Personality Correlates of Illusory Correlation in Clinical Judgment

Robert J. Lueger

Loyola University Chicago

Recommended Citation
http://ecommons.luc.edu/luc_theses/2706

This work is licensed under a Creative Commons Attribution-Noncommercial-No Derivative Works 3.0 License.
Copyright © 1974 Robert J. Lueger
PERSONALITY CORRELATES OF
ILLUSORY CORRELATION IN CLINICAL JUDGMENT

by

Robert J. Lueger

A Thesis Submitted to the Faculty of the Graduate School of Loyola University of Chicago in Partial Fulfillment of the Requirements for the Degree of Master of Arts

February

1974
ACKNOWLEDGMENTS

The author wishes to express his gratitude to Dr. Thomas Petzel for his help and encouragement during the preparation of this thesis. A note of thanks is also due Dr. Ronald E. Walker for his suggestions, ideas, and helpful comments. Finally, the author would like to thank the 92 individuals who, as introductory psychology students, served as subjects for this study.
VITA

Robert J. Lueger was born on January 20, 1949, in Axtell, Kansas.

He attended grade and high schools in Seneca, Kansas, and was graduated from high school in Baileyville, Kansas in 1967. He received his Bachelor of Arts degree in Psychology from St. Benedict's College, Atchison, Kansas in 1971. In September, 1971 he began graduate studies in clinical psychology at Loyola University, Chicago, Illinois.

He served as a graduate research and teaching assistant until June, 1973. From June, 1972 to September, 1972, he served his clerkship at the Kansas Neurological Institute, Topeka, Kansas. He began his clinical psychology internship at Topeka State Hospital, Topeka, Kansas in September, 1973.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACKNOWLEDGMENTS</td>
<td>ii</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>iv</td>
</tr>
<tr>
<td>I. INTRODUCTION TO THE PROBLEM AND A REVIEW OF RELEVANT LITERATURE</td>
<td>1</td>
</tr>
<tr>
<td>Process of Clinical Inference</td>
<td>3</td>
</tr>
<tr>
<td>Systematic Error in Clinical Report</td>
<td>5</td>
</tr>
<tr>
<td>Personality Correlates of Interest</td>
<td>13</td>
</tr>
<tr>
<td>Choosing an Experimental Paradigm</td>
<td>18</td>
</tr>
<tr>
<td>II. METHOD</td>
<td>20</td>
</tr>
<tr>
<td>Subjects</td>
<td>20</td>
</tr>
<tr>
<td>Task Materials</td>
<td>20</td>
</tr>
<tr>
<td>Experiment I</td>
<td>21</td>
</tr>
<tr>
<td>Procedure</td>
<td>22</td>
</tr>
<tr>
<td>Experiment II</td>
<td>23</td>
</tr>
<tr>
<td>Procedure</td>
<td>24</td>
</tr>
<tr>
<td>Measures</td>
<td>25</td>
</tr>
<tr>
<td>III. RESULTS</td>
<td>27</td>
</tr>
<tr>
<td>Experiment I</td>
<td>27</td>
</tr>
<tr>
<td>Experiment II - Scale Comparisons</td>
<td>29</td>
</tr>
<tr>
<td>Experiment II - Personality Measures</td>
<td>33</td>
</tr>
<tr>
<td>IV. DISCUSSION</td>
<td>40</td>
</tr>
<tr>
<td>SUMMARY</td>
<td>46</td>
</tr>
<tr>
<td>REFERENCES</td>
<td>48</td>
</tr>
</tbody>
</table>
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Reliability Coefficients and Percentages of Popular and Total Responses Repeated on Both Testings</td>
<td>28</td>
</tr>
<tr>
<td>2. Percent of Subjects Reporting Each Characteristic According to Symptoms</td>
<td>30</td>
</tr>
<tr>
<td>3. Comparison of Percentages of Characteristics Reported By Chapmans' Clinicians and Students and Present Experimental Group</td>
<td>32</td>
</tr>
<tr>
<td>4. Means, Standard Deviations, and t-Scores of Selected Measures for Low and High Reporters of Illusory Populars.</td>
<td>34</td>
</tr>
<tr>
<td>5. Means and Standard Deviations of Selected Measures According to Number of Illusory Populars Reported</td>
<td>36</td>
</tr>
<tr>
<td>6. F-Values for the Difference Between Means of Selected Measures According to Number of Illusory Populars Reported</td>
<td>37</td>
</tr>
<tr>
<td>7. Mean Number of &quot;No Relationship&quot; Responses Reported for High and Low Scorers on Selected Personality Measures</td>
<td>39</td>
</tr>
</tbody>
</table>
CHAPTER I

Introduction to the Problem and a Review of Relevant Literature

Previous investigations of the process of clinical assessment have centered around the validity of the clinicians' instruments as well as the validity and reliability of the clinicians' own judgmental process (Dohrenwend & Dohrenwend, 1965; Gough, 1962; Hathaway, 1955; Little & Shneidman, 1959; Meehl, 1960). Inherent in these studies has been the controversy of actuarial versus clinical prediction (Gough, 1962; Meehl, 1960; Sawyer, 1966; Taft, 1959). In many of these studies, the performance of clinicians in judgmental situations has not exactly been a call for optimism. Quite demonstrative of this was Little and Shneidman's study in which they compared individual blind diagnoses of experienced clinicians from several tests of personality and found that the clinicians agreed only slightly above chance. Similarly, Sawyer summarized a number of previous studies in comparing clinical prediction with mechanical or actuarial prediction, and found that users of a combination of the two methods seemed to outperform users of the pure clinical method alone.

Also involved in the controversy over the validity of clinical judgment is the disconcertion of not a few researchers with clinicians -- and particularly psychodiagnosticians of projective tests -- who continue to report clinical observations and make clinical inferences which, according to the research evidence, are clearly erroneous. Chapman and Chapman (1967, 1969) have pointed out this phenomenon on two projective tests, the Draw-A-Person (DAP) and the Rorschach, as have Roeback (1968) and Swenson
(1968) more extensively on the DAP. The disconcertion of the researchers has led Meehl (1960) to comment,

Personally I find the cultural lag between what the published research shows and what clinicians persist in claiming to do with their favorite devices even more disheartening than the adverse evidence itself. (p. 26)

Advocates of clinical judgment, intuition, or the inferential process, have countered (with some justification) with the claim that the validity of the research evidence may be brought to question, and that the comparative techniques of the researchers take clinical inference out of context and apart from the assessment process. Moreover, practicing clinicians maintain that their observations and inferences are substantiated through consensual validation with the similar observations of other clinicians.

More recent investigations then have centered around the process of clinical inference. These studies have attempted to simulate the cognitive processes of the clinician as he makes his judgmental decisions. This research takes two related but somewhat different forms. First, those studies investigating the accumulation of information according to a linear additive or configural process (Anderson, 1972; Goldberg, 1968), and secondly, those studies which investigate factors influencing the clinician's selection and utilization of cues in diagnostic judgment (Chapman & Chapman, 1967, 1969; Hunt & Walker, 1971). It is this second, albeit ambiguous, distinction with which this research is concerned, all the while keeping an eye to the earlier presented argument about the validity of clinical inference.
Process of Clinical Inference

Thorne (1961) sees diagnosis as essentially a problem-solving process with the classic pattern being: 1) data collection, 2) data reduction, 3) identification of pathognomic cues, 4) classification and categorization coding, 5) etiological diagnosis, 6) pathological diagnosis, 7) differential diagnosis, 8) nosological diagnosis, and 9) prognostic diagnosis. Thorne has noted that of these, data reduction and identification of pathognomic cues are the two most critical steps, and it is with these two processes that this research is most concerned.

In the problem-solving step of data reduction, the judge utilizes both inductive and deductive reasoning. Inductively, he scans the test stimulus complex to gather cues and formulate an hypothesis. Deductively, having formulated an hypothesis about the personality of the subject, he attempts to find cues which verify or disconfirm this hypothesis.

In the problem-solving step of identifying the pathognomic cues, the judge must discriminate between those cues present which are indicative of a particular personality description and other irrelevant cues which, though present, are not indicative of that personality description.

Thorne has here made a point highly relevant to the validity problem and to the clinician-versus-researcher controversy. A tentative explanation for the lack of validity and reliability in clinical judgment may be the fact that clinical judges are not able to differentiate valid cues. This is, as will later be pointed out, the major theme of the Chapmans' studies, and it will also lie at the crux of the research design to be presented.
Hunt and Jones (1961) also spoke of the identification of appropriate cues under the heading of stimulus identification. They noted that the stimulus elements (cues) in a clinical judgment situation are largely unknown and ill-defined. This is immediately obvious in a brief glance at the research literature. Many of the hypotheses forwarded by projective test constructors have remained unexplored, and on quite a number of the hypotheses which have been researched, there is evidence both substantiating and disconfirming the importance of a particular sign as a pathognomonic cue. The clinician then, in making his judgment, must deal with global and complex stimulus situations, and he must identify the specific part stimuli which will provide him with the essential cues for making a judgment.

Sarbin, Taft, and Bailey (1960) differentiated four stages in the cognitive activity of a clinician attempting to formulate an inference: 1) scanning, 2) scrutinizing, 3) probing, and 4) soliciting. Scanning is defined as an unbiased sweeping of the entire set of cues to detect cues relevant to the current enterprise. Scrutinizing involves a more active search for the relevant cues, while probing and soliciting are more advanced forms of inferential investigation. In this instance, the element of prime concern, again, is with how the judge scans the cue complex to detect relevant cues.

Gardner (1961) has noted individual differences in this scanning principle. He has found the ability to focus on relevant cues and to not attend to irrelevant cues to be independent of the cues in the stimulus complex. He calls this a "field articulation ability," or the ability to concentrate on certain cues and to ignore illusion-producing cues.
Studies on the factors influencing the selection and utilization
of cues have investigated several variables, among them number of cues
utilized, effects of immediate feedback, and the conceptual frame of
reference of the clinician. It is to this last problem of clinician bias
and original judgmental frame of reference that Chapman and Chapman (1967,
1969) relate. In a series of studies on factors influencing clinical
judgment, they have demonstrated how prior expectations of the relationships
between cues and criteria can lead to faulty observation and inference.
Moreover, their results indicate that judges who systematically make
invalid associations of pathognomicity to certain cues derived from pro-
jective protocols, may in fact be reacting to the prior expectation of
cue (sign) and symptom based on word association, rather than to the valid
and relevant cues present in the protocol.

Systematic Error in Clinical Report

Examples of systematic error of various kinds have been reported by
psychologists investigating other behavioral phenomena. Thorndike (1920a)
proposed the halo effect after studying the ratings of officers made by
their superiors in the army. Thorndike defined his halo effect as a
marked tendency or constant error toward suffusing ratings of specific
features with a positive quality belonging to the individual as a whole.
Another example, the Muller-Lyer illusion (Krech et al., 1969) occurs
in perceptual phenomena. The classical Muller-Lyer includes two horizontal
lines of the same length which appear unequal because of the opposite
directions of arrowheads attached to either end. The Muller-Lyer illusion
is an example of a part being dependent on the nature of a whole. Such a
part-whole relationship generalizes to problem solving and to stereotyping
and is usually associated with the perceptual effects referred to as one's "frame of reference." Loevinger (1965) has pointed to the presence of systematic error running through items of rating scales. The effect of the systematic error here is to raise item homogeneity and to thus increase and inflate the correlation of that test with other tests which have the same error. Loevinger noted that this correlation seduces many psychologists into clinging to a weak item form even when the incurring weakness of such an effect is pointed out.

The Chapmans believe that the principles of systematic error can be generalized regardless of the subject matter involved. Thus, errors in clinical observations are the same as superstitions, beliefs in magic, and social prejudice in that all are based on systematic error in the reports of observations of a supposed correlation between two classes of events. The Chapmans call this report of a correlation not warranted by the facts an "illusory correlation." "Correlation" here is not taken in the statistical sense, but is taken to mean co-occurrence; that is, an occurrence of one event tends to imply the occurrence of another. An illusory correlation may include both systematic and random error, and can occur under one or more of three conditions: 1) when two classes of events are not correlated, 2) when two classes of events are correlated to a lesser extend than reported, and 3) when two classes of events are correlated in the opposite direction of that reported.

As applied to clinical judgment, illusory correlation includes erroneous reports of correlations between symptoms of patients and their performance on diagnostic tests. This is particularly apparent when
performance on these tests is interpreted in terms of signs which can be derived from the stimulus complex of test responses. For example, on the DAP, "eye" or "elaboration of the eye" appears to have some unmeasured relationship of a cognitive or semantic nature -- and thus some associative value -- to "paranoia."

Loren Chapman (1967) noted the presence of reported illusory correlations between pairs of words which have associative properties and properties of distinctiveness. Two words may have associative properties when they are similar in meaning and associatively related. For example, "table" and "chair" have a high associative value because of their frequent occurrence with each other, as do "flag" and "banner" which are similar in meaning. Properties of distinctiveness are those properties which easily differentiate a word pair from other words of a list. An example would be a pair of three syllable words among pairs of one syllable words.

Chapman and Chapman (1967) generalized this systematic error to the clinical situation with six studies on the DAP. The purposes of their strategies were to determine the extent of illusory correlation in clinical observation, to investigate the bases of these errors, and to investigate the conditions under which these errors occur.

In order to test their hypothesis of illusory correlation in a quasi-clinical situation, the Chapmans needed to use a test and a symptom for which both valid and invalid signs have been reported by clinical researchers. They noted the controversy over sign interpretations of the DAP and selected that projective test. Thus, they set out to prove that the popular meanings being reported by clinicians were illusory correlations
based on verbal associative connection of the test sign to the symptom, rather than on the clinicians' valid observations.

First of all, they needed to know how clinicians used the DAP. Therefore, they sent a survey to 110 practicing psychodiagnosticians to learn the characteristics of the drawings that the diagnosticians reported that they had observed. Their survey took the form of six symptom statements: 1) he is worried about how manly he is; 2) he is suspicious of other people; 3) he is worried about how intelligent he is; 4) he is concerned with being fed and taken care of by other people; 5) he has had problems of sexual impotence; 6) he is worried that people are saying bad things about him. For each of these six symptom statements, the diagnosticians were asked to assume that the patient was a man and to list what the drawings of each would be characterized by. With this data they were able to compare reports made by clinicians with similar observations made by undergraduate observers performing the role of "beginning clinicians" in a quasi-clinical situation.

In their first study of this strategy, the Chapmans presented to undergraduate observers a series of DAP drawings, each of which was paired with two contrived symptoms of the alleged patient who drew it. By looking through the drawings (for a brief period of 30 seconds each) the observers were supposedly accumulating "clinical experience" as to the meaning of different aspects of performance on the DAP. The drawings and symptom statements were paired in such a way that there was no relationship between the occurrence of any symptom and any drawing characteristic (sign) viewed as its correlate in conventional clinical
practice. The drawings which were used were obtained from 35 diagnosed psychotics and ten graduate students.

Two of the six symptom statements given above appeared paired together on each card. Each of the six symptom statements was paired with every other statement in the set, thus making 15 possible pairs. A set of 45 different cards was used. Thus, each symptom statement was paired with every other symptom statement three times in the set of 45 cards, and each pairing appeared on a different card.

The hypothesis to be tested was that naive observers would "rediscover" the same widely accepted correlates of the six symptoms that clinicians erroneously use. The alternate hypothesis to follow was that if naive observers report the same correlates, shared systematic errors must be inferred. Moreover, if these were the same correlates that clinicians used, one might suspect that clinicians would show these same systematic errors.

After viewing the pictures, the subjects were asked to list the characteristics of each picture for each of the six symptom statements. The results (partially given in Table 3) indicated that the observers showed massive illusory correlation and that the illusory correlations that they reported showed remarkable similarity to the correlates that clinicians had reported from their clinical practice. Moreover, statistical cross-checks revealed that the observers saw each drawing characteristic as occurring with one symptom more than others, that the observers agreed on which symptom was with which characteristic, and that the correlates reported were not actually present in the test data.
In order to get at the associative basis of this illusory correlation, the Chapmans distributed a questionnaire to subjects who did not participate in the first task. The purpose of this questionnaire was to measure the associative strength between the problem area and the parts of the body referred to in various drawing characteristics. Each of the six problem areas was paired with seven body parts, giving 42 pairings in all. The format of the questionnaire ran as follows: "The tendency of SUSPICIOUSNESS to call to mind HEAD is . . . (a) very strong, (b) strong, (c) moderate, (d) slight, (e) very slight, (f) no tendency at all." The responses from (a) to (f) were assigned values from six to one respectively. The results were congruent with the earlier Chapman (1967) study; namely, the strongest associative connections were between signs and symptoms of the most often reported illusory correlations.

In the other five studies of the series on the DAP, the Chapmans repeated the task to the same observers on three consecutive days, presented the symptom statements without the accompanying drawings to determine amount of prior expectation, introduced negative correlations, and attempted to maximize the motivation of the observers. The illusory correlation phenomenon showed great resistance and was not markedly affected by these procedures.

The Chapmans (1969) also applied the illusory correlation or systematic error phenomenon to Rorschach content analysis. They selected erroneous signs (those which the literature failed to support), valid signs (those which the literature did support), and essentially neutral signs from Wheeler's 20 signs of homosexuality. They again surveyed
practicing clinicians, determined the word associative strength of the
sign and symptom pairs, and generally found the reports of clinicians
and undergraduate observers to agree. Moreover, both clinicians and
observers preferred invalid signs to valid signs in reporting illusory
correlates on the invalid signs. Their 13 conditions with almost 700
judges repeatedly verified the presence of illusory correlation, and further
testified to its hardiness.

Starr and Katkin (1969) investigated the phenomenon of illusory
correlation using Rotter's Incomplete Sentences Blank (ISB). They surveyed
practicing clinicians to determine what responses these clinicians had
observed with particular personality characteristics. They then presented
to clinical graduate students, non-clinical graduate students, and under­
graduate students ISB sentence stems and completions with paired statements
from a pool of five problem statements. Only five subjects (of 60)
correctly maintained that there was no relationship between the sentence
completions and the problem statements, and thus were not susceptible to
the illusory correlation. Moreover, the reported correlates were quite
similar to those reported by the clinicians in the survey.

Golding and Rorer (1972) investigated illusory correlation with
Rorschach responses. They chose Chapman and Chapman's (1969) paradigm
using valid and invalid signs of homosexuality. Golding and Rorer
determined the a priori expectations of undergraduate judges for a cue
to be associated with a particular symptom, and then tested two types of
feedback conditions -- simultaneous and predictive. Simultaneous feedback
occurred when the statements appeared on the cards with the pictures as in the Chapmans' designs (1967, 1969). In this feedback condition, the subjects were able to study the cue and symptom simultaneously on each trial. In prediction feedback, each subject was given a cue and was required to predict the symptom before he was allowed to see the true value of the symptom. This training resulted in a decrease in the report of illusory correlation. Nevertheless, the reports of illusory correlation were still far greater than chance. Nor were there any differential effects between the types of feedback given.

The Chapmans have noted that their findings of illusory correlation as reported by clinicians do not in themselves give evidence of personality qualities leading to deficits of judgment, but that they do clearly point out some of the difficulties inherent in the clinician's task. Golding and Rorer, did find consistent and strong differences among subjects in the nature of their response behavior. Subjects significantly differed from one another in the probabilities that they assigned to cue-symptom relationships as well as in their response to the conditions of training which were aimed at eliminating the report of illusory correlation. Fiske (1971) has noted that human judgment has deficiencies and that these deficiencies must be assessed to determine generalizability over judges, situations, and conditions. Therefore, it seems a wise course to pursue the illusory correlation phenomenon of the Chapmans to see if this error is indeed generalizable to all judges or whether judges highly susceptible to the error and who make unwarranted associations of sign and symptom present personality characteristics which reliably differ from judges who are not as susceptible to systematic error.
Personality Correlates of Interest

The amount of research on personality characteristics of the clinical judge is not impressive in sheer volume. Literature specific to the characteristics involved in the reading of the cues and to the association of cues with symptoms of pathognomonicity is even more noticeably lacking. Most of the studies which are available consider the judgmental process of the clinician on a more global scale and thus report more global personality characteristics of various qualities of judges. Adams (1927) and Vernon (1933) distinguished the good judge of others and the good judge of self. However, their reported characteristics did not include measures of specific abilities. Kelly and Fiske (1951) performed one of the more extensive investigations of the abilities and personality characteristics of the clinician. However, here again, the correlation of specific traits with the personality of the good judge were not remarkably high. Nor did Kelly and Fiske present measures which would address the specific ability to respond to valid or invalid cues.

Thus, it may be necessary to look at cognitive processes and to infer the possible influence of these processes on clinical judgment. First to be considered are field dependence and field independence. In considering whether a judge responds to a present relationship between cue and symptom, or whether or not he selects relevant cues, the ability to detect the cue in the stimulus complex would seem to be of prime importance. Witkin (1962) has demonstrated that people differ in cognitive style and that they have characteristic, self-consistent ways of functioning in their perceptual and intellectual activities. He has demonstrated that some people have an ability to analyze a complex configuration and to respond to specific parts of it, ignoring other parts. To these people he attributes the trait
of field independence. Conversely, people who have difficulty eliciting a specified part from the whole complex are labeled field dependent.

Witkin and later researchers have also found particular personality characteristics to be associated with each of these types of cognitive styles. Field dependent persons have been found to be in need of more support and guidance from others. They are thus particularly attentive to facial characteristics and expressions which provide cues to other people's attitudes and moods. Field dependent persons characterize themselves and others in terms of "external" attributes. Messick and Damarin (1964) found field dependent people to have a better incidental recall of photographs and to be relatively better at recognizing people they have seen only briefly before. Field independent persons, however, are better at incidental recall of items given in a stimulus complex.

If, as the Chapmans and others have suggested, reporters of illusory correlations or utilizers of invalid cues are relying on prior expectations, perhaps they are doing so because of an inability to select some part from the complex stimulus configuration of the projective test protocol. Field dependent persons should more readily form proper cue-symptom relationships because of their attentiveness to expressed affective cues of others. However, it may be that their global handling of the affective cues interferes with their observation of frequency of cue-symptom pairings. Thus, one might hypothesize that those more susceptible to the systematic error of word association of sign and symptom are more field dependent than are those who are not as susceptible and who are thus free to see cues and their symptoms in proper relationships.
A second variable to be considered is creativity. In their book on creative thinking, Wallach and Kogan (1965) considered Guilford's (1967) mode of divergent thinking. Divergent thinking can roughly be described as thinking in different directions, and as searching for variety. Considered within this domain are the cognitive units of originality, expressional fluency, ideational fluency, word fluency, and associational fluency. One aspect of the capacity to generate these cognitive units is the limit of a person's behavioral repertoire of cognitive productions. Thus, Wallach and Kogan define creativity as the capacity to store cognitive elements.

Mednick (1962) differentiated two variables which reflect individual differences in creativity. First of all, there is the total number of associations of which a person is capable; and secondly, is the relative uniqueness that his associations possess. Gradients of associative response strength are different for high and low creative persons, and these gradient slopes represent the emission potential of the person. To use Mednick's example, if the word "table" is presented, the highly stereotypic response would be "chair." The stereotypic responses are high in a person's hierarchy, while the unique associates are lower in the hierarchy. The sharply-sloped gradient of available responses which defines the less creative person contains the responses of greatest stereotypy. The shallow-sloped gradient of available responses which defines the more creative person represents a larger number of available cognitive elements. Thus, a person who can be represented by the shallow gradient will most likely offer stereotyped associates to begin with, just as the person operating under the sharp gradient. However, when the less creative person (repre-
sented by the sharp slope) exhausts his repertoire of responses, the more creative person (represented by the shallow slope) will continue responding with increasingly unique associates. Thus, the word association behavior of the highly creative individual should be characterized by less stereotypy and commonality.

The Chapmans (Chapman, 1967; Chapman & Chapman, 1967) have presented strong evidence testifying that one basis of illusory correlation is the word association between symptom and cue. The most oft-reported illusory correlates are those with the strongest cue-symptom word association. Thus, one might suspect that those persons most susceptible to the illusory correlation phenomenon are less creative than those who are less susceptible to the illusory correlation.

Another variable to be considered is social intelligence. The ability to read and to react to affective cues from another person and to respond with appropriate affect is the essence of the ability termed social insight, or social intelligence (Guilford, 1967; Thorndike, 1920b). This ability appears to be closely related to the cognitive processes of creativity in that both are prominent correlates of sensitivity to physiognomic properties. These physiognomic properties are those of a cue presentation which provide information about the affective or emotional significance (sadness, fear, weakness, etc.) of such stimuli. Reading the stimulus cues to detect physiognomic properties requires the positing of a symbol-referent relation between a visual pattern and an emotional state. Thus, a particular visual configuration signifies a particular affect. The ability to posit such
a relationship between cue and affect is the product of a learned inference, and it thus varies from person to person. Creativity and social intelligence differ in that creativity refers to the capacity for production of relationships between visual patterns and emotional states, while social intelligence refers to the ability to understand the social consequences of pattern-state relationships and to then produce behaviors acting in accordance with such an understanding. The judge high in social intelligence possesses more of the ability to detect the social consequences and affective tone of a cue configuration. Consequently, when attempting to infer personality characteristics from the cues presented on a test protocol, the person with greater social intelligence should make more valid associations of cue and symptom. In addition, if such a person is more readily aware of the social consequences, one might suppose that he is more aware of external cues as well. Thus, one might suspect that a person high in social intelligence is likewise aware of frequency of cues and that he therefore would be less susceptible to the illusory correlation phenomenon.

A fourth personality-related dimension which might properly be considered is the confidence with which a judge designates a cue or set of cues to be indicative of a symptom. This confidence should reflect the degree to which that association is learned (Hunt & Jones, 1961). If the Chapmans are correct that illusory correlation is a learned verbal association and that such a learned verbal association represents greater stereotypy, then one would expect in this instance that the judges who report invalid signs would express more confidence in their choices.
A fifth consideration is that the number of cues utilized may be related to success. Wescott (1968), in studying intuitive problem solving with his Test of Intuitive Problem Solving, failed to find a positive correlation between success and number of cues utilized. Hunt and Walker (1971) did find success and number of cues to be positively related. Westcott noted that some subjects apparently possessed the ability and did extract more information than did others. His findings did indicate that the probability of success increased with additional information. If reporters of illusory correlations are reporting stereotypic learned relationships, then it can be expected that they would utilize a lesser number of cues in making a judgmental inference.

Choosing an Experimental Paradigm

Determining the personality characteristics and abilities of the highly error-prone judge and less error-prone judge requires a controlled situation in which all judges are reviewing the same test stimuli with the same amount of background information. In order to simulate a clinical situation in which judges review test stimuli and then must make a decision about the personality of the patient, the paradigm used by the Chapmans will be employed. Goldberg's (1968) acclamation of the Chapmans' studies as "one of the most ingenious studies of clinical judgment ever made" certainly singles out their methodology for consideration.

If one attempts to replicate the Chapmans' discovery of illusory correlations in clinical observations, there is a choice between their design on the DAP or their design on the Rorschach. While the Rorschach design offers the advantage of differentiating valid and invalid signs,
the DAP more closely resembles a judgmental situation involving clinical inference based on subjective associations of cues and symptoms. The essential elements of investigation are the personality characteristics of the judges or observers. The DAP will suffice to differentiate those judges who report illusory correlations (cue-symptom relationships which are not present) of a highly stereotypic, associative nature, and those judges who report relationships of a less associative nature. The former judges will be designated as those highly susceptible to the illusory correlation phenomenon, while the latter judges will be designated as those less susceptible to illusory correlation.

The hypotheses to be tested are that judges who are highly susceptible to illusory correlation: 1) are more field dependent, 2) score lower on tests of creativity, 3) score lower on tests of social intelligence, 4) are more confident in their choice of cues, and 5) utilize a lesser number of cues than judges who are less susceptible to illusory correlation.
CHAPTER II

Method

Subjects

Ninety-two male undergraduates enrolled in an introductory psychology course at Loyola University served as subjects. They were questioned for previous experience with the DAP, but no subjects reported any prior experience with this instrument. Twenty-two subjects participated in Experiment I, the reliability design, but were not permitted to participate in Experiment II, the personality correlates design. Twenty-one of those twenty-two subjects returned for the second testing. Seventy subjects participated in Experiment II and are henceforth referred to as judges.

Task Materials

The materials used and the procedure of stimulus administration approximated those of the Chapmans as closely as possible. The drawings were, in fact, the same as those employed in the Chapman design,* which were collected from ten nonpsychotics and 35 psychotics. Xerox reproductions were obtained of each drawing, and each drawing was backed by cardboard and covered with clear plastic to protect it from the effects of handling.

Six symptom statements were used for pairing with the drawings. The six symptom statements were:

1. He is worried about how manly he is.
2. He is suspicious of other people.
3. He is worried about how intelligent he is.
4. He is concerned with being fed and taken care of by other people.
5. He has had problems of sexual impotence.
6. He is worried that people are saying bad things about him.

* The author would like to thank Loren Chapman for generously sharing his stimulus drawings.
Printed on the same sheet of paper as each drawing were a pair of the above statements purporting to represent symptoms of the patient who made the drawing. All possible pairs (15) of the six symptom statements were used. Each pairing appeared three times to make a total set of 45 cards. Thus, each symptom statement appeared on 15 different cards. In order to prevent a relationship between the occurrence of each symptom and any of the drawing characteristics expected (that is, any of the 14 drawing characteristics which the Chapmans obtained from their survey of clinicians), the Chapmans had had each drawing ranked according to the presence of each characteristic. Thus, assigning symptom statements to drawings was done in such a way that no systematic relationship existed between the occurrence of a symptom and any drawing characteristic. This was accomplished by pairing the symptom statement equally often with the drawings having each of various degrees of possession of the drawing characteristic. For example, the symptom statement, "He is worried about how intelligent he is." was paired as often with drawings having small heads as with large or medium-sized heads. This prevented any actual correlation between cue and symptom.

**Experiment I**

Chapman and Chapman (1967) indicated that a group of judges reported similar percentages of illusory correlates for the same cues in the repetition of the task on three consecutive days. Nevertheless, this did not clearly establish that the same judges were reporting the same illusory correlates on each of these three presentations. To test the reliability of this phenomenon for individual judges over successive administrations, Experiment I was employed.
Procedure

In Experiment I, twenty-two subjects were tested in a group setting. The experimenter first presented a brief description of the DAP and its clinical use, explaining that psychologists make interpretations about patients' emotional problems from the nature of their drawings. No examples of drawing characteristics were offered.

The following instructions were then given:

Now we want to test your powers of judgment and observation. I am going to show you some drawings made by men with various emotional problems. Together with each drawing you will find two statements that describe the emotional problems of the man who made the drawing. Many of the men have some of the same problems. Please study the pictures and the statements carefully because when you are through I am going to ask you about the characteristics of the drawings that were made by men with each kind of problem.

Each subject was handed a drawing face down. At a signal, the subjects looked at the card for 30 seconds and on signal passed the card on, receiving a new card in the process. This was repeated every 30 seconds until each of the subjects had seen each of the 45 drawings once.

After all the subjects had seen all of the 45 drawings, a questionnaire was distributed to each subject with items of the following format:

Some of the pictures which you saw were drawn by men with the following problem:

"He is worried about how manly he is."
Did you notice any general kind of thing or things which most often characterized the drawings made by men with this problem?
Yes ____ No ____
If your answer is yes, the pictures drawn by these men were more often characterized by:

......................
......................
......................
......................
A format like the above was constructed around each of the other five symptom statements as well.

When all of the subjects had completed all six of the items, the subjects were asked to list the amount of confidence they felt in making the associations of signs and symptoms for each of the six statements on a scale of:

(1) Guessed, (2) Rather uncertain, (3) Fairly certain, (4) Rather certain, (5) Positive.

Two weeks later, the subjects were recalled. Twenty-one subjects returned and performed the same task with identical materials, instructions, and procedure.

**Experiment II**

The studies on illusory correlation indicate that the phenomenon is massively reported in nearly all conditions. Very few judges correctly indicate that there are no cue-symptom relationships present. Since there are no valid cue-symptom relationships present, any reporting of a relationship may be considered to be erroneous. Evidence seems to indicate that most judges are imbued with prior expectations of a cue-symptom relationship and that this prior expectation may be the cause of the illusory phenomenon (Chapman & Chapman, 1967, 1969; Golding & Rorer, 1972). In addition, the most popular illusory correlates are those of the highest associative strength, and, thus, are those of greatest prior expectation. Evidence from these studies suggests that the most popular illusory correlates are those which interfere with the reporting of valid
cue-symptom relationships. Thus, those judges who report the most popular illusory correlates may be identified as those highly susceptible to illusory correlation, while those judges who report no relationships or who do not report the most popular illusory correlates may be identified as those less susceptible to illusory correlation. The purpose of Experiment II then, was to differentiate between those judges highly susceptible to the illusory phenomenon and those judges less susceptible to the illusory phenomenon, and to compare scores of these highly susceptible and less susceptible judges on personality-ability measures of field dependence-independence, social intelligence, and creativity.

Procedure

In Experiment II the seventy subjects were tested in groups varying in size from eight to seventeen. The materials, instructions, and procedure were the same as those of the task in Experiment I. The most commonly reported illusory correlations for each symptom statement in both Experiments I and II were compared with the most often reported correlates given by surveyed clinicians published in Chapman and Chapman (1967).

Utilizing the Chapmans' information as criteria, scoring standards for two scales were formulated, one employing the most frequently reported (most "popular") drawing characteristic for each symptom statement as given by surveyed clinicians, and one employing the most popular responses given by the Chapmans' experimental undergraduate subjects. Thus, according to each scale, the range of erroneous populars which an individual
subject could report observing across the six symptom statements extended from zero to six; that is, potentially one for each of the six symptom statements. Those judges who reported a large number of illusory populars (four, five, or six) were considered highly susceptible to the illusory phenomenon, while those who reported no or few populars (zero, one, or two) were considered to be less susceptible to the illusory phenomenon. Since the results of the personality measures are to be generalized to practicing clinicians, more attention has been paid to the scale of the clinician responses (hereafter referred to as the "Clinician Scale") as opposed to the scale of the student responses (hereafter referred to as the "Student Scale").

Measures

Five tests representing three personality characteristics and abilities were then group administered. The characteristics and measures are as follows:

a) A test of field dependence, the Hidden Figures Test (HFT, Cf-1) (Witkin et al., 1962) was used to determine scores on a continuum of field dependence-independence. The HFT, Cf-1 resembles the Witkin Embedded Figures Test in that it measures the ability to recognize a figure hidden among other lines.

b) Two of Wallach and Kogan's (1965) creativity tests were utilized to determine level of creativity. Two items of Alternate Uses ("cork," "chair") and two items of Similarities were employed ("watch-typewriter," "curtain-rug").
c) Two of Guilford's tests (O'Sullivan & Guilford, 1966), Cartoon Predictions and Expression Grouping, were used to measure social intelligence or the ability to draw implications from social relationships.

Two additional scores obtained from an analysis of responses on the experimental task were:

d) Level of Confidence was determined from the judges' subjective report as obtained from the confidence rating scale.

e) Average Number of Cues Utilized was determined by counting and averaging the number of signs reported by each subject for each symptom statement.
CHAPTER III

Results

Experiment I

The results of the test-retest reliability design are presented in Table 1 for the twenty-one subjects who completed both tasks. The results are given according to the Clinician Scale and according to the Student Scale. (Each of these Scales was scored on first and second testing according to the number of popular responses reported, with a possible range of zero to six.) A test-retest correlation of 0.65 was obtained for the Clinician Scale (most frequently presented clinician responses) using the Pearson product-moment. A test-retest correlation of 0.70 was obtained for the comparable Student Scale (most frequently presented student responses). Table 1 also indicates the percentage of individual responses which were repeated on the second testing. For example, if the characteristic "head emphasis" was reported for the symptom statement "He is worried about his intelligence," on both test and retest, this was scored as a repetition. Seventy-five percent of such popular responses were repeated by subjects on second testing utilizing the Clinician Scale, while 72% of the popular responses were repeated by the subjects on second testing utilizing the Student Scale. This can be compared to the 56% of first responses, popular and non-popular, repeated on second testing for all the subjects.

Additional data from the test-retest design revealed that the average number of responses given for each symptom statement on first testing was 1.24, while a comparable average of 1.27 responses was given
Table 1

Reliability Coefficients and Percentages of
Popular and Total Responses Repeated on Both Testings

<table>
<thead>
<tr>
<th></th>
<th>Clinician Scale (^a)</th>
<th>Student Scale (^a)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reliability (^b)</td>
<td>(0.65)</td>
<td>(0.70)</td>
</tr>
<tr>
<td>(Number of Popular Responses)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percentage of Popular Responses Repeated on Both Testings</td>
<td>(75%)</td>
<td>(72%)</td>
</tr>
<tr>
<td>Percentage of Total Responses Repeated on Both Testings</td>
<td></td>
<td>(56%)</td>
</tr>
</tbody>
</table>

\(^a\) Formulated from Chapman and Chapman (1967) data.

\(^b\) Pearson product-moment correlation coefficient.
for each symptom statement on second testing. The average confidence level per response on first testing was 3.60, while the figure for second testing was slightly less at 3.53. The average number of "No Relationship" responses per subject of first testing was 1.14; on second testing, 1.00 "No Relationship" responses were reported by each subject.

Experiment II -- Scale comparisons

The responses of the 70 questionnaires of Experiment II are given in Table 2. The results are presented in terms of percentage of subjects who reported each characteristic across the symptom statements. Only the most frequently reported characteristics are presented. In addition, the percentage of subjects reporting "No Relationship" is also given. Cochran Q analyses were computed for each of the descriptive statements to determine if the response characteristics were randomly distributed among the respective symptom statements. The distribution of characteristics across each symptom statement was found to be non-random for five of the statements (p< .01 in each case). These results show that the judges saw one or more drawing characteristic as occurring with a statement more often than any other characteristic.

A comparison of Table 2 with the Chapmans' data (Table 1, page 196, Chapman & Chapman, 1967) shows that the same characteristics tended to be reported for the six statements in both studies. One of the Chapmans' fifteen characteristics, "phallic limbs," was not reported by the judges
Table 2  
Percent of Subjects Reporting Each Characteristic According to Symptoms (n=70)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Manly, muscular</td>
<td>68</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>15</td>
</tr>
<tr>
<td>2. Feminine, child-like</td>
<td>32</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>39</td>
</tr>
<tr>
<td>3. Detailed drawing</td>
<td></td>
<td>42</td>
<td>12</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Ears atypical</td>
<td></td>
<td></td>
<td></td>
<td>20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Facial expression atypical</td>
<td>18</td>
<td>20</td>
<td>15</td>
<td></td>
<td></td>
<td>31</td>
</tr>
<tr>
<td>7. Head emphasized</td>
<td>16</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>59</td>
</tr>
<tr>
<td>8. Mouth emphasized</td>
<td>15</td>
<td></td>
<td></td>
<td>15</td>
<td></td>
<td>8</td>
</tr>
<tr>
<td>9. Passive posture outstretched arms</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>18</td>
</tr>
<tr>
<td>10. Sexual area</td>
<td>16</td>
<td>15</td>
<td>15</td>
<td></td>
<td></td>
<td>59</td>
</tr>
<tr>
<td>11. Clothing</td>
<td></td>
<td>15</td>
<td>15</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12. Arms, hands</td>
<td></td>
<td></td>
<td></td>
<td>15</td>
<td></td>
<td>15</td>
</tr>
<tr>
<td>13. Profile</td>
<td></td>
<td>15</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No Relationship</td>
<td>4</td>
<td>8</td>
<td>10</td>
<td>18</td>
<td>10</td>
<td>25</td>
</tr>
<tr>
<td>Cochran Q^a</td>
<td>36.09^b</td>
<td>14.00^b</td>
<td>24.50^b</td>
<td>11.02^b</td>
<td>43.57^b</td>
<td>5.60</td>
</tr>
</tbody>
</table>

^a Cochran Q method was utilized with the three characteristics ("No relationship" responses not included) most frequently reported for each of the symptom statements. The Q value represents the probability that the frequencies of the responses differ from random distribution. If the observed Q value is equal to or greater than the critical value (for p < .01), the implication is that the frequencies of the reported characteristics differ significantly among themselves.

^b p < .01
in the present experiment. In addition, the characteristics, "sexual area emphasized" and "sexual area deemphasized" have been combined into one category "sexual area" due to the difficulty in clearly differentiating judges' responses into two categories of a sexual nature.

Table 3 combines summaries of the Chapmans' results and the results of the present study. A comparison is made between the reports of the Chapmans' surveyed clinicians, the reports of the Chapmans' experimental subjects, and reports of the present experimental judges. Only the three characteristics most frequently reported for each statement by judges in each of the above three groups have been included. One can see a high amount of congruity between the reports of the Chapmans' students and the reports of the present student judges. This suggests experimental reliability in both the Chapmans' and the present design. Also, once again, there is substantial congruity between the reports of surveyed clinicians and the reports of undergraduate students. Utilizing the most frequently reported response for each of the six statements, these responses were summed to get scores for both the Clinician Scale and the Student Scale. In other words, the reports of the 70 judges were compared in terms of scoring on the Clinician Scale and scoring on the Student Scale. A correlation of 0.79 was obtained between these two scorings with the Pearson product-moment. This suggests high but not complete agreement between clinicians and student judges. Inspection of Tables 2 and 3 indicates that results from four of the six statements (Numbers 1, 2, 3, and 5) are more congruent between clinicians and students than are results from
### Table 3
Comparison of Percentages of Characteristics Reported by Chapmans' Clinicians and Students and Present Experimental Group

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Clinicians&lt;sup&gt;a&lt;/sup&gt; n = 44</th>
<th>Students&lt;sup&gt;a&lt;/sup&gt; n = 108</th>
<th>Experimental Group n = 70</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Muscular manly</td>
<td>80</td>
<td>76</td>
<td>68</td>
</tr>
<tr>
<td>2. Feminine, childlike</td>
<td>23</td>
<td>22</td>
<td>32</td>
</tr>
<tr>
<td>3. Sexual area</td>
<td>14</td>
<td>5</td>
<td>16</td>
</tr>
<tr>
<td>No Relationship&lt;sup&gt;b&lt;/sup&gt;</td>
<td>--</td>
<td>--</td>
<td>4</td>
</tr>
<tr>
<td>1. Eyes atypical</td>
<td>91</td>
<td>58</td>
<td>42</td>
</tr>
<tr>
<td>2. Ears atypical</td>
<td>55</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>3. Facial expression atypical</td>
<td>18</td>
<td>44</td>
<td>18</td>
</tr>
<tr>
<td>No Relationship</td>
<td>--</td>
<td>--</td>
<td>8</td>
</tr>
<tr>
<td>1. Head large or emphasized</td>
<td>82</td>
<td>55</td>
<td>53</td>
</tr>
<tr>
<td>2. Facial expression</td>
<td>2</td>
<td>21</td>
<td>20</td>
</tr>
<tr>
<td>3. Detail</td>
<td>34</td>
<td>13</td>
<td>12</td>
</tr>
<tr>
<td>No Relationship</td>
<td>--</td>
<td>--</td>
<td>10</td>
</tr>
<tr>
<td>1. Mouth emphasized</td>
<td>68</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>2. Feminine, childlike</td>
<td>32</td>
<td>39</td>
<td>39</td>
</tr>
<tr>
<td>3. Passive posture</td>
<td>36</td>
<td>21</td>
<td>18</td>
</tr>
<tr>
<td>No Relationship</td>
<td>--</td>
<td>--</td>
<td>18</td>
</tr>
<tr>
<td>1. Sexual area</td>
<td>73</td>
<td>35</td>
<td>59</td>
</tr>
<tr>
<td>2. Manly, muscular</td>
<td>25</td>
<td>31</td>
<td>15</td>
</tr>
<tr>
<td>3. Feminine, childlike</td>
<td>23</td>
<td>25</td>
<td>8</td>
</tr>
<tr>
<td>No Relationship</td>
<td>--</td>
<td>--</td>
<td>10</td>
</tr>
<tr>
<td>1. Ears atypical</td>
<td>64</td>
<td>7</td>
<td>14</td>
</tr>
<tr>
<td>2. Facial expression atypical</td>
<td>18</td>
<td>52</td>
<td>31</td>
</tr>
<tr>
<td>3. Eyes atypical</td>
<td>43</td>
<td>26</td>
<td>20</td>
</tr>
<tr>
<td>No Relationship</td>
<td>--</td>
<td>--</td>
<td>25</td>
</tr>
</tbody>
</table>

<sup>a</sup> From Chapman and Chapman (1967).

<sup>b</sup> The Chapmans' data did not include a "No Relationship" response.
the other two statements (Numbers 4 and 6). This is suggested by agreement between clinicians and students on the most popular response on these four statements as well as highly significant Cochran Q values, and finally by the tendency of experimental judges to give fewer "No Relationship" responses on these four statements. In contrast, for the two weakest measures of illusory correlation, Statements 4 and 6, respectively 18% and 26% of the judges reported observing "No Relationship." While the Cochran Q value was significant at the .01 level for Statement 4, no characteristic was reported by more than 39% of the judges.

Experiment II -- Personality measures

Utilizing the Clinician Scale (i.e., the characteristic most frequently reported by clinical psychologists for each of the six statements) the number of illusory populars was computed for each subject on a potential scale of zero to six. The top horizontal columns of Tables 4 and 5 categorize the 70 judges according to the number of populars they gave. The modal number of illusory populars reported was two. No reports of six illusory populars were obtained in Experiment II. A look at the distribution of reported illusory populars over the Clinician Scale suggests a positively skewed distribution of responses. In addition, in Tables 4 and 5, the mean scores and standard deviations of five personality-ability measures and two measures derived from the experimental task are given for the different values on the Clinician Scale.

Table 4 presents a division of the judges into low (zero, one, or two reported populars) and high (three, four, five, or six reported populars) reporters of illusory correlation. This division of the
Table 4

Means, Standard Deviations, and t-Scores of Selected Measures for Low and High Reporters of Illusory Populars

<table>
<thead>
<tr>
<th>Measure</th>
<th>Low n = 39 0-2 Populars Reported</th>
<th>High n = 31 3-6 Populars Reported</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Hidden Figures Test</td>
<td>Mean 11.14 SD 6.24</td>
<td>Mean 12.21 SD 6.51</td>
<td>0.70</td>
</tr>
<tr>
<td>2. Expression Grouping</td>
<td>18.87 2.47</td>
<td>19.50 2.49</td>
<td>1.06</td>
</tr>
<tr>
<td>3. Cartoon Predictions</td>
<td>23.44 2.37</td>
<td>23.38 1.90</td>
<td>1.00</td>
</tr>
<tr>
<td>4. Similarities</td>
<td>12.95 5.21</td>
<td>13.32 4.37</td>
<td>0.31</td>
</tr>
<tr>
<td>5. Alternate Uses</td>
<td>14.10 5.41</td>
<td>16.42 5.42</td>
<td>1.78a</td>
</tr>
<tr>
<td>6. Average Number of Responses Per Item</td>
<td>1.42 0.56</td>
<td>1.70 0.62</td>
<td>1.99a</td>
</tr>
<tr>
<td>7. Average Confidence Per Response</td>
<td>3.21 0.60</td>
<td>3.36 0.46</td>
<td>1.11</td>
</tr>
</tbody>
</table>

a $p < .1$
distribution of reported populars gives 39 low reporters versus 31 high reporters. The means and standard deviations for each of the five personality measures and the two task-derived measures are given, and differences in the means of these measures between low and high reporters are compared using Student's t-test. Two measures, Alternate Uses and Average Number of Responses Per Item, were found to approach significance at the 0.10 level (two-tailed), suggesting a trend for each of these measures. No other significant differences in means were found.

The same results are rearranged and presented in Table 5 according to four categories of reported populars (zero-one, two, three, four-six). As Table 5 indicates, categories zero and one, as well as four, five, and six have been combined due to the low frequencies of reports. Once again, the means and standard deviations of the five personality measures and the two task-derived measures are given. A one-tailed F-test was used to test for significant differences in means of personality measures across number of reported illusory populars, with a Newman-Keuls analysis (unequal n correction) of the differences between all possible comparisons of means. Table 6 presents the F values of these comparisons for the seven dependent measures. For Expression Grouping, a measure of social intelligence, significant differences were obtained between the means of zero-one and two reported illusory populars (p < .005), zero-one and four-six reported illusory populars (p < .005), and zero-one and three reported illusory populars (p < .025). A significant difference (p < .05)
Table 5
Means and Standard Deviations of Selected Measures According to Number of Illusory Poplars Reported

<table>
<thead>
<tr>
<th>Measure</th>
<th>0-1 n = 13</th>
<th>2 n = 26</th>
<th>3 n = 18</th>
<th>4-6 n = 13</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Hidden Figures Test</td>
<td>11.03 5.95</td>
<td>11.20 7.04</td>
<td>13.32 6.78</td>
<td>10.68 6.04</td>
</tr>
<tr>
<td>2. Expression Grouping</td>
<td>16.92 2.83</td>
<td>19.85 2.47</td>
<td>19.36 2.37</td>
<td>19.69 2.56</td>
</tr>
<tr>
<td>3. Cartoon Predictions</td>
<td>23.05 1.51</td>
<td>23.64 2.63</td>
<td>23.91 1.90</td>
<td>24.05 1.95</td>
</tr>
<tr>
<td>5. Similarities</td>
<td>12.15 4.00</td>
<td>13.35 5.17</td>
<td>13.83 4.03</td>
<td>12.62 5.51</td>
</tr>
<tr>
<td>6. Average Number of Responses Per Item</td>
<td>1.38 .47</td>
<td>1.44 .60</td>
<td>1.54 .50</td>
<td>1.92 .68</td>
</tr>
<tr>
<td>7. Average Confidence Per Response</td>
<td>3.34 .44</td>
<td>3.15 .66</td>
<td>3.34 .38</td>
<td>3.38 .60</td>
</tr>
</tbody>
</table>
Table 6

F-Values\(^a\) for the Difference Between Means of Selected Measures

According to Number of Illusory Populars Reported

<table>
<thead>
<tr>
<th>Measure</th>
<th>Mean Comparisons of Categories of Illusory Populars Reported</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0-1</td>
</tr>
<tr>
<td></td>
<td>to</td>
</tr>
<tr>
<td></td>
<td>2</td>
</tr>
<tr>
<td>1. Hidden Figures Test</td>
<td>.17</td>
</tr>
<tr>
<td>2. Expression Grouping</td>
<td>2.93 (^b)</td>
</tr>
<tr>
<td>3. Cartoon Predictions</td>
<td>.59</td>
</tr>
<tr>
<td>4. Alternate Uses</td>
<td>.88</td>
</tr>
<tr>
<td>5. Similarities</td>
<td>1.20</td>
</tr>
<tr>
<td>6. Average Number of Responses Per Item</td>
<td>.06</td>
</tr>
<tr>
<td>7. Average Confidence Per Response</td>
<td>.19</td>
</tr>
</tbody>
</table>

\(^a\) Newman-Keuls analysis (unequal n's)
\(^b\) \(p < .005\)
\(^c\) \(p < .025\)
\(^d\) \(p < .05\)
was also found between the means of Average Number of Responses Per Item for zero-one reported illusory populars and four-six reported illusory populars. No other differences were significant.

The results of Experiment II were further analyzed to determine whether a relationship existed between scores on each of the five personality measures and the number of "No Relationship" responses reported by the experimental judges. Table 7 presents the mean number of "No Relationship" responses for high and low scores (median split) on each of the five personality measures, and the corresponding t-values. None of the differences were significant.
Table 7

Mean Number of "No Relationship" Responses Reported for
High and Low Scorers on Selected Personality Measures

<table>
<thead>
<tr>
<th>Measure</th>
<th>Low Scorers a n = 35</th>
<th>High Scorers a n = 35</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Hidden Figures Test</td>
<td>.77</td>
<td>.66</td>
<td>.55</td>
</tr>
<tr>
<td>2. Expression Grouping</td>
<td>.83</td>
<td>.60</td>
<td>.99</td>
</tr>
<tr>
<td>3. Cartoon Predictions</td>
<td>.66</td>
<td>.77</td>
<td>.55</td>
</tr>
<tr>
<td>4. Alternate Uses</td>
<td>.80</td>
<td>.63</td>
<td>.82</td>
</tr>
<tr>
<td>5. Similarities</td>
<td>.83</td>
<td>.60</td>
<td>1.10</td>
</tr>
</tbody>
</table>

* Median split.
CHAPTER IV

Discussion

The results of the present study replicate and add validity to the data of Chapman and Chapman (1967). A high amount of congruity can be observed between the results of the present design and the results presented for the Chapmans' student subjects. Furthermore, the reports of the present 70 student judges are once again in high agreement with the reports of the Chapmans' surveyed clinicians. A relatively high correlation (.79) between clinicians' and students' reports was obtained by further investigating the present data with the construction of a Clinician Scale and a Student Scale of the most frequently reported characteristics. Thus, the evidence of the present study adds further support to the illusory correlation hypothesis offered by the Chapmans that many clinical interpretations of projective test material have their origins in illusory correlations based on naively and intuitively formed associative connections.

The test-retest correlations suggest that the report of an illusory relationship is a relatively stable phenomenon. Reliability coefficients of .65 (Clinician Scale) and .70 (Student Scale) were obtained for the most popular illusory responses. As the data in Table 1 suggest, the illusory popular is more reliably reported than are nonpopular reports of relationships. According to Clinician-Scale scoring and Student-Scale scoring, respectively, 75% and 72% of the popular responses reported on first administration were again reported by individual
subjects on a second administration of the DAP materials. For the total responses, including both popular and nonpopular responses, only 56% of first responses were again reported on second testing. This difference in stability of report for popular responses and total responses encourages the explanation that the symptom-characteristic associative relationship of an illusory popular is stronger than the symptom-characteristic associative relationship of a nonpopular report. In addition, there is evidence that the associative strength of characteristic to symptom differentially varies for each of the six symptom statements. This suggests that some characteristic-symptom populars have stronger associations than other characteristic-symptom populars. For example, the illusory populars "head-intelligence" and "sexual area-impotence" were reported by a greater percentage of the subjects than were the populars "mouth-fed and cared for" or "eyes-suspicious." Likewise, there was less tendency to give a "No Relationship" response for the symptom-characteristics of greatest associative strength.

The most interesting finding regarding personality-ability measures was the marked lack of significant differences between high and low reporters of illusory correlation. As the information on Table 4 suggests, the high-low dichotomy yielded only trend differences for two measures, Alternate Uses and Average Number of Responses per Item. In addition, these trend indications were opposite to those offered in the original hypotheses; namely, it was hypothesized that low reporters of illusory correlation would score higher on these variables than would high reporters.
of illusory correlation. Since Alternate Uses, as a measure of creative productivity, is related in concept to Average Number of Responses per Item, what the data seem to suggest, then, is that high reporters of illusory correlation tend to see a greater number of characteristic-symptom relationships in the stimulus material, or, generalizing to the clinician's task, in projective test protocols.

Since a high-low dichotomy may include curvilinear data and thus shield true differences, an extended analysis of the number of reported illusory populars and personality measures was performed. Tables 5 and 6 present this information. Here, the number of reported populars were grouped into four categories -- zero-one, two, three, and four-six reported populars. Only one personality-ability variable, Expression Grouping, which is a measure of social intelligence, was found to differ significantly across the number of illusory populars reported. Those who reported zero or one popular significantly differed from those reporting two populars ($p < .005$), three populars ($p < .025$), and four, five, or six populars ($p < .005$). Once again, the original hypothesis that high reporters would be more stereotypic and less socially sensitive to cues, and thus less socially intelligent was contradicted. According to the present evidence, the opposite situation may be suggested; namely, that those judges who report few illusory populars (zero-one) manifest less social intelligence than judges highly susceptible to the illusory phenomenon. A recent article (Kelley, 1973) discussing the Chapmans' illusory phenomenon within the realm of attribution theory and implicit personality theory lends some
support to this suggestion. If the learned associative basis of the illusory correlation phenomenon is a result of experience stored as an implicit cognitive organization, then perhaps the stronger and more abundant the associative cue relationships, the greater the amount of social intelligence, which is also sometimes thought to be dependent on relatively covert cues.

Tables 5 and 6 also indicate a trend difference in Average Number of Responses per Item. As noted and discussed above, the original hypothesis was not supported; the opposite situation was suggested in that high reporters of illusory populars gave more responses than did low reporters of illusory populars. In utilizing a technique of self-report other than an open-ended questionnaire, one might be able to discern whether the number of populars reported is merely a function of increased responses, or whether, as the present evidence suggests, is indeed dependent upon differential associative response strengths. For example, with a multiple-choice, one-response-per-item format on the questionnaire the number of responses would be the same for all subjects. The probability of reporting an illusory correlation would not be inflated by the opportunity to give more associations. If results similar to the present results were obtained, one could assert with more certainty that the probability of reporting an illusory correlation depends upon the associative strength of the characteristic and symptom.

Inherent in the definition of illusory correlation is the assumption that judges who report "No Relationship" do not see a characteristic-symptom relationship in the material and thus are not susceptible
to the illusory phenomenon. This implies that judges who give a large number of "No Relationship" responses should present more of the positive qualities associated with the personality-ability measures. However, the results presented in Table 7 do not support such an hypothesis. No significant differences are evident between high and low scorers on these personality-ability measures and the number of "No Relationship" responses given.

The variation in associative strengths of the illusory popular characteristics for each of their respective symptom statements (collectively identified as the Clinician Scale or the Student Scale) suggests that Number of Reported Populars may not represent an homogenous index of measurement. If this is indeed the case, the negative findings of the majority of the personality measures may not be a correct representation of the true case. In order to correct for this, in future research utilizing Number of Reported Populars on a DAP task such as the one employed here, it is recommended that only the four strongest of the six symptom statements be utilized. In addition, experimentation might be undertaken to arrange a method of more objective scoring such as a multiple selection technique rather than an open-ended questionnaire.

Future research might also examine the relationship between the operational definitions offered by the Chapmans on the DAP, on the Rorschach, and by Starr and Katkin (1969) on the Incomplete Sentences Blank. It is recommended that the investigation of the illusory correlation phenomenon be extended to other clinical interpretative tasks such as the
Thematic Apperception Test (and other TAT-type tests), to qualitative interpretations of WAIS responses, and to interview material itself.

Finally, it seems more research is needed like that of Golding and Rorer (1972) in investigating the resistance of illusory correlation to corrective training attempts. Contrary to the expectations of this study, the present results add little new information on the important question of the personality characteristics of the effective clinician. The answer may lie more in training than in personality differences. Perhaps good clinicians are made, not born.
Summary

This study has investigated the relationship of various personality-ability measures (field dependence, creativity, and social intelligence) and the phenomenon of illusory correlation in the context of clinical judgment. Chapman and Chapman (1967, 1969) have defined and investigated illusory correlation in the clinical judgment of projective test protocols. Illusory correlation has been defined as the tendency to report characteristic-symptom relationships in diagnostic material which in fact do not exist in that material or which exist to a lesser extent than that reported. The report of an illusory correlation is thought to depend upon the characteristics of the stimulus material and upon the perceptive qualities of the judge.

The present study attempted to differentiate judges according to the number of illusory correlates they reported -- that is, judges highly susceptible to the illusory phenomenon and judges less susceptible to the illusory phenomenon -- and to see if high and low reporters of this phenomenon differed on five personality-ability measures and two task-derived measures. It was hypothesized that low reporters of this illusory phenomenon would score higher on these personality measures than would high reporters of illusory correlation.

DAP drawings were viewed by 70 undergraduate, male subjects who were then divided into groups on the basis of the number of times they reported the illusory phenomenon in a subsequent questionnaire. These subjects were then given the following measures: Hidden Figures Test (field dependence), Expression Grouping and Cartoon Predictions (social
intelligence), and Alternate Uses and Similarities (creativity). The amount of confidence with which a judge gave a response (Average Confidence Per Response) and the number of responses given for each symptom statement (Average Number of Responses Per Item) were determined from responses on the questionnaires.

The original hypotheses were not supported. Trend indications in the direction opposite to that hypothesized were found for Alternate Uses (creative productivity) and Average Number of Responses Per Item. No other differences were significant. Further investigation revealed significant differences between the scores of infrequent reporters and more frequent reporters ($p < .005$, $p < .025$; both one-tailed) for Expression Grouping, a measure of social intelligence. Once again, Average Number of Responses Per Item was the only other measure with at least a trend indication. The stability of reports of the illusory phenomenon was also determined. With test-retest reliabilities of 0.65 and 0.70 having been obtained for two different methods of scoring. Implications of these essentially negative results were discussed, and it was suggested that training rather than personality differences may be of more importance in determining whether judges are susceptible to illusory correlation. Suggestions were also made for future research, including the development of other methods of self-report in the questionnaire, the investigation of illusory correlation on other diagnostic tasks, and the synthesis of present operational definitions of illusory correlation.
References


Little, K. B., & Shneidman, E. S. Congruencies among interpretations of psychological test and anamnestic data. Psychological Monographs, 1959, 73 (Whole No. 476).


Walker, R. E. Social intelligence: A psychological concept with a long history, limited development, but promise as a late bloomer. Unpublished manuscript, Loyola University, 1971.


The thesis submitted by Robert J. Lueger has been read and approved by the following Committee:

Dr. Thomas P. Petzel  
Assistant Professor, Psychology, Loyola

Dr. Ronald E. Walker  
Professor, Psychology and  
Dean, Arts and Sciences, Loyola

The final copies have been examined by the director of the thesis and the signature which appears below verifies the fact that any necessary changes have been incorporated and that the thesis is now given final approval by the Committee with reference to content and form.

The thesis is therefore accepted in partial fulfillment of the requirements for the degree of Master of Arts.

8/20/73  
Date

[Signature]  
Director's Signature