily a matter of temporary changes in the mobile features of the face, whereas the personality traits are regarded as more permanent characteristics of the person and are judged largely from the structural features of the face. The literature on neither of those topics will be reviewed here since they are more or less ancillary to the present research and the general form of the experiments has been described only for the purpose of placing the present research in proper perspective partly by contrasting it with these earlier studies. However an excellent review and evaluation of the literature prior to 1954 is presented by Bruner and Tagiuri (1954).

More recent research has tended to develop along a somewhat different line; a more phenomenological approach has been adopted following a suggestion that to have come from several sources almost simultaneously. Bruner and Taiguri accredit it to recommendations by Mac Leod (1947), Krech

and Crutchfield (1948), and Lewin (1947), while Secord, Dukes and Bevan attribute it to the two symposia edited by Blake and Ramsey (1951) and Bruner and Krech (1950). The outstanding example of this approach is to be found in the initial study by Secord, Dukes, and Bevan (1954). And the trend is continued in the succeeding research by Secord and Muthard (1955) and Stritch and Secord (1956). The distinction between these studies and the kind of research described in the preceding paragraph is that the interest has shifted from a concern over the accuracy of judgments, as indicated by agreement with objective measurement, to an interest in the amount of agreement found among different judges' impressions of a given person. The research reported by Secord, Dukes and Bevan is especially significant in that it demonstrated there was consensus among judges in the personality traits they attribute to the persons viewed. A less surprising but nevertheless important finding in the same investigation was that judges agreed in their appraisal of physiognomic characteristics of persons viewed. The authors devised an index analogous to a reliability coefficient to measure the amount of agreement between judges on each trait. According to the size of the indices obtained, the extent of agreement was quite similar for physiognomic and personality traits. All of their findings were later confirmed in the research by Stritch and Secord, and Secord and Muthard.

One feature that the more recent studies have in common with the earlier studies is that in all of them the subjects, or judges, had a relatively passive role. The appraisals of the stimulus were made in a highly structured situation; trait names were provided and the subject

was required merely to indicate a rating on each trait by marking a scale. This procedure serves well for the indentification of the phenomenal dimensions of the stimulus, and in the scaling of stimuli, but it does not provide direct information regarding the function of these various dimensions in a different setting. It does not indicate, for example, which of two available stimulus dimensions will be most effective as a cue in a learning situation. This is a characteristic of many experiments of the traditional perception variety: they discover what aspects of a stimulus <u>can</u> be perceived, or responded to, but not necessarily what <u>will</u> be responded to if a different procedure is employed. Bruner and Taiguri have pointed to this gap in our knowledge concerning the "perception of people". They noted there are no systematic studies devoted to discovering what features of others are most likely to be noticed by people in various situations (Bruner & Taiguri, 1954, p. 648). The point here is closely akin to the distinction between "potential similarity" and "psychological similarity" made by Wallach (1952). Wallach employed the terms to explain the fact that a series of objects may have certain properties in common and yet not be responded to similarly. In which case they would be potentially similar, but not psychologically similar.

The present research was designed to obtain information regarding psychological similarity of faces as opposed to potential similarity. It sought to discover if selected attributes of persons' faces function effectively as cues in a learning situation and to ascertain the extent of transfer of the learned response to new faces, similar to the first with respect to the selected attribute.

Information pertaining to transfer of training where the training involved is learning to respond to people has particular relevance for the theory and practice of psychotherapy, for as has be pointed out by a host of writers the process of psychotherapy can be conceived of as a learning procedure (Cameron, 1947; Dollard & Miller, 1950; Margaret, 1950; Shoben, 1949). Even Freud made the correspondence between therapy and learning explicit. In the article he wrote for the Encyclopedia Britannica Freud describes psychoanalytic treatment as "a second education of the adult, as a correction to his education as a child" (Freud, 1954). Some writers have been even more specific, designating psychotherapy as primarily a problem in social learning (Cameron, 1947; Rotter, 1954). To illustrate how the present research may be related to the conventional concepts of clinical psychology consider the phenomenon of transference which as the word is used in clinical psychology refers to the patient behaving toward the therapist in ways that the patient had behaved toward other significant people in the past. Translated into the common terminology of learning theory transference becomes an example of "generalization" (Dollard & Miller, 1950, p. 260), which is a term refering to the frequently observed fact that a response which a subject has learned to make in the presence of a specific stimulus, can then be evoked by a variety of stimuli not identical with the original, but similar to it in some way (Osgood, 1953, p. 350). In other words, the subject has learned to respond to a class of stimuli instead of a particular stimulus. "Transference", then, appears to be a special case of generalization in which the stimulus objects are other persons. This particular translation of the concept of transference into the generalization concept of

learning theory has been more carefully explained by Bueno (1955). Bueno investigated the generalization of a conditioned autonomic response, where the conditioned stimulus was a picture of a person, and the "test" stimuli were pictures of persons differing from the conditioned stimulus in age or sex. Unfortunately Bueno's research was plagued by technical difficulties which prevented drawing unequivocal conclusions from the results. But his study appears to be the only attempt made heretofore to study the generalization of a learned response where the stimuli were pictures of persons. The present research was aimed at gaining information about how pictures of human faces function as cues for a learned, voluntary response, and whether or not selected attributes, shared by faces, make them similar in the sense that a response learned to one transfers to another which shares the attribute.

Also because there is some evidence (Secord & Muthard, 1955) to support the popular belief that the sexes differ in their ability to discriminate among persons' faces, the research was designed to permit investigating the performance of each sex separately.

Statement of the Problem

Specifically the present research sought answers to the following questions:

1. Can persons learn to respond to faces on the basis of physiognomic traits characterizing a group of faces?

2. Can persons learn to respond to faces on the basis of personality traits which judges agree characterizes a group of faces?

3. If the answer to one and two is affirmative, is learning to respond to one kind of trait in faces more readily accomplished than learning to respond to the other kind of trait?

4. Can the subjects verbailze accurately concerning the particular trait in the faces to which they learn to respond? While not an essential part of the study, the information is easily obtained and may furnish useful auxiliary data.

5. Finally, is there a difference between males and females in the ability to do any or all of the things described in questions one through four?

It should be noted that the research on all of these questions was limited to college students as subjects and the use of black and white photographs of faces as stimuli.

CHAPTER II

METHOD

Overall Plan

The general plan of the research is to study peoples' ability to learn to respond to selected attributes of a stimulus. The particular kind of response chosen for use in the study was a simple indentification or labeling response. The Ss were required merely to assign the stimulus objects to one of two categories. Recommendation of this method as a particularly useful "defining operation" for psychological similarity has been made by Walla ch (1952). Furthermore Leeper (1951) has presented a convincing argument that discrimination learning, conditioning and some inductive concept formation may be properly, grouped together as examples of a single process. Leeper's point is that even in a very simple example of learning such as is found in classical conditioning experiments, the term "stimulus" is misleadingly simple. The total stimulus situation contains a great deal more than the single stimulus. For example consider the classical conditioning procedure, in which a dog comes to salivate in response to a signal light (the so-called conditioned stimulus) after the signal light has regularly preceded the presentation of meat powder (reinforcer). Actually the signal light is only a small part of the total laboratory arrangement confronting the animal. The gross situation has many features which are common to both the conditioned stimulus and the neutral stimulus

(the period between trials): These are such things as the walls of the room, restraining straps, odors, noises, etc. On the other hand the gross situation containing the conditioned stimulus is never exactly identical from one presentation to the next: The animal may have changed its posture, so that the light comes from a slightly different angle; or some of the background features may have been inadvertently changed, an extraneous noise may have ceased since the previous trial, etc. Because of the true complexity of the situation Pavlovian conditioning is more properly regarded as a process in which an organism eventually comes to respond to a variety of situations which share a particular common property. It also must learn not to respond to every feature of the stimulus situation, since many of these are common to the "neutral" stimulus. Thus it is very simular to the process in which a human learns to separate objects into classes on the basis of a property common to some, but not present in others. In view of the essential similarity between the various kinds of learning the present study is not identified as an example of concept formation, as opposed to discrimination learning and/or transfer-of-training; nor will the identification as an example of any of these be denied. The same experimental paradigm is often employed for investigating each of these processes, (Osgood, 1953; Vinacke, 1951; Woodworth & Schlossberg, 1954), and in the present case wherever results of previous studies expressed in the terminology of any one of these frames of reference were deemed relevant to the present research they were cited. The general procedure employed in these experiments requires S to learn to classify the stimulus

objects of a training set or series correctly according to some principle decided by E. Usually it involves simply a dichotomous classification. After learning to classify the objects of the training set, S is presented with a second set of objects which can be dichotomized on the same basis as the training set according to E, and the degree to which S is able to perform the second task at a level greater than that expected by chance, or to achieve mastery of the second task more rapidly than he did the first task, is taken to be evidence of "concept formation," or "transfer of training," or "generalized responding," depending on the conceptual framework preferred by E.

The similarities among the various procedures employed in investigating learning are becoming increasingly evident. For example, a recent experiment reported by Green (1955) is titled Concept formation: a problem in human operant conditioning. Green conducted his study using an apparatus which can be accurately described as a Skinner box for humans. He presented cards bearing different patterns of dots as stimuli and S "classified" them by pressing a telegraph key if the card was a "correct" card, and not pressing the key if the card was "incorrect." Correct responses were reinforced by awarding "points" to S, and S could earn more than one point by repeatedly pressing the key in the presence of the "correct" stimulus. In Green's experiment the correct cards had a common pattern of dots which did not appear on the incorrect cards.

The significance of Green's research for the present study is that it illustrates the close correspondence between "learning" and the process referred to as "concept formation."

The method and results of the research reported here will be discussed primarily in the terminology common to most learning theory: The terms transfer of training and generalization will be employed to describe what others might call an instance of concept formation.

In the present study, the objects to be classified are pictures of the faces of adult male humans, and the basis of classification in one case will be marked presence or absence of a personality trait which judges agree is expressed in the faces. In a second case, and for a different group of Ss, the basis of classification will be a physiognomic trait that serves to distinguish the faces--this trait also will be determined from judges' ratings. Finally a third group of Ss will learn a similar classification of faces, but one where there is no systematic distinction: that is, no particular trait by which S can distinguish the members of one class from the other. The group in the last condition will serve as a control group.

The overall design of the experiment can be represented as a two factorial design with one factor designated sex of subjects, and the second factor the type of task, as described in t he preceding paragraph. The experiment may also be correctly described as a two by three factorial, with independent Ss in each cell. The form of the design described in this manner is represented by the arrangement shown in Figure 1. At the risk of overcomplicating the matter at this point it might be indicated that the design can be most adequately respresented if the training series and test series, (or first learning and second learning) are viewed as constituting a third factor, and the experiment conceived of as a three

Sex	Type of Task: Personality Trait	Physiognomic Trait	Control
Male	10 Ss	10 Ss	5 Ss
Female	4 Ss	4 Ss	2 Ss

Figure 1. The basic design of the experiment.

factor design: two by three by two. Although strictly speaking this is not a true third factor, the distinction in this case is not thought to be to important. The design corresponds to what Edwards (1950) has called "design involving repeated measurements." The three factor model best describes the shape of the design employed in obtaining data in the final stages of the study.

The overall program of research divides naturally into three successive stages: 1. Selection and preparation of the stimulus materials; 2. Scaling the stimuli, selection of suitable trait dimensions to be used in the final phase of the experiment, and assembling the sets of pictures to be employed in the final phase; 3. Conduct of the learning and transfer experiments followed by analysis and interpretation of the data.

Each of these steps requires description in some detail.

Selection and Preparation of Stimuli

The decision to use black and white photographs was based primarily on the line of reasoning presented by Secord, Dukes, and Bevan (1954). The essence of the argument is that photographs are a satisfactory compromise; They permit better control of the stimulus than is the case when live humans are used, but, on the other hand, they lose some of the abundance of cues provided by the latter. The use of photographs as stimuli places the experiment somewhere between the artificiality of the usual laboratory situation and the real-life situation.

It was further decided to utilize only faces of adult males, in order to minimize extraneous variables and permit exploring the independent effects of physiognomic traits and personality traits. The photographs were selected from random issues of Time magazine published within the period April 1957 to April 1958, according to the following set of criteria: Full face or three-quarter profile photographs, approximately two by three inches in size, head to shoulders, not appearing to show a strong emotion (although some were allowed whose expression might be described as bordering on anger), of persons not so well known that the average college student would be likely to recognize them.

Every photograph that met these requirements was taken for inclusion in the sample until a total of fifty photographs had been obtained. It was recognized that the sample obtained by this procedure would be representative of only a selected part of the population, namely adult males who, for whatever reason, have their pictures appear in news magazines. But within this limited group, it is believed the procedure produced a fairly random sample.

The background of the photograph was cut away from the figure, but the portions of the photographs showing the clothing were left in place. This was a change from the practice of some of the previous experiments, but in the present study it was not believed necessary to deprive the Ss of whatever cues are provided by a person's dress. The preparation was completed by mounting each photograph on a plain, white, four by five

inch card.

Scaling the Stimuli and Preparing the Sets

This part of the research constitutes a sub-experiment. The operations used in this stage of the research derive almost entirely from the method and results reported by Secord, Dukes, and Bevan.(1954). It was deemed advisable to make full use of the information already available from the previous studies in order to accomplish the scaling as directly and economically as possible. It will be recalled that the rationale for the present research rests in a large part on the results of the research reported by the forementioned authors.

A rating sheet was prepared for use by the judges. The sheet contained a list of twenty-two traits upon which each photograph was to be rated. The list included eighteen personality traits and four physiognomic traits. The list is reproduced in Table I and a sample copy of the rating sheet is provided in Appendix A.

The traits were selected in the following manner from the original list presented by Secord, Dukes and Bevan. Those authors reported the results of a cluster analysis performed by them which revealed six groupings of personality traits; four of these groups appeared to be relatively independent. The traits for use in the present study were selected to represent each of the four clusters in the belief that trait ratings within one cluster might well be considered ratings of a single, more basic, trait. From within the trait clusters, traits were selected on the criterion of high "objectivity" indices (analogous to high reliability), reported by Secord, Dukes, and Bevan.

TABLE I

TRAITS UPON WHICH PHOTOGRAPHS WERE RATED. PERSONALITY TRAITS ARE GROUPED BY TRAIT-CLUSTERS. PHYSIOGNOMIC TRAITS SHOW ANCHOR WORDS. TRAITS SELECTED FROM LIST PRESENTED BY SECORD, DUKES & BEVAN (1954)

Personality Traits	Physiognomic Traits
Warmth-Tolerance Cluster	Age (young face - old face)
Cheerful Appearance	
Sense of Humor	Complexion (light - dark)
Likeable	
Honest Face	
Kind Face	Fullness of Lips (narrow, thin - thick, full)
Warmhearted	
Moral-Social Responsibility Cluster	Eyebrows (light - heavy)
Intelligent Look	
Conscientious	
Air of Responsibility	
Air of Refinement	
Distinguished Look	
Forcefulness Cluster	
Self Confident	
Alert Expression	
Determined Look	
Energetic	
Aggressive Look	
Aloofness Cluster	
Proud	
Reserved	

In addition, it should be noted that two of the choices were made in part on an intuitive basis; the traits so selected were Determined Look and Aggressive Look and the inclusion of these two traits in the cluster labeled Forcefulness was not indicated by the previous work. Instead these "traits-reflected" appeared in a cluster labled Meekness. Ultimately neither of the two traits were used in the final phase of the study, so the liberty taken in this instance is not really due any consideration in evaluating the final results of the present study.

The selected traits, grouped according to clusters, and the cluster names suggested by Secord, Dukes, and Bevan, are presented in Table I. Also the groupings are indicated on the Record Sheet for Photographs which was prepared to facilitate recording the accumulated data for each photograph; a sample copy of this record sheet is available in Appendix B.

The four physiognomic traits were selected on the basis of high objectivity indices, and one further consideration: Complexion was included because it, like Age, is a characteristic of the whole face rather than some relatively minute detail or single feature. Since the personality traits appear to represent the whole face it was believed that overall physiognomic traits would be more suitable for use in the final phase of the research. The four physiognomic traits selected for scaling were: Age, Complexion, Fullness of Lips, and Heaviness of Eyebrows; these also appear in Table I. The total of twenty-two traits in the completed list was based upon previous

experience with this type of instrument, which had shown that this number of judgments constitutes a task of acceptable size for the average rater. It was also expected that not all of the traits would prove sufficiently reliable to permit scaling, and use of a large number of traits at this stage appeared advisable so that there would be some "spares."

The final form of the rating sheet consisted of the rater's instructions at the top of the sheet beneath which appeared the list of the trait names each followed by a seven point scale. A sample copy appears in Appendix A. Raters were required to rate each photograph they judged on every one of the traits; Each rater worked on one photograph at a time, independently of the other raters and photographs. As an additional check on the selection criteria of the photographs, the raters were asked to indicate recognizing a face by noting that fact on the rating sheet. The use of this seven point scale and rating technique was dictated by the use of this scale and technique in the forementioned work of Secord, Dukes and Bevan, and the assumption that the fewer the deviations from the demonstrated effective technique, the better the chances of obtaining consistencies in judgments. Unfortunately, it is a technique that does not lend itself very well to constructing a scale, which is a disadvantage in the present study, as is indicated a little later in the report. The rating procedure is fairly standard and will not be described in any detail. It was, of course, necessary to designate the polarity of the scales representing the physiognomic traits. This was done by appending an appropriate word to each end of

the numerical scale on the rating sheet, eg. for Age the low end of the scale was designated "young" and the upper end designated "old". The anchor words for the other traits are shown in Table I.

Volunteer students recruited from the undergraduate psychology courses at the University of Tennessee served as raters in rating the photographs. A total of sixty-two raters, forty-three females and nineteen males, participated, but it was not possible to have every rater rate every photograph. To do so would have been too much of an imposition on the raters. It required approximately forty-five minutes for a rater to complete ratings on ten photographs. Therefore the plan followed consisted of obtaining judgments from each of the first group of nineteen Ss (who were fortunately available for two one hour sessions) on every one of the first nineteen photographs. Then the data obtained from this trial group was assembled and inspected in order to ascertain the degree of consistency obtained with the measuring device. Since the ratings made by the trial group showed an acceptable amount of agreement, it was decided to utilize the remainder of the available Ss by having each S rate ten photographs. This plan was carried out, but unfortunately due to difficulties in scheduling the work the raters were not distributed uniformly over the photographs. As a result the number of raters was not exactly the same for all photographs but there were no fewer than ten raters for any photograph, and some photographs were rated by as many as twenty-four persons.

The ratings obtained were compiled on a separate record sheet for each photograph. The actual data were considered too bulky for

inclusion in this report, but it is available. An inspection of the frequency distribution of ratings on each trait by photographs showed sufficient agreement among the ratings on the personality traits within the clusters labeled Warmth Tolerance, Moral-Social Responsibility, and Forcefulness, to warrant a further attempt to scale the photographs on these factors. The ratings in the cluster labeled Aloofness failed to show much agreement; The cluster was therefore dropped from further consideration in this study. The ratings on the physiognomic traits exhibited marked agreement.

There appears to be very little in the way of a standard procedure for combining ratings to produce a single scale value for each stimulus from the data obtained by this simple rating method. Guilford (1954) has suggested an analysis of variance technique, but the method assumes every rater rates every item, and for even a small number of raters, items, and traits the computational labor required is enormous. For the present problem it would have been prohibitive.

Since the research as planned did not necessarily call for stimulus photographs scaled on an interval scale, but instead only required establishing sets of stimuli which were clearly divergent on the various bipolar traits, a method for accomplishing this was devised. The median rating was decided upon as an estimate of the locus of a photograph on the seven-point scale of each trait. As it developed, the personality traits and physiognomic traits posed different scaling problems. The median rating on every trait was computed for every photograph. Inspection of the medians revealed that within the personality

trait clusters there was a decided tendency for a high rating on any one trait to be associated with a high rating on all of the other traits and conversely. A frequent exception to this rule was the Honest Face trait in the Warmth-Tolerance Cluster. Since Secord, Dukes and Bevan also expressed some surprise over the placement of this trait in the cluster, the decision to ignore it as a factor in the present study was made at this point. In view of the impressive evidence that the remaining personality subtraits within clusters were not independent, it became necessary to regard the subtraits as different indices of a single underlying trait. Consequently a consistency measure was chosen as the most appropriate method of indentifying photographs that were high or low on these bipolar traits: The Warmth-Tolerance trait the Moral-Social Responsibility trait, and the Forcefulness trait. The rationale involved is quite simple: If the medians of the subtraits are all estimates of a single dimension, and if all five of the five estimates for a given photograph deviate in the same direction from the overall median for all photographs, it can be stated with considerable confidence that the photograph is distinctly different on that dimension, since the probability of such consistency resulting by chance alone is .031.

If the same procedure is followed to detect photographs that are either high or low on the trait, the two-sided probability value is more appropriate, which means simply doubling the previous value, .031, yielding .062.

The operation as described was employed to identify the photographs representing the high and low categories on each of the personality traits.

The grand median for use in the operation was estimated for each of the traits from an overall frequency distribution of ratings taken over raters, photographs, and subtraits. The actual frequency distributions obtained can be found in Appendix C.

The operation identified twenty-nine photographs which were distinguished on the Warmth-Tolerance trait: thirteen high, and sixteen low. Twenty-five photographs were distinguished on the Moral-Social Responsibility trait, thirteen high and twelve low. But only fifteen photographs were found to be distinguished on the Forcefulness trait. The Forcefulness trait was therefore dropped from further consideration in the experiment, since it was desired to have at least twenty photographs from which to form two sets of ten photographs each, capable of being sorted dichotomously on a given bipolar trait, for use in the final stage of the study.

The operation for identifying photographs for the high and low categroies on the physiognomic traits was somewhat different, but it is believed to be an approximately equivalent operation. The operation was designed to establish classes of photographs as divergent on the physiognomic dimensions as were those obtained on the personality dimensions. A technique equivalent to converting the medians to standard scores was employed. The method consisted of constructing an overall frequency distribution of the ratings for all photographs on each trait, calculating the "grand median" and the variance for each of the overall distributions, and from these establishing 95 per cent confidence limits for medians based on samples of sizes corresponding to the number of judgments on the individual photographs. These limits were computed

according to the formula:

$$\text{Grand Median} \pm 1.96 \frac{1.253\ \sigma}{\sqrt{n}}$$

Because of the varying number of judgments made on the different photographs it was necessary to compute a set of limits for each case, but the general procedure was to assign a photograph to the high or low category on a physiognomic trait if its median rating on the trait fell outside the 95 per cent confidence limits.

Since only two personality traits had been found suitable for use in the final stage, only two physiognomic traits would be needed; therefore photographs were tested only on the Age and Complexion traits. The method described identified twenty photographs that were distinguished on the Age trait; but, unfortunately, twelve were high (old) and only eight were low (young). The explanation for this is quite apparent; persons whose faces appear in news magazines tend to be older persons. Since it was essential that there be at least ten photographs available in each category, the records of the individual photographs were examined, and four were selected for which median ratings on Age were extremely low and upon which the raters had tended to agree. Inclusion of the two photographs which were eventually selected from these four for use in the final phase, involving ten photographs, amounted, in effect, to utilizing 93 per cent instead of 95 per cent confidence limits as criteria. This still compared favorably with the criteria used for the personality traits, which it may be recalled amounted

to 93.8 per cent confidence limits.

In the case of the Complexion trait the results were somewhat less satisfactory. The first selection based upon 95 per cent limits yielded only twelve photographs, six high (dark) and six low (light). In order to obtain full sets of ten in each category it was necessary to reduce the criteria to approximately the 70 per cent confidence limits. And in view of this, an additional procedure for establishing the stimulus sets appeared advisable. The procedure employed is described in the following paragraph. For this added procedure it was desirable that there be at least one "spare" photograph available in each category. For the Complexion trait, in order to include eleven photographs in each category, it was necessary to reduce the criteria to the 60 per cent confidence limits.

Because the procedure for selecting photographs to comprise a set had become somewhat less than uniform, as a precautionary measure an additional test was made on every set of stimuli to be certain they were dichotomous with respect to the intended trait. This was further indicated as a useful course of action since with at least one "spare" photograph available for each category of every set, a further refinement of the sets might be effected. These tests were conducted in a straightforward fashion. The photographs constituting a trial set were given to individual judges who were instructed to sort them into two equal piles on the basis of the appropriate trait. None of these judges were persons who participated in making the original ratings on the photographs. The trait was carefully described to the judge and he was allowed to sort and re-sort the photographs until he was satisfied

that the classes were correct. However he was not permitted to spread the entire set of photographs out before him and view them simultaneously; he was permitted to see no more than four photographs at one time. This approximated the manner in which the stimuli were displayed to the Ss in the final phase of the experiment.

In all, four judges sorted each set of photographs; Two of the judges were persons familiar with the study being conducted (these included the writer) and two were not. A photograph was retained for use in the set if at least three of the four judges placed it correctly. Actually reasonably high agreement was obtained in these initial sortings: For the classifications based on Warmth-Tolerance, Moral-Social Responsibility and Age, each of which contained twelve photographs per category, the proportion of disagreement by all four judges was only 10/96, 14/96 and 2/96 respectively; for the classification based on Complexion, for which there were only eleven photographs per category, the proportion of disagreement was 8/88. Eliminating the photographs which failed to meet the specified criterion (accurate placement by at least three judges) reduced the amount of disagreement to 4/80, 4/80, 0/80 and 5/80 respectively. This was taken to be sufficient evidence that the sets of photographs could be consistently sorted. Additional evidence to support the conclusion was obtained by spreading the photographs on the table, so that the members of the sets could all be compared simultaneously. The four judges upon viewing the photographs displayed in this manner, agreed without exception on the correctness of the classifications.

Thus were established from among the fifty original photographs, alternate arrangements constituting four sets of photographs. Each set contained twenty photographs. And each set could be systematically dichotomized on the basis of either a bipolar personality trait, or a bipolar physiognomic trait.

Each set of twenty photographs were then separated into two sets of ten which contained five "high" and five "low" on the given trait. Or to put it another way, each set of twenty could be viewed as consisting of ten pairs of photographs one "high" and one "low" on the given trait. These were then divided into two sets of five pairs each.

The assignment of photographs to the sets in the last step was done in a manner that provided approximate matching of the sets. Photographs were assigned to alternate sets on the basis of the median rating on the trait in the case of physiognomic traits, or according the number of subtraits with medians outside the 95 per cent confidence limits in the case of the personality traits.

Two of these sets of ten photographs each capable of being dichotomized on the basis of the same trait provided suitable stimulus materials for use with an experimental group in the final phase of the research. In the final part of the experiment each S was required to learn to sort one set of ten photographs into the two classes of five each, and was subsequently tested with the second set of ten, to ascertain whether or not the training would transfer from one set to the other.

There remained only the task of constructing "control sets" of photographs, which by definition whould have no factor systematically distinguishing between the two classes. What was required were two sets of ten photographs with each set of ten arbitrarily divided into two groups of five in such a way that there would be no identifiable group differences. These sets were formed by carefully selecting photographs from among the previously established sets so that the "known" factors at least were ambiguous. For example, if one "old" face appeared in one group of five photographs and "old" face was included in the companion group of five so that Age would not be a distinguishing factor between the groups of photographs constituting a control set. Subsequent inspection by judges revealed no systematic phenomenal difference operating to distinguish one "class" of the control set from the other.

The Conduct of the Learning and Transfer Experiments

The final design of the experiment depended in part upon the kind of materials made available by the preceding operations. These gave the final phase of the experiment the following form. It was possible to designate four experimental conditions and a control condition according to the traits serving to separate the stimuli. Therefore there were four experimental groups, identified, for convenience, by the name of the trait involved in the sorting: A Warmth-Tolerance group, a Moral-Social Responsibility group, an Age

group, and a Complexion group. Among the experimental conditions the first two are properly located under the more general heading of personality traits, and the last two under the heading of physiognomic traits. The final design of the experiment can be reconstructed from Table III, p. 39. The sex factor was included as a second independent variable. Table III can beeseen to be the same as Figure 1 except that in the former specific experimental conditions are indicated.

The experiment was conducted on an individual basis, with each S tested separately; but the procedure was the same for each S in every group. It was, as has already been described, learn task I to a criterion, then be tested on task II, in this instance, by learning task II to the same criterion of mastery. (For a more detailed discussion concerning the use of this experimental paradigm see p. 13 of this report.)

In the present experiment a modified card sorting procedure was decided upon as the most appropriate kind of task in which to measure learning and transfer. This type of task has been discussed at leggth by Berg (1948). Essentially it is a compromise between the proceedure in which objects comprising the classes to be discriminated are presented singly or successively, and the method of simultaneous presentation. It was believed in this instance that the successive method would produce a very difficult learning task, and the simultaneous method would present too easy a task. The literature in this regard is not absolutely conclusive; Some investigations has shown one method superior and others the reverse (Woodworth & Schlossberg, 1954, p. 588), but in the absence

of incontrobertible evidence, the position described in the foregoing sentences appears most tenable.

The method used in the present experiment also permitted S to manipulate the stimulus objects to some extent and direct participation by S in this manner has been shown to facilitate learning (Davidon, 1952).

One other factor had to be considered in choosing the method; this was the number of stimuli representing the classes between which S must learn to discriminate. In the present case there were five stimuli representing each class, which is a relatively small number. It has been established that extensive rather than intensive training facilitates training with these kinds of materials (Reed, 1946). That is to say, the greater the variety of objects representing each class, the more readily the distinction is learned.

Taking all of these factors into consideration the particular procedure used here was designed to provide a task of suitable difficulty. "Suitable difficulty" means a task sufficiently difficult so that the effects of the experimental variables might be detected.

The procedure used in the experiment was as follows. S was seated before a small, plain table upon which were located two place cards marked "A" and "B". S was presented with one of the prepared decks of ten photographs, the stacked deck being placed face up upon the table before S. S was instructed to sort the faces into two sets of five labeled simply group A and group B. S assigned photographs to categories by taking the top photograph from the deck and placing it before the appropriate place card. To facilitate recording the responses

S was asked also to call out his choices as he made them. After each choice E announced whether is was "Correct" or not. If S's choice was incorrect, S was instructed to move the misplaced photograph to the correct stack. The situation resulting from this procedure enabled S to see at most four photographs at any time, and usually no more than three. He could see the next photograph to be assigned and he might see the last photograph assigned to each category. The latter, when they were available, provided correct examples of faces belonging to each category, that is a correct "A-face," and a correct "B-face".

Ss in the experimental groups were told that there existed a principle for distinguishing between As and Bs and futhermore that it was something about the faces that separated the As from the Bs. S was urged to try to discover the difference between the As and the Bs. Providing S with the information that a principle is involved was based on a recommendation by Woodworth & Schlossberg (1954, p. 613).

After each sorting, while E was rearranging the photographs for the next sorting, S was asked to express, in writing, any ideas he had concerning the distinguishing features of the stimuli.

When S had completed a sorting of the ten photographs E took the set, rearranged them according to a preplanned randomized series, and returned them to S to be sorted again. This procedure was continued until S had sorted the entire set correctly on two successive occasions, or until a total of ten sortings had been made. In either case training on the initial set was terminated and S was then presented with

the second or transfer, set of ten photographs which could be correctly dichotomized on the same basis as the first set. S was informed that the same principle involved in sorting the original set could be applied to sort this second set correctly, and he was instructed to continue with the second set in the same manner as before, sorting it into As and Bs. The same procedure that had been used with the first set was followed with S being required to sort the photographs until two successive errorless sortings were achieved, or until a total of ten sortings had been made.

In every case S was urged to try to learn the task as quickly as possible. The verbatim instructions used in the experiment are in Appendix D. Throughout the experiment, except while arranging the stimuli, E was stationed a few feet to the side of the sorting table and recorded the responses on the record sheet. The record sheet was shielded from S's view.

Ss in the control groups were treated exactly like the experimental Ss, except that they were explicitly informed that there would be no systematic difference between the As and Bs. They were told the assignment of photographs to classes had been made in a completly arbitrary manner, and it would be necessary to solve the problem by memorizing. It may be recalled that the photographs for use in the control condition were selected so that there would be no trait distinguishing one "class" from the other, and information given the Ss in this regard was essentially correct.

For the Ss in each group, the order of utilizing the two sets

of stimuli was varied in a manner intended to minimize, in the overall results, any possible sequential effects due to differences between the sets. The set serving as the training set for some Ss, served as the test set for others and vice-versa.

The order of appearance of the photographs within a set on each trial was a randomized series with the restriction that no more than three consecutive stimuli belong to the same class.

Ss were permitted to work at their own pace during the sorting. The interval between trials varied between one and two minutes depending on E's speed in rearranging the stimulus materials.

The experiment was conducted in a private, air conditioned room.

Following the experiment, Ss were asked a few questions in a rather informal manner. With all of the stimulus photographs of the set with which he had worked spread out in view before him, S was asked if he could then verbalize the difference between the classes (if he had not already done so).

For the Ss in the control group the informal questioning and inspection of the stimulus photographs was aimed at discovering if S had been able, or was then able, to find some systematic difference between the classes. It was, of course, believed that he would not, but the additional information served as a further check on the adequacy of the technique used in preparing the stimulus materials.

Thirty-six Ss participated in the learning and transfer part of the experiment. All were students from undergraduate psychology courses at the University of Tennessee. These were not the same courses from

which Ss for the first part of the experiment were drawn. None had been raters in the first part of the experiment, and all denied having knowledge of the first part of the experiment. There were ten female Ss and twenty-six male Ss. The data for one male S assigned to the Moral-Social Responsibility experimental group was discarded on the grounds that he failed to follow the experimental instructions. It appeared to E that the S became distracted from the task during the process of learning the second set, which he did indeed fail to master. Inquiry after the session was completed, revealed that S had by his own admission "lost interest" in the problem at about that point, and made no further attempt to learn. This was taken to be adequate evidence that he had deliberately ignored the experimental instructions requiring S to make an effort to learn to classify the stimuli. For the remaining thirty-five cases, seven Ss, five males and two females, were assigned to each of the four experimental conditions and the control condition.

CHAPTER III

RESULTS

The results of the main part of the experiment (learning and transfer) in terms of trials to criterion on the first and second sets for all 35 Ss is presented in Table II. Total errors before reaching the criterion appears in Table III. Table IV shows the number correct on the first trial of the second set for each S in every group, a standard measure in transfer of training experiments.

These tables show the major aspects of the data which have some bearing on the questions asked at the outset of this research.

Since "Trials to Criterion" and "Total Errors" may be regarded as alternative indices of performance, and since there is substantial correlation between them as is apparent in the data shown in Tables II and III. only the data in Table III was subjected to statistical analysis. The choice to use "Total Errors" instead of "Trials to Criterion" is based on the fact that the former provides a wider range of possible scores, and generally speaking, the greater the number of steps on psychological scale, the greater the reliability (Ferguson, 1941), and consequently the possibility of detecting the effects of the experimental factors, if there are effects is increased.

The features of the data in Table III which are especially worthy of attention are: The differences between the groups in scores on the first set; these scores reflect the relative difficulty of the problems, which in turn can be interpreted as a function of the faces

TABLE II

TRIALS TO CRITERION IN CONSECUTIVELY LEARNING TO SORT TWO SETS OF PHOTOGRAPHS ON THE BASIS OF A PERSONALITY TRAIT, OR A PHYSIOGNOMIC TRAIT, OR IN A CONTROL CONDITION. 10 INDICATES FAILURE TO REACH CRITERION IN TEN TRIALS, THE MAXIMUN NUMBER ALLOWED

Personality Traits				Physiognomic Traits				Control	
Warmth-Tolerance		Moral-Social Responsibility		Age		Complexion			
Set 1	Set 2	Set 1	Set 2	Set 1	Set 2	Set 1	Set 2	Set 1	Set 2
				Males					
1	0	9	2	2	0	1	3	10+	10+
1	2	3	1	3	0	6	5	2	5
1	3	1	1	4	1	10+	2	9	10+
5	1	1	2	0	0	3	2	6	3
4	2	3	2	2	1	8	2	10+	7
				Females					
1	0	3	2	1	1	7	2	10+	6
1	0	5	2	2	2	9	3	10+	10+
				Group Means					
2.1	1.1	3.6	1.7	2.0	0.7	6.3	2.7	8.1	7.3

TABLE III

TOTAL ERRORS BEFORE REACHING CRITERION IN LEARNING TO SORT TWO SETS OF PHOTOGRAPHS ON THE BASIS OF A PERSONALITY TRAIT, OR A PHYSIOGNOMIC TRAIT, OR IN A CONTROL CONDITION. + INDICATES SUBJECT FAILED TO REACH CRITERION IN TEN TRAILS, THE MAXIMUN NUMBER ALLOWED

Personality Traits				Physiognomic Traits				Control	
Warmth-Tolerance		Moral-Social Responsibility		Age		Complexion			
Set 1	Set 2	Set 1	Set 2	Set 1	Set 2	Set 1	Set 2	Set 1	Set 2
				Males					
1	0	25	8	4	0	2	2	20+	18+
1	1	11	3	4	0	16	11	5	16
1	3	4	1	16	2	17+	7	26	27+
8	1	3	3	0	0	12	5	11	7
6	1	4	3	3	2	19	6	30+	18
				Females					
1	0	9	6	4	1	15	8	25+	18
1	0	19	4	3	3	19	6	30+	22+
				Group Means					
2.7	0.9	10.7	4.0	4.9	1.1	14.3	6.4	21.0	18.0

NUMBER CORRECT ON THE FIRST SORTING OF THE SECOND SET FOR MALE AND FEMALE SUBJECTS ON DIFFERENT SORTING TASKS BASED ON PERSONALITY TRAITS, PHYSIOGNOMIC TRAITS, AND A CONTROL CONDITION. TEN IS THE MAXIMUM POSSIBLE SCORE

Personality Traits		Physiognomic Traits		Control
Warmth-Tolerance	Moral-Social Responsibility	Age	Complexion	
		Males		
10	5	10	9	6
10	7	10	6	6
9	8	10	6	7
10	9	8	6	6
10	8	8	5	6
		Females		
10	5	9	4	5
10	7	8	7	6
		Group Means		
9.9	7.0	9.1	6.1	6.0

being distinguished on the designated trait. Secondly the difference stands out between the scores for each S on Set 1 and Set 2; this difference represents the amount of "transfer," or the extent to which learning on the first task generalized to the second.

The scores on Set 1 in Table III reveal an apparent difference in the difficulty of the problems for the various conditions. To test this difference statistically the scores from Set 1 were analyzed separately. This portion of the data was treated as if it were an independent experiment.

A technique described by Walker & Lev (1953, p. 381) for use with data produced by this type of design was employed. It is an approximation to the exact analysis of variance and it is useful when there are unequal numbers in the sub-groups. In the present case there were unequal subgroups due to the number of male Ss exceeding the number of female Ss. Use of this technique permits testing the difference in performance in the various tasks, the differences in performance by the sexes across tasks, and the possible differences between the sexes in performing certain tasks.

An extension of the analysis of variance described by Senders (1958) and Walker and Lev (1953) permits making separate comparisons among the various groups. These comparisons will be explained in more detail in the interpretation of the analysis which follows.

The results of the analysis are presented in Table V. The F ratios obtained indicate that there was no significant difference in the overall performance of the sexes. Nor was there any evidence

of difference in performance that could be attributed to the combination of subject's sex and task. But the hypothesis that there was no difference between performances on the various tasks can be rejected with a high degree of confidence.

Partitioning the sum of squares due to "Tasks" enabled making the following orthogonal comparisons, which are also shown in Table V. The first compares the performance in the control condition against that in all other conditions combined. This comparison pits the performance in the task where no distinguishing trait was available against performance on tasks where there was some trait on which the faces can distinguished. The difference in this case was found to be highly significant, P = .01. The difference was in the direction of more rapid learning by the groups for whom a distinguishing trait was available. The size of the difference, in terms of errors, is indicated by the mean score of the experimental groups, 8.14, compared to the mean score of the control groups, 21.00.

The second analysis compares performance on tasks for which the distinguishing feature of the faces is a personality trait with the performance on tasks for which the distinguishing feature is a physiognomic trait. There is practically no evidence of a difference between these types of tasks.

The two final comparisons, under tasks, each tests the difference between the performances on the individual tasks within one of the general types of tasks. The performances of the two groups both of which learned a classification based on personality traits were compared, and the performances of the two groups whose tasks were based on phsiognomic traits were compared. In both cases the difference in

TABLE V

RESULTS OF AN APPROXIMATE ANALYSIS OF VARIANCE PERFORMED ON THE PORTION OF THE DATA FROM TABLE 3 REPRESENTING SET 1. * AND ** DENOTE VALUES SIGNIFICANT AT OR BEYOND THE .05 AND .01 LEVELS RESPECTIVELY

Source	df	MS	F
Tasks	(4)	(140.03)	(8.17**)
Experimental vs. Control	1	340.47	19.86**
Personality vs. Physiognomic	1	15.96	.93
Warmth-Tolerance vs. Moral-Social Responsibility	1	90.25	5.26*
Age vs. Complexion	1	113.42	6.62*
Sexes	1	17.43	1.02
Sex X Tasks Interaction	4	11.62	.68
Error	25	17.14	
Total	34		

performance on the tasks, which were based on the same type of trait, proved to be highly significant; P = .01 in each case. Learning to separate the photographs on the basis of Warmth-Tolerance was found to be less difficult than learning the corresponding classification based on Moral-Social Responsibility. The mean error score for the Warmth-Tolerance group was 2.71; for the Moral-Social-Responsibility group it was 10.71. Similarly, learning to respond to the faces on the basis of Age proved to be more easily accomplished than learning to respond to the faces on the basis of their Complexion. For the Age group the mean error score was 4.86, for the Complexion group it was 14.29

The four sub-comparisons under tasks are orthogonal, and completely account for the Variance due to tasks.

The fundamental assumptions required for the analysis of variance are somewhat tenuous in the present case. Tests for homogeneity of variance do not appear very appropriate when some of the sub-groups consist of only two scores. Sub-groups of seven scores each were obtained by ignoring sex as a factor, and the Hartley F-maximum test applied to the estimates of variance based on these sub-groups of seven each failed to reject the homogeneity assumption at the .05 level. There were clearly not enough scores to permit testing the assumption of a normal distribution in the underlying population.

In summary, the inferences suggested by the analysis of the results obtained on Set 1 (the original learning of the different tasks) are these: Learning to respond differentially to pictures of faces is facilitated when the faces are distinguished on the basis of

either judged personality traits, or judged physiognomic traits; but neither general type of trait appears to function more effectively as a cue than the other. Instead, there were found to be wide differences in the effectiveness of the cues provided by different from within either category. One personality trait proved to be an effective cue in learning, another did not. Similarly one of the physiognomic traits proved to be a highly effective cue, while the other did not.

There was no evidence that the different traits function more effectively as cues for one sex than the other.

However the foregoing conclusions attribute a somewhate greater degree of generality to the present findings than can be justified; The statements should be restricted to adults in the college population.

In order to test the significance of the transfer effects from Set 1 to Set 2 which are suggested by the scores in Table III a second analysis of variance was performed which took into account all of the scores in the table. However, no technique of analysis was available which would compensate for the unequal size of the subgroups in this more complex design; there appears to be no technique corresponding to the Walker and Lev method employed on the less complex data. In view of this and the fact that the number of Ss of each sex is balanced across the tasks, and furthermore since the previous analysis had failed to show any difference due to sexes, the sex variable was ignored as a factor in conducting the second analysis. When sex is not taken into account, the subgroups are all equal in size.

A preliminary inspection of the total distribution of scores

revealed a pronounced skewness and heterogeneity of variance, and in order to rectify this situation a log standard transformation was made. A constant of one was added to all scores in order to eliminate the

$$X' = \log_{10} (X + 1)$$

zero scores, for which there would be no logarithm. The transformed scores exhibited a distribution similar to the normal, and the Hartley F-max test indicated the homogeneity of variance assumption could be retained, P = .05.

A table of the transformed scores appears in Appendix E.

The analysis of variance on the transformed scores was performed according to the method described by Edwards (1950) under the title of "Repeated Measurements Design." The results of the analysis appear in Table VI. On the basis of the analysis, several statements can be made. There is convincing evidence that the overall performance differs on the various tasks; the F-ratio yielded by this comparison is significant at well beyond the .01 level. This is a confirmation of the finding in the earlier analysis, but it also indicates that the difference in difficulty found previously on Set 1 continues for the tasks in Set 2. Because the trend of the data is apparently the same in both Sets- - the relative standing of the tasks is the same for Set 2 as it was on Set 1- - the comparisons between tasks which were made in the previous analysis were not repeated here.

The highly significant F-ratio produced by the mean square associated with Sets shown in Table VI, warrants the statement that there was transfer from Set 1 to Set 2 in all types of tasks. Referring

TABLE VI

RESULTS OF ANALYSIS OF VARIANCE PERFORMED ON LOG-TRANSFORMATION OF SCORES IN TABLE III ** DENOTES VALUE SIGNIFICANT AT OR BEYOND THE .01 LEVEL

Source	df	MS	F
(Subjects, Ss)	(34)		
Tasks (T)	4	2.0265	19.30**
Ss within Tasks, 1st error term	30	.1050	2.99**
Sets	1	1.1290	321.65**
(Interaction, Ss X Sets)	(34)		
Tasks X Sets	4	.0610	1.74
Ss within Tasks X Sets, 2nd error term	30	.0351	
Total	69		

back to Table III it can be seen that the overall transfer was postive. The significance of this transfer effect can be demonstrated in a more direct fashion by a simple non-parametric test. In twenty-eight of the thirty-five cases, the total errors decreased from Set 1 to Set 2, and comparing this to the null hypothesis that changes in either direction were equally likely and binomial distribution shows the result to be significant at the .005 level. Within the experimental groups postive transfer occured in twenty-three of the twenty-eight cases, which is statistically significant at the .005 level. Within the control group positive transfer occurred in five of the seven cases which fails to achieve statistical significance but nevertheless strongly suggests the presence of positive transfer. The existence of postive transfer in the control condition can be explained as due to broad factors resulting in improvement in performance on the second task. These broad factors include such things as becoming more at ease in the test situation, improved concentration, greater facility in handling the test materials, etc. For the experimental conditions there was, in addition to these broad factors, a narrow or specific factor which served as a basis of transfer, the trait which distinguished the two classes of faces.

Turning next to the F-ratio associated with the interaction of Tasks and Sets, the value obtained failed to achieve statistical significance. The implication in this instance is that the data do not provide sufficient evidence for a statement that there is more transfer for some kinds of tasks than others. The remaining F-ratio shown in

Table VI, labeled Ss within tasks seems relatively unimportant. It shows only that there are individual differences among Ss.

In summary, the more complete analysis served primarily to demonstrate that the overall transfer effect, which was apparent in the data, was statistically significant. It also revealed a statistically significant difference in the scores on the different tasks, confirming one result of the previous analysis and extending it to apply to the data for Set 2, and although a significant transfer effect was revealed, it was not shown that the transfer effect differed significantly among the tasks.

Because the data in Table III do strongly suggest the existence of differential transfer effects among the groups, further analysis was made on this aspect of the data A "savings score" was adopted as a suitable index of transfer. This savings score is a function of the two error scores available for each S in Table III. Using the pairs of scores for each S the transfer indices were computed according to the formula:

$$\frac{TE_1 - TE_2}{TE_1} \times 100$$

where TE_1 = total errors on Set 1
and TE_2 = total errors on Set 2

A special case was made for the fourth S in the Age group for whom both error scores were zero. Transfer in this case was regarded as 100.0 since there were no errors in learning Set 2, and the absence of errors on Set 1 was known from S's verbal report to have resulted from an accurate "guess" concerning the principal involved. S stated

that he had noted the discrepency in ages in the first two photographs he viewed, and suspected that this was the basis of classification. Responding to this attribute of the faces, he was also fortunate in assigning the very first photograph to the correct letter-label. It may be recalled that the categories had been designated simply A and B; and even though the principle was known, the category had to be matched to the proper letter. The latter step could only be accomplished by trial and error. There was no logical relationship between the letter label and the category.

The transfer indices produced by this operation are shown in Table VII. Heterogeneity of variance and non-normal distributions prohibited the use of the analysis of variance technique with these data. But it was possible to construct two by two contingency tables which compared each experimental group to the control group, separately, on the basis of the number of scores in each group above and below the grand median of the entire distribution. The Fisher-Yates Exact Test applied to each of the contingency tables representing the four comparisons indicated that the hypothesis of no difference in transfer effects between the experimental group and control group could be rejected at better than the .05 level of significance for three of the experimental groups: Warmth-Tolerence, $P = .010$; Moral-Social Responsibility, $P = .035$; Age, $P = .010$. In the case of the Complexion group the data approached the .05 significance level, the actual P value being .066.

No further comparisons among the groups on the basis of the transfer indices appeared warranted. It was quite obvious that for the two groups representing each type of experimental tasks, Personality Traits and Physiognomic Traits, transfer is high in one case and low in the other.

The results of the generalization analysis presents much the same order among the tasks as was found by the analysis concerning relative difficulty of the tasks. Transfer was found to be greater in every case where the faces could be distinguished on the basis of a "perceived" personality trait or a perceived physiognomic trait than it was in the condition where no trait was available. However, neither general type of trait was shown to serve more effectively as a basis of transfer than the other. Instead there was found to be considerable difference between traits of the same general type.

The data presented in Table IV p. 41 the number correct on the first sorting of Set 2, is also a measure of transfer. Unfortunately, in the present case, it is not possible to give an exact interpretation of these data. Due to the fact that S was informed as to whether his choices were right or wrong during the early part of the first sorting of Set 2, there is no absolute assurance that S achieved a high score on the first sorting of Set 2 on the basis of transfer from the previous training alone. He might also simply have learned a new principle early in the first sorting. Interpretation of these data would have been somewhat easier had the first sorting of Set 2 been carried out as a extinction procedure, making it more nearly similar to the conventional

TABLE VII

TRANSFER OF TRAINING INDICES FOR MALE AND FEMALE SUBJECTS ON DIFFERENT SORTING TASKS BASED ON PERSONALITY TRAITS, PHYSIOGNOMIC TRAITS, AND A CONTROL CONDITION

Personality Traits		Physiognomic Traits		Control
Warmth-Tolerance	Moral-Social Responsibility	Age	Complexion	
		Males		
100.0	68.0	100.0	0.0	10.0
0.0	72.7	100.0	31.2	-220.0
200.0	75.0	87.5	58.8	-3.8
87.5	0.0	100.0[a]	58.3	36.4
83.3	25.0	33.3	46.1	40.0
		Females		
100.0	33.0	75.0	46.7	28.0
100.0	78.9	0.0	68.4	26.8
		Group Medians		
87.5	68.0	87.5	46.7	26.8

[a]Special case, explained in text.

paradigm for measuring stimulus generalization. This matter is dicussed in detail by Brogden (1951).

With the qualification imposed by the foregoing consideration, the data in Table IV exhibits some striking features. The overall number correct for a group of seven Ss, on the basis of chance, would be 38.5 or 5.5 for each individual. The overall number correct in every group is above the chance-expectancy. For two of the groups, the Control group and the Complexion group, the deviation from chance is quite small. For the Moral Social Responsibility group it is moderate, and for the Warmth-Tolerance and Age groups it is marked.

Testing the hypothesis that the obtained proportions correct might be the result of sampling error, that is, that they are only chance deviations from the hypothetical proportions correct of .55, yielded the following results: Warmth-Tolerance, P = .001; Moral-Social Responsibility, P = .012; Age, P = .001; Complexion, P =.312; Controls, P = .401.

An analysis of variance performed on the scores in Table IV showed the difference among the groups to be significant at well beyond the .01 level. And application of the Tukey Gap Test (Tukey, 1949) as described by Federer (1955) identified two distinct groups the first consisting of the Age and Warmth-Tolerance groups, and the other containing the Moral Social Responsibility, Complexion and Control groups.

The conclusion suggested by the data and analysis, confirms the results of the analysis made on the transfer indices. Significant transfer occurred on the basis of one of the percieved personality traits and one of the perceived physiognomic traits, but did not occur for the

other traits representing these classes. In the present analysis transfer in the experimental groups showing low transfer was not shown to be significantly different from the control group.

One additional datum was collected in the experiment. This was the Ss' verbalizations of the principle involved in sorting the photographs. For the purposes of analysis only the final verbalization made during the learning of Set 1 was taken into consideration. Verbalizations were available only for the experimental groups, and of course the Ss' statements were seldom identical with experimenter's description of the principle. In fact it was quite apparent that each statement obtained could well be regarded as falling somewhere along a continuum extending from absolutely incorrect to perfectly correct according to the degree to which it was equivalent to the principle stated by the experimenter.

To illustrate what is meant here, a S in the Complexion group might have described the principle as "swarthy persons versus fair persons" and this would certainly be equivalent to E's wording of "dark complexioned persons versus light complexioned persons." But compare this to another S's statement that "It's something about their skin." Obviously the last statement is not quite so accurate as the first, and yet it is more nearly correct than "no statement," which, in turn, is not so incorrect as the statement "persons with hair parted on left side versus persons with hair parted on right side."

Since it did seem reasonable to conceive of the statement as

varying along the dimension of correctness, a method was devised for assigning each statement a value on this dimension. Each statement was rated on a five point scale according to how accurately it described the principle involved in the experimental condition. The scoring system employed was as follows:

5 = Identical, or equivalent.
4 = Close.
3 = Similar, or related.
2 = (no statement).
1 = Different or incorrect.

Submitting the statements to three independent judges (including the writer) for scoring revealed an average intercorrelation among raters of .87, indicating that the scores obtained are quite reliable. The scores are presented in Table VIII, from which it can be seen that the scores are closely related to the scores presented in Table IV, the number correct on the first sorting of Set 2. The Pearson product-moment correlation coefficient between the two sets of data is .842, which is excellent agreement for 28 pairs of scores and is significant at well beyond the .001 level.

These data support previous results. It can be said Ss were able to verbalize the principle involved quite accurately when the faces were distinguished on the Warmth-Tolerance trait or the Age trait. This repeated the pattern already reported in each of the previous analyses. It may be added that the number correct on the first sorting of Set 2 which constituted a score in Table IV appears to be measuring very much the same thing as the Verbalization Scores. This was indicated by the substantial correlation between the two

TABLE VIII

SCORES INDICATING SUBJECT'S ACCURACY IN VERBALIZING THE PRINCIPLE INVOLVED IN PERFORMING THE TASK, FOLLOWING INITIAL LEARNING OF A SORTING TASK. SCALE IS EXPLAINED IN TEXT

Personality Traits		Physiognomic Traits	
Warmth-Tolerance	Moral-Social Responsibility	Age	Complexion
	Males		
4	3	3	3
4	2	5	2
5	5	5	1
4	2	5	1
5	1	5	1
	Females		
5	4	5	2
4	3	5	3
	Group Medians		
4	3	5	2

sets of scores. This finding tends to support an assumption made in the design of the experiment: that in many instances inductive concept formation and discrimination learning are basically the same process.

CHAPTER IV

DISCUSION OF THE RESULTS

Summary of the Findings

On the basis of the evidence obtained in this research the following answers can be made to the questions asked at the outset.

1. Can persons learn to respond to faces on the basis of some physiognomic trait which characterizes a group of faces?

The answer appears to be yes. It was demonstrated that when faces were distinguished on the basis of a physiognomic trait, Age, learning to respond to the faces on the basis of the cue provided by this trait was facilitated compared to learning where no trait was available. It was also demonstrated that the learned response transferred to a second set of faces which were distinguished on the trait. The evidence suggested that a second physiognomic trait, Complexion, also functioned as a cue in learning, but not nearly as effectively as Age.

2. Can persons learn to respond to faces on the basis of some personality trait which judges agree characterizes a group of faces?

The answer appears to be yes; but here as in the answer to the previous question, one of the personality traits employed in the study proved much more effective as a cue than did the other.

3. Is learning to respond to one kind of trait in faces more readily accomplished than learning to respond to the other kind of trait?

The answer to this question seems to be no. At least the evidence obtained in this study provides no basis for saying that there is a difference. Instead it was found that traits of the same kind varied widely in their effectiveness as cues.

4. Can the Ss verbalize accurately concerning the facial traits to which they learn to respond?

The answer is yes for some traits: Here too the major differences found occurred between traits of the same kind. Ss were found able to verbalize quite accurately concerning one of the personality traits and one of thephysiognomic traits, but not for the others.

5. Do the sexes differ in the ability to do any of the things described in questions one through four?

The evidence from the present research did not show any difference between the sexes in their performance on those tasks.

Some Implications for Theory and Recommendations for Future Research

The present research was conducted from an S-R behavioristic frame of reference; However it was designed to produce information which would have relevance for social psychology and clinical psychology, as well as eventual importance for general psychological theory.

The relevance for social psychology resides in the fact, demonstrated by the research, that a response one has learned to make to a particular person may also be made in response to a new person who is similar to the first with respect to a perceived personality trait or a perceived physiognomic trait. At the outset of this research it was suspected that one of these general kinds of attributes of persons might take precedence over the other as a cue, but this was not shown to be the case. In view of this feature of the results, no general principle can be provided to the social psychologists for predicting along which particular dimension behavior will transfer. However it has served to show that both kinds of attributes are available as bases of transfer, and this is perhaps equally important information. The knowledge that these kinds of unifying concepts exist for social stimuli may possibly enable the social psychologists to find regularities in their data that might otherwise have escaped notice.

The writer makes no claim for the originality of the suggestion that such unifying concepts, perceived personality traits and physiognomic traits, exist. What the present research does provide is a firmer factual foundation to justify using these concepts if they are found useful in explaining social behavior, particularly learned social behavior. The social psychologist may employ these factors with more confidence than he could have in the absence of impirical evidence.

A similar argument can be presented for the significance of

the research to clinical psychology. To the degree that psychotherapy is conceived of as a learning process, and particularly if it is regarded as a process of social learning, the argument would be identical to that provided in the preceding paragraphs. The possible relationship between psychotherapy and social learning has been mentioned in some detail in the introduction to this report, and the discussion will not be repeated at this point. But in so far as clinical psychology emphasises the learning of "emotional responses" the findings of the present research would have to be used cautiously. Emotional responses are generally conceded to be involuntary responses, associated primarily with activities of smooth muscle and glands. The present research has been limited to the investigation of the effectiveness of different ones in connection with a voluntary response, and there is as yet no assurance that the laws of learning are the same for the voluntary and involuntary systems. The latter question provides an interesting lead for possible future research; the experiment could be repeated using an involuntary response to discover if the relationships found here hold for a wider range of behavior. Another consideration which limits the significance of the findings of the present study for psychotherapy is due to the static nature of the stimuli used in the experiment. As was pointed out in the introduction of the present report, psychotherapy is concerned with behavior made in response to the actions of other people, perhaps especially responses to the verbal activity of other people, and not with responses made to the appearance of other people with action frozen as it is in still photographs. In this regard, the

present study was intentionally removed from the dynamic situation existing in real life in order to achieve better control of the stimuli, while recognizing that a certain amount of realism was lost some degree of artificiality introduced. Complete ecological validity simply could not be attained with the available facilities, and some compromise was necessary. It might also be pointed out that psychotherapy deals with responses made to whole persons and not to faces alone, even though the face may be the most significant part of persons as stimulus objects. Here, too, it must be admitted that the present research was somewhat artificial.

The foregoing limitations are not the only restirctions which must be placed on the data provided by this research. It must also be remembered that it utilized pictures of adult male faces instead of actual human faces in general and it was performed with only a select portion of the general population as subjects. Both of these limitations suggest possible directions for future research which might provide information of greater generality.

The possible implications of this research for general psychological theory are not nearly so obvious as its significance for more specialized branches. However, attention is drawn to the fact that the two traits that were found to serve most effectively as cues, Age and Warmth-Tolerance, are both attributes which must be inferred from the physical features of the face. (Despite the face that Age was listed as a physiognomic trait.) It was not anticipated that such higher-order unifying factors would prove to be more readily responded

to than a more concrete feature like Complexion. The distinction suggested here between concrete and high-order factors is that in the latter the thing which unites the faces within a category appears to be labeling response which persons in our culture learn to make to certain kinds of faces. It is conceivable that there may be several different kinds of faces in such a category, alike only in that all evoke the same labeling response, e.g. "young." Miller and Dollard (1941) have explained in some detail how new learning might generalize to a series of stimuli by means of a previously learned labeling response. They speak of the process as the "acquired equivalence of cues" which they describe as follows. A common response becomes connected to each individual stimulus, and once such a response is acquired through a number of learning situations, cues produced by this common response can serve as the common stimuli necessary for generalization.

A possible implication of the present research is that it indicates the myriad ways in which culturally determined labeling responses can be involved in shaping human behavior. Even so the research findings will not support an unqualified statement that higher-order factors always provide a more effective cue for generalization than do concrete features. For despite the fact that both of the cues found to be most effective in this research were of the higher-order variety, one of the cues found to be least effective, Moral-Socail Responsibility, is also of this type. In view of the equivocal nature of the findings on this point, the writer prefers to dismiss the implications for theory

with this statement. It appears that a fully satisfactory theory of human behavior will have to include some provision to allow for the operation of such higher-order factors among stimulus objects, but the nature of these higher-order factors, including their particular effectiveness as cues, will probably remain a matter to be determined by empirical investigation (of which this research provides an example).

CHAPTER V

SUMMARY

An experiment in human learning was conducted in which subjects learned to sort black and white photographs of adult male faces. The effectiveness of two different kinds of attributes of the faces as cues for the sorting task was investigated. The kinds of attributes were personality traits and physiognomic traits. Previous research by other investigators had demonstrated that college students agree in attributing these kinds of traits to photographs of faces.

In the present study photographs were rated on a number of personality traits and physiognomic traits by a group of college students. Then, on the basis of these ratings, sets of photographs were selected that were dichotomous with respect to a bipolar personality trait or a bipolar physiognomic trait. Each of these two classes was represented by two traits employed as cues in separate sorting tasks which constituted the major part of the experiment. The personality traits employed were designated Warmth-Tolerance and Moral-Social Responsibility; The physiognomic traits were Age and Complexion. In addition there was a control condition in which the subjects learned to sort a like number of photographs into an artificial dichotomy. Thus there were five conditions and a different group of subjects was employed in each, forming an "independent-groups" type of experimental design.

In every case experimental subjects were required to learn to sort correctly first one set of photographs and then a second set,

both of which were dichotomized on the basis of the same trait. The control subjects merely learned to sort two consecutive sets of photographs in which there was no cue or principle available to distinguish between the two categories. This procedure produced measures of cue effectiveness in both original learning and in transfer of training.

Thirty-five college students served as subjects in the learning and transfer experiment. None of these were persons who participated in the preliminary part of the experiment in which the photographs were rated. The subjects were assigned to the various conditions so that each group contained five males and two females. The resulting factorial design permitted testing for possible differences in performance that might be attributed to sex of the subjects.

The results showed that college students can learn to respond to personality traits and to physiognomic traits in photographs of faces. Learning was facilitated significantly in all experimental conditions where traits were available as cues in contrast to the control condition where traits were not available as cues. However, comparisons among the experimental groups produced no evidence that one type of trait serves any more effectively as a cue than the other type. Instead it was found that traits of the same type differed widely in their effectiveness as cues.

Results on the transfer tasks were consistent with the findings in the original learning. Postive transfer occurred in all conditions, but, in general, there was no evidence of greater transfer for tasks based on physiognomic traits. There was strong evidence of greater

transfer in every case where traits were available as cues than in the control condition where no trait was available. But there were also sizeable differences in the amount of transfer associated with traits of the same kind.

The subjects' ability to verbalize accurately concerning the principle involved in sorting the photographs was found to correlate highly with the amount of transfer which occured. Again the results presented the pattern found before, suggesting no consistent difference in the cue value of the two types of traits for the verbal process either, but sizeable differences between traits of the same type.

The experiment produced no evidence that the sexes differ in performance on any or all of these tasks.

REFERENCES

REFERENCES

1. Berg, E. A. A simple objective technique for measuring flexibility in thinking. J. gen. Psychol., 1948, 39, 15-22.
2. Blake, R. R., & Ramsey, G. V. Perception: an approach to personality. New York: Roland Press, 1951
3. Bruner, J. S., & Krech, D. (Eds.) Perception and personality. Durham: Duke Univ. Press, 1950.
4. Bruner, J. S., & Tagiuri, R. The perception of people. In: G. Lindzey (Ed.) Handbook of social psychology. vol. II. Cambridge: Addison-Wesley, 1954.
5. Bueno, L. F. Generalization and extinction in an experimental therapeutic situation. Unpublished doctoral dissertation, Univ. of Tenn., 1955.
6. Brogden, W. J. Animal studies of learning. In: S. S. Stevens (Ed.) Handbook of experimental psychology. New York: Wiley and Sons, 1951. Ch. 16, 568-612.
7. Cameron, N. The psychology of the behavior disorders. Boston: Houghton Mifflin, 1947.
8. Davidon, R. S. The effect of symbols, shifts, and manipulation upon the number of concepts attained. J. exp. Psychol., 1952, 44, 70-79.
9. Dollard, J., & Miller, N. E. Personality and psychotherapy. New York: McGraw-Hill, 1950.
10. Edwards, A. L. Experimental design in psychological research. New York: Rinehart, 1950.
11. Federer, W. T. Experimental design. New York: MacMillan, 1955.

12. Ferguson, L. W. A study of the Likert technique of attitude scale construction. J. soc. Psychol., 1941, 13, 51-57.

13. Freud, S. Psychoanalysis. In: Encyclopedia britannica. vol. 18. Chicago, London, Toronta: Encyclopedia Britannica Inc., 1954, 668.

14. Green, E. J. Concept formation: A problem in human operant conditioning. J. exp. Psychol., 1955, 49, 175-180.

15. Guilford, J. P. Psychometric methods. (2nd ed.) New York: McGraw-Hill, 1954.

16. Hilgard, E. R., & Marquis, D. G. Conditioning and learning. New York: Appleton-Century, 1940.

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

18. Krech, D., & Crutchfield, R. S. Theory and problems of social psychology. New York: McGraw-Hill, 1948.

19. Leeper, R. Cognitive processes. In: S. S. Stevens (Ed.) Handbook of experimental psychology. New York: Wiley and Sons, 1951. Ch. 19, 730-757.

20. Lewin, K. Frontiers in group dynamics: I. Concept, method, and reality in social science, social equilibria and social change. Hum. Relat., 1947, 1, 5-41.

21. Macleod, R. S. The phenomenological approach to social psychology. Psychol. Rev., 1947, 54. 193-210.

22. Margaret, Ann. Generalization in successful therapy. J. consult. psychol., 1950, 14, 64-70.

23. Miller, N. E., & Dollard, J. Social learning and imitation. New Haven: Yale Univ. Press, 1941.

24. Munroe, Ruth L. Schools of psychoanalytic thought. New York: Dryden Press, 1955.

25. Newcomb, T. M. Social psychology. New York: Dryden Press, 1950.

26. Osgood, C. E. Method and theory in experimental psychology. New York: Oxford Univ. Press, 1953.

27. Reed, H. B. II. The influence of the length of series. III. The origin of concepts. J. exp. Psychol., 1946, 36, 167-179.

28. Rotter, J. B. Social learning and clinical psychology. New York: Prentice Hall, 1954.

29. Secord, P. F., Dukes, W. F., & Bevan, W. Personalities in faces: An experiment in social perceiving. Genet. Psychol. Monogr., 1954, 49, 231-279.

30. Secord, P. F., & Muthard, J. E. Individual differences in the perception of women's faces. J. abnorm. soc. Psychol., 1955, 50, 238-242.

31. Senders, Virginia L. Measurement and statistics. New York: Oxford Univ. Press, 1958.

32. Shoban, J. E., Jr. Psychotherapy as a problem in learning theory. Psychol. Bull., 1949, 46, 366-392.

33. Stritch, T. M., & Secord, P. F. Interaction effects in the perception of faces. J. Pers., 1956, 24, 272-284

34. Sullivan, H. S. The meaning of anxiety in psychiatry and in everyday life. Psychiatry, 1948, 11, 1-13.

35. Time, the Weekly News magazine. New York: Time, Inc.

36. Tukey, J. W. Comparing individual means in the analysis of variance. Biometrica, 1949, 5, 99-114.

37. Vinacke, W. E. The investigation of concept formation. Psychol. Bull., 1951, 48, 1-31.

38. Walker, Helen M., & Lev, J. Statistical inference. New York: Holt, 1953.

39. Wallach, M. A. On psychological similarity. Psychol. Rev., 1952, 65, 103-116.

40. Wells, W. S. Galvanic skin response as a function of the experimenter and manifest anxiety. Unpublished doctoral dissertation, Univ. of Tenn., 1956.

41. Woodworth, R. S., & Schlossberg, H. Experimental psychology. New York: Holt and Co., 1954.

APPENDIXES

APPENDIX A

Replica of Rating Sheet for Photographs

Judge _____ Photo No. _____ Date _____

Rate each picture on each of the traits listed below. Indicate your ratings by encircling the apprppriate number from 1 to 7. In general the number 1 will mean that the face shows very little of the trait and the number 7 will mean that the face shows a lot of the trait. The number 4 will indicate an amount half-way between. For some particular traits additional information is provided to guide you in making the ratings.

Cheerful Appearance	1	2	3	4	5	6	7	
Sense of Humor	1	2	3	4	5	6	7	
Likable	1	2	3	4	5	6	7	
Honest Face	1	2	3	4	5	6	7	
Kind Face	1	2	3	4	5	6	7	
Warmhearted	1	2	3	4	5	6	7	
Intelligent Look	1	2	3	4	5	6	7	
Concientious	1	2	3	4	5	6	7	
Air of Responsibility	1	2	3	4	5	6	7	
Air of Refinement	1	2	3	4	5	6	7	
Distinguished Look	1	2	3	4	5	6	7	
Self Confident	1	2	3	4	5	6	7	
Alert Expression	1	2	3	4	5	6	7	
Determined Look	1	2	3	4	5	6	7	
Energetic Look	1	2	3	4	5	6	7	
Aggressive Look	1	2	3	4	5	6	7	
Proud	1	2	3	4	5	6	7	
Reserve	1	2	3	4	5	6	7	
Age (young face)	1	2	3	4	5	6	7	(old face)
Complexion (light)	1	2	3	4	5	6	7	(dark)
Fullness of Lips (narrow, thin)	1	2	3	4	5	6	7	(thick, full)
Eyebrows (light)	1	2	3	4	5	6	7	(heavy)

APPENDIX B

Replica of Rating Sheet for Photographs

Number of Judges _____ Date _____

(CLUSTER I: WARMTH, TOLERANCE)

Cheerful Appearance	Sense of Humor	Likeable	Honest Face	Kind Face	Warm-hearted
7	7	7	7	7	7
6	6	6	6	6	6
5	5	5	5	5	5
4	4	4	4	4	4
3	3	3	3	3	3
2	2	2	2	2	2
1	1	1	1	1	1

(CLUSTER II: MORAL-SOCIAL RESPONSIBILITY)

Intelligent Look	Conscientious	Air of Responsibility	Air of Refinement	Distinguished Look
7	7	7	7	7
6	6	6	6	6
5	5	5	5	5
4	4	4	4	4
3	3	3	3	3
2	2	2	2	2
1	1	1	1	1

(CLUSTER III: FORCEFULNESS)

Self-Confident	Alert Expression	Determined Look	Energetic	Aggressive Look
7	7	7	7	7
6	6	6	6	6
5	5	5	5	5
4	4	4	4	4
3	3	3	3	3
2	2	2	2	2
1	1	1	1	1

(CLUSTER IV: ALOOFNESS)		(PHYSIOGNOMIC TRAITS)			
Proud	Reserved	Age	Complexion	Fullness of Lips	Eyebrows
7	7	7	7	7	7
6	6	6	6	6	6
5	5	5	5	5	5
4	4	4	4	4	4
3	3	3	3	3	3
2	2	2	2	2	2
1	1	1	1	1	1

APPENDIX C

Frequency Distributions of Ratings for Two Personality Trait Clusters

Rating	Frequency
Warmth-Tolerance	
7	246
6	565
5	824
4	869
3	779
2	534
1	383
Moral-Social Responsibility	
7	270
6	775
5	899
4	757
3	419
2	225
1	155

APPENDIX D

Standard Instructions for Experimental Subjects

This is an experiment in sorting photographs.

I am going to give you a set of ten photographs of men's faces. These are photographs which appear in recent issues of a newsmagazine, but they are not very well known persons so you are not likely to recognize any of them. However if you do recognize a face please tell me.

What I want you to do is sort these photographs into two sets of five each. We will call them A-s and B-s, and I have placed cards with these letters on them on the table to indicate where to place the photographs in each set.

You are to work in this manner. Place the stack of photographs, face up, in front of you on the table. Pick up the top photograph and place it on either the A or B stack as you choose. I will tell you whether your choice is right or wrong; if it is wrong please move the photograph to the correct stack.

There is a rule for separating the A-s from the B-s. There is something about the faces that distinguishes the A-faces from the B-faces. As we go along see if you can discover what it is.

After you have sorted the photographs I will mix them up and give them to you to sort again. We will continue to do this until you have sorted them twice in succession without making an error. In between trials while I am shuffling the pictures I will give you a

APPENDIX D (Continued)

chance to write out any hunches or ideas you have about how to separate the A-s from the B-s.

So you really have two tasks: You must learn to classify each photograph correctly, and you must also try to find out what it is that distinguishes the A-s from the B-s.

Try to learn these as quickly as possible.

Are there any questions? If not begin.

(After the criterion is reached and oppertunity to verbalize is given, present the second set of photographs and say:)

Now here is a second set of ten pictures that can be sorted into A-s and B-s on exactly the same basis as the first set. The very same factor that distinguished the A-s from the B-s in the first set distinguishes the A-s from the B-s in this set also. Please go through the procedure of sorting as before, and again you will do it until two consecutive correct sorts have been made.

Any questions? If not begin.

APPENDIX E

Log Transformations of Error Scores From Table III

Personality Traits				Physiognomic Traits				Control	
Warmth-Tolerance		Moral-Social Responsibility		Age		Complexion			
Set 1	Set 2	Set 1	Set 2	Set 1	Set 2	Set 1	Set 2	Set 1	Set 2
.30	.00	1.42	.95	.70	.00	.48	.48	1.32	1.28
.30	.30	1.08	.60	.70	.00	1.23	1.08	.78	1.23
.30	.60	.70	.30	1.23	.48	1.26	.90	1.43	1.45
.95	.30	.60	.60	.00	.00	1.11	.78	1.08	.90
.85	.30	.70	.60	.60	.48	1.30	.85	1.49	1.28
.30	.00	1.00	.85	.70	.30	1.20	.95	1.42	1.28
.30	.00	1.30	.70	.60	.60	1.30	.85	1.49	1.36