ASSUMED SIMILARITY AND ACCURACY IN OBSERVING PEOPLE

Thesis for the Degree of M. A.
MICHIGAN STATE UNIVERSITY
ROBERT STUART FORSYTHE
1970

ABSTRACT

ASSUMED SIMILARITY AND ACCURACY
IN OBSERVING PEOPLE

By

Robert Stuart Forsythe

The purpose of this study was to investigate how accuracy in observing people is affected by differences in sex, scholastic ability, and assumed similarity. Assumed similarity (empathy) is defined as the extent to which one perceives another's personality as being similar to his own. The four hypotheses tested were these:

1. Accuracy in observation is unrelated to academic skills.
2. Women are more accurate observers than men.
3. People who empathize highly with a target person are more accurate observers of that person than those who are low empathizers.
4. A person observes more accurately those people with whom he has high empathy; he observes less accurately those people with whom he has less empathy.

The Cline color film interviews with three women were used to present the visual stimulus. The subjects were 147 undergraduates attending Michigan State University. After viewing the film, subjects estimated how similar they

were to each interviewee in terms of attitudes, beliefs, values, and personality traits. They then answered questions about the appearance of the three females interviewed.

The results indicated that academic ability, as reflected by course grade, has no effect on one's accuracy in observation. The second hypothesis was also strongly supported. Women were significantly better observers than men. Hypothesis 3 was not supported. Differences in assumed similarity across subjects were uncorrelated with differences in observational accuracy. The results did provide some confirmation of Hypothesis 4, however. Within subjects, differences in assumed similarity for the three interviewees were associated with corresponding differences in observational accuracy.

The low reliabilities of the three subtests on observation and the unknown reliability of the empathy scale are the major limitations of this study, affecting particularly the testing of the last two hypotheses.

Approved Henry C. Smith

Date February 10, 1970

ASSUMED SIMILARITY AND ACCURACY IN OBSERVING PEOPLE

By

Robert Stuart Forsythe

A THESIS

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

MASTER OF ARTS

Department of Psychology

1970

ACKNOWLEDGMENTS

I would like to express sincere appreciation to Dr. Henry Clay Smith. His encouragement, patience, and invaluable advice made the writing of this thesis both enjoyable and instructive. I would also like to thank Dr. Donald Johnson and Dr. Paul Bakan for their assistance and helpful suggestions.

TABLE OF CONTENTS

	Page
INTRODUCTION	1
HISTORY	3
The Measurement of Assumed Similarity	4
The Measurement of Observational Accuracy	6
Factors Affecting Accuracy in Observation	8
The Effect of Actual Similarity on Observation	10
The Effect of Assumed Similarity on Accuracy in Informal Observation	14
PROBLEM	16
METHOD	18
Subjects	18
Development of the Assumed Similarity Scale	18
Development of the Observational Accuracy Test	20
Procedure	22
RESULTS	23
DISCUSSION AND SUMMARY	27
BIBLIOGRAPHY	32
APPENDICES	34

LIST OF TABLES

Table		Page
1.	Mean observation scores for subjects who ranked each woman first, second, or third on empathy	15
2.	Mean total accuracy scores for males and females	24
3.	Correlations between accuracy scores and empathy scores for males and females	24
4.	Frequency of positively correlated and negatively correlated orderings of observation and empathy scores	26

INTRODUCTION

Since Johannes Muller in 1826 first set forth his notions concerning the function of the nerves in relation to the Sensorium, a number of theorists and researchers have concerned themselves with the problem of why people perceive things as they do. The locus of investigation has moved rapidly from the armchair to the laboratory, and the focus has shifted from perceptual aspects that were universally true for all individuals to phenomena in which individual differences in perception occur. One such phenomenon that has been studied is that of selective attention. Given a set of potential stimuli, different people may selectively attend to different subsets of the whole (Bakan and Leckart, 1966).

It is the purpose of this paper to investigate the effects of certain variables on the perception of people. Although the effects of sex and mental skills will be studied, the main thrust of investigation is to determine the relationship between assumed similarity and observational accuracy. Assumed similarity, or empathy, is defined as the degree to which an individual X perceives another's (Y's) personality as being similar to his own. Observational

accuracy is the degree to which X remembers visual details about Y. It is hypothesized that the greater X's assumed similarity to Y, the greater X's observational accuracy for Y. The nature of the relationship among sex, intelligence, empathy and accuracy of observation has not yet been firmly established. Hopefully, the present study helps to define this relationship.

HISTORY

Although the instruments used to measure empathy and observational accuracy are unique to the present study, the theoretical problem investigated represents an extension of research begun by Levine and Murphy in 1943. In that classic experiment, they examined the effects of having different political ideologies on the learning of messages congruent or incongruent with these beliefs. Since then, several experimenters have studied how attention and retention are affected by differences in actual similarity between observer and observed. The relationship of observational accuracy to personality and demographic variables has also been investigated, and with the introduction of Cline's filmed interviews in 1955, attempts were begun to develop a standardized instrument for measuring skills in observing these interviews. Researchers in other areas were concurrently exploring the concept of empathy, or assumed, rather than actual similarity between people. In 1968 Smith formally questioned how empathy affects accuracy in observation, which led to the present investigation.

The review of the literature will first familiarize the reader with the concepts and measurement of assumed similarity and observational accuracy, and then present

findings on the effects of sex, scholastic achievement, and certain personality traits on accuracy in observation. The general historical development of the two major hypotheses will follow, beginning with the study by Levine and Murphy, and concluding with Smith's investigation.

The Measurement of Assumed Similarity

Smith (1966) has defined empathy or assumed similarity as "the tendency of a perceiver to assume that another person's feelings, thoughts, and behavior are similar to his own." He feels that the desire to "find out how much and in what ways a person is similar to or different from ourselves" is an important motive in forming impressions of people. (Smith, 1968)

The role of similarity has received considerable attention in the field of psychology, but most of the research concerns actual similarity between people, rather than assumed similarity. The distinction between these two concepts is important. The former implies that an actual correspondence exists between two individuals, whereas the latter implies that one is making an assumption about this correspondence, and this assumption may or may not be accurate. Although some studies have indicated that actual similarity is positively related to assumed similarity (Silkiner, 1962), the latter concept is the individual's subjective evaluation and would therefore be more

psychologically operative in influencing any other subjective perceptions.

The standard method for measuring degree of empathy is to present a subject with a number of statements describing specific behaviors or personality traits. The subject is instructed to indicate which items apply to himself and which apply to some specified other person. The total number of items the subject answers identically for both himself and the other person constitute an estimate of assumed similarity.

"Everyone assumes some similarity between himself and others, but some assume more than others." (Smith, 1966) In 1960, Chance and Meaders studied personality differences between high and low empathizers. They had each subject listen to a pair of interviews in which the two interviewees had been rated as opposites in either Autonomy or Succorance needs using scores from the Edwards Personal Preference Scale. After listening to the interviews, subjects filled out the EPPS as they thought the two interviewees had. These answers were compared with the subject's own responses to the EPPS, and the two corresponding estimates of assumed similarity were calculated for each subject. Chance and Meaders found that very high empathizers had significantly higher scores on the needs Deference, Affiliation, Succorance, and Order than very low empathizers. They also had much lower scores on the needs Exhibition,

Autonomy, and Dominance. Even though these differences were statistically significant, the use of extreme groups limits the generalization of results.

The Measurement of Observational Accuracy

Accuracy of observation will be defined as the degree to which a person can recall visual stimuli. Obviously, one also "observes" sounds, smells, and tactile sensations, as well as sights. For the purpose of this investigation, only visual perceptions will be studied. Also, although observation is a complex sequential process, accuracy will be estimated by measuring only the degree of recall. Thus, for a subject to be categorized as an accurate observer, he must attend to, perceive, and be able to recall certain visual stimuli.

Harris, in 1962, constructed the first standardized film test for estimating accuracy in observing people. Using the six film interviews developed by Cline in 1955 as the visual stimulus, he constructed thirty true-false descriptive statements about each of the six interviewees and had seventy subjects, mostly male, respond to each test immediately after the corresponding film was over. The statements were about each interviewee's conversation and appearance, and therefore tapped the modality of hearing in addition to sight. For all 180 items, the estimated internal consistency was .67.

Bruni, in 1963, used the same films as Harris, and developed a shorter and slightly less reliable observation test. Since three of the interviews were with males and three with females, he constructed a separate 60 item test for each sex. Half of these sixty items concerned conversation, and half were statements about appearance. Statements were chosen so that each item would be descriptive of only one of the filmed interviewees. Subjects were instructed that they would be required to match each statement with the appropriate interviewee and that some of the statements would not apply to anyone seen in the films. For all six interviewees, the sixty four-choice items on appearance had an odd-even reliability coefficient of .49, while the items on conversation had a reliability of .73. The reliability of the test on observation of males was .68, and the coefficient for the female test was .57.

Both of the above instruments informed subjects of the nature of the visual task before they viewed the films. Scores on these tests indicate differential accuracy in formal observation. Smith (1968) developed a measure of the accuracy of informal observation by not pre-instructing subjects that they would be tested on visual observation. Instead, they were informed that they were to evaluate each interviewee's personality traits, behavior, and attitudes. After viewing the films, subjects responded to 48 true-false

statements about aspects of each interviewee's appearance. The instrument in the present study is a revised version of this test.

Factors Affecting Accuracy in Observation

Unlike empathy, which Chance and Meaders found to be related to several personality variables, accuracy in formal observation seems to be unaffected by personality differences. Scores on Bruni's sixty item test of observation were correlated with scores on twenty-two personality scales developed by Smith. (See Hershey, 1958) Although the scales are very reliable, none of the correlations with observational accuracy scores were significant, and ranged from -.14 to +.08. Bruni also reported that scores on such variables as stereotype accuracy, empathic drive, consideration, initiation of structure, and psychological distance were uncorrelated with observational accuracy. Harris (1962), using a very similar design, also reported no relation with scores on total stereotype accuracy or empathic drive.

Several studies have examined general intelligence as a possible correlate of observational accuracy. Harris found that scores on his observation test were correlated .39 with course grade of the subject. This finding is inconsistent with the results reported by Bruni, and Witryol and Kaess (1957). The former discovered no significant

correlation between accuracy scores and course grade, grade point average, or scores on the College Entrance Examination. Witryol and Kaess investigated the relation of differences in intelligence and sex to performance on three social memory tasks. Two of these tasks, the George Washington University Memory for Names and Faces, and the KW Memory for Names and Faces, instructed subjects to memorize the names printed under several picture-portraits of individuals. Later, they had to correctly match the names and pictures. The third memory task involved assigning different aliases to a small group of persons and then later having each person match the alias with the person to whom it had been given. Although females scored significantly higher than males on all three tasks, a modified Alpha Exam given to subjects revealed no differences in intelligence scores between males and females. The correlation of these scores with scores on the KW Memory Test was .01 and -.02 for males and females, respectively. These results support the contention by Bruni that general intelligence is unrelated to accuracy in observation. They also present consistent and fairly conclusive evidence that sex does play an important role in observation. Witryol and Kaess explain this phenomenon results from women being required to learn greater social facility than men as part of their female role.

The Effect of Actual Similarity on Observation

Although most of the literature in this area is concerned with the effects of actual similarity on formal observation, reviewing these investigations should prove helpful in understanding how the present study came about. In their classic study, Levine and Murphy (1943) had college students who proclaimed themselves to be either procommunist or anticommunist in their thinking undergo a learning task in which each subject listened to both a very anticommunist passage and a procommunist passage. Each essay contained a number of distinct ideas, and subjects were instructed to try to reproduce both passages from memory. Although there were several trials spaced over a four week period, the results of the first immediate recall session seem most relevant to the present study. Procommunist students recalled significantly more ideas from the procommunist passage than anticommunist students, and anticommunist students recalled more ideas from the anticommunist passage than procommunist students. Two aspects of this study which may limit the validity of their findings are the small sample size used (n=10) and the fact that the intent of the experiment was not kept hidden from the subjects. Taft, in 1954, employed a somewhat similar paradigm. He read a passage to thirty Negro youths and thirty white youths which contained ideas rated as favorable (to Negroes), unfavorable, ambiguous, or neutral. Upon immediate recall, Negroes

remembered significantly more ideas than whites, both favorable and unfavorable. Negroes also remembered more favorable items than unfavorable items, but in direct proportion to the frequency of their occurrence in the passages. In contrast to Levine and Murphy's results, these findings indicate no tendency for attitudes to differentially affect the recall of favorable and unfavorable statements.

In 1966, Waly and Cook undertook three separate experiments in which they had college students attempt to learn verbatim twelve statements, six of which expressed prosegregationist opinions, and six, antisegregationist opinions. On the basis of responses to written attitude questionnaires, subjects were similarly classified as being either prosegregationist, neutral, or antisegregationist. The first study supported the findings of Levine and Murphy. Using undergraduates from a few Southern border colleges, they found that by the fifth learning trial, antisegregationists had learned antisegregationist statements better than prosegregationist statements. The opposite was true for the prosegregationist students. The other two studies used a similar design, but selected subjects from colleges in the North and in the South. These latter findings revealed no significant effect of attitudes by statements, and supported the authors' contention that Levine and Murphy's findings resulted in part from factors other than differences in attitude.

The next two studies investigated the relationship between observer-observed congruity and observer viewing time. Thirty-six male neuropsychiatric patients served as subjects in Christiansen's 1961 study. Nurses were trained to rate patients on their tendency to approach or avoid women (the nurses). Each of these patients was asked to view sixty slides at whatever pace he chose, under the guise that he would be helping to select pictures for other patients to see. The experimenter unobtrusively recorded the amount of time spent viewing each slide. Thirty of the slides had neutral content. Thirty others were classified as possessing heterosexual content: fourteen of these represented affiliative social contacts, and sixteen portrayed physical sexual contacts. Using the time spent viewing the neutral slides as a baseline, the ratio of time spent viewing heterosexual slides to neutral slides correlated .50 with the nurses' ratings on approach-avoidance to women. Patients rated as avoiding heterosexual contact spent relatively less time viewing slides with heterosexual content than did patients rated as approaching this type of contact.

Bakan and Leckart in 1966 administered the Maudsley Personality Inventory to undergraduates and on the basis of the extroversion scores, classified sixty students as extroverts, ambiverts, or introverts. These subjects examined pictures whose contents had been similarly

categorized by judges as extrovert, ambivert, or introvert. An observer recorded the time subjects spent examining each picture. The experimenters hypothesized that "normal S's will tend to select or attend to stimuli which are congruent with their personality characteristics." The results were not clearcut. Extroverts tended to look longer at extroverted pictures than at introverted pictures. Introverts looked at both groups of pictures equally long. However, extroverts, ambiverts, and introverts looked longest at ambiverted pictures. The researchers explained this occurrence on the basis that the ambivert pictures had more complex stimulus content, which in itself would elicit a longer looking time.

Kagan's 1967 investigation incorporated different ideas from many of the above studies, and in doing so more nearly approaches the design of the present study. He classified 56 Radcliffe students as either Academic or Social Types using consensual ratings by peers. The Academic Types were involved with studies much more than with social activities, and the opposite was true for the Social Types. Subjects were instructed to decide which of two poems were the more creative by listening to a tape recording of each poem. One of the poets was described as being a "social activist" and the other an "academic grind." When asked afterward to reproduce both poems verbatim, Social Types recalled more of the poem when told it was

written by the social activist rather than the academic grind. The reverse was true for the Academic Types. Although the design of the experiment was amenable to statistical analysis, no details were given concerning any significance testing, score distributions, or the measurement of the dependent variable. For this reason, one can say nothing about the strength of the relationship, the size of the absolute differences between groups, or errors in the measuring technique. Nevertheless, Kagan's study is an important stage in the development of the present investigation.

The Effect of Assumed Similarity on Accuracy in Informal Observation

The investigation by Smith in 1968 was the first to examine the relationship between assumed similarity and informal observational accuracy. He had 121 students in a sophomore level psychology course observe the Cline filmed interviews of the three women, D, N, and P. They were instructed that they were to note their similarity to the interviewees. Following the presentation of all three interviews, subjects ranked D, N, and P on empathy, assigning a "1" to the woman they empathized most with, a "2" to the woman they empathized next best with, and so forth. They were then asked to respond to a 48 item true-false

observation test over what they had seen during the interviews. The instruments used are found in Appendix A.

In tabulating the data, Smith examined the mean accuracy scores for those subjects who ranked each interviewee first, second, and third. This data is presented in Table 1.

Table 1.

Mean observation scores for subjects who ranked each woman first, second, or third on empathy.

		Ranking by subject		
		First	Second	Third
	D	9.66	9.07	8.70
Woman	N	9.72	8.53	8.49
	P	10.32	10.39	10.23

In the cases of D and N, the relationship between observation scores and rank appears slightly positive. However, the number of entries varies so much across cells that an analysis of variance is not appropriate for statistical analysis. Furthermore, the low reliability of the subtests makes doubtful any conclusions of systematic variance. The present study was intended as a refinement of Smith's original investigation, with emphasis on the development of a more reliable observation test and the use of appropriate statistical analysis.

PROBLEM

The purpose of this investigation is to identify some of the factors related to informal observational accuracy. Although the roles of sex and intelligence in formal observation were examined by Witryol and Kaess, their relation to informal observation is uncertain. It is this writer's contention that general mental skills play an insignificant role in informal observation, and that other skills are more relevant. As suggested by Witryol and Kaess, certain aspects of the female role seem to provide her with observational skills that are superior to those of men. The first two hypotheses concern these problems.

1. Accuracy in observation is unrelated to academic skills.
2. Women are more accurate observers than men.

The two major hypotheses of this study concern the effect of empathy on observational accuracy, but investigate two different phenomena. The first proposition examines differences in empathy across observers or how observers empathize differently with the same target person. Some will assume more similarity to this person, while

others will assume much less. Hypothesis 3 postulates a difference in observational accuracy for these groups of people.

3. People who empathize highly with a target person are more accurate observers of that person than those who are low empathizers.

The second phenomenon deals with differences within an observer, which may be explained as the differences in empathy one person feels for several others. He may assume a great deal of similarity to one person, but very little similarity to another. How is the observer influenced by different feelings of empathy for different people? The final hypothesis proposes an answer to this question.

4. A person observes more accurately those people with whom he has high empathy; he observes less accurately those people with whom he has less empathy.

METHOD

In this section the subjects and general procedures used will be described. Since the instruments used to measure both empathy and accuracy of observation were developed in the course of the investigation, they will be discussed in detail.

Subjects

Data was collected February, 1969, from 174 students attending a regular class meeting of a sophomore level course in psychology at Michigan State University. Of those students present, 147 completed all items on the two scales. Data from these 68 males and 79 females was used in the analysis.

Development of the Assumed Similarity Scale

All but one of the studies discussed earlier looked at specific attitudes or personality characteristics in determining congruence between subject and observed stimuli. In all cases, the judgment of actual similarity or congruence was made by the experimenter. Such a procedure may be appropriate for stimuli such as written passages, in

which content can be controlled. However, when the stimulus is another human being, he presents quite a complex picture, and any specific categories imposed by the investigator may vary in appropriateness. The fact that a target person is anticommunist may be a strong evaluative point for one observer, but another observer might weigh more heavily intelligence and friendliness in making assumptions about similarity.

Smith's study in 1968 overcame this difficulty. Subjects were instructed to rank the three interviewees on similarity in "personality traits, and in her religious, social, and political behavior and attitudes." (See Appendix A) Although the method of ranking tends to eliminate the problems of level, spread, and central tendency inherent in rating methods, it provides the experimenter with little indication of the extent to which the property ranked is present. One subject may have judged all three women as being very dissimilar to himself, and consequently, his rank of "1" would be assigned to the least dissimilar of the three. Another may have found all three to be very similar to himself, and his rank of "1," although equal numerically to the former's, would belie the psychological equivalence of the ranks. For this reason, a rating scale was employed.

The present study employed instructions similar to those used by Smith and instructed each subject, "...your

task will be to evaluate each woman as to how similar she is to you in her beliefs, attitudes, values, and personality." Subjects then indicated their empathy for each woman by checking which of seven Likert type alternatives best described their degree of assumed similarity. Response choices ranged from "unlike each other in most ways" to "like each other in most ways," and were weighted from one to seven, respectively. The use of these particular category names was influenced by Smith's statement that people desire to "find out . . . in what ways a person is similar to or different from ourselves." (1968) One major disadvantage in using a scale that previous researchers in the area have not used is that no estimates of the reliability of measurement are available.

Development of the Observational Accuracy Test

One of the major goals of this study was to improve the reliability of Smith's observation test. In his study, the test consisted of 48 true-false items. Items 1 through 16 were statements about D; 17 through 32 concerned N; and 33 through 48 concerned P. Subjects were instructed to answer "True" if a statement were true and "False" if either a statement were false or its factualness could not be determined from the film.

An estimate of the homogeneity of the 48 item test was provided by application of Kuder-Richardson Formula 20,

which yielded a reliability coefficient of zero. The three individual subtests had reliabilities of .22, .03, and -.10 for D, N, and P, respectively.

The first step in constructing a more reliable instrument was a decision to use three-choice items rather than true-false items. Increasing the number of choices from two to three reduces the error variance associated with guessing, and in turn increases the reliability. (Guilford, 1956) The second step involved an item analysis of the original 48 items to determine which discriminated between high and low scorers. Those items that were good discriminators were reworded to fit the new three choice structure; those that did not were eliminated. After the creation of other new items, the revised test contained fourteen statements about D, N, and P, for a total of 42 items. (See Appendix C)

Application of Kuder-Richardson Formula 20 yielded the following estimates of internal consistency for the new test:

For the 14 items on D, $r_{xx} = .37$

For the 14 items on N, $r_{xx} = .16$

For the 14 items on P, $r_{xx} = .54$

For the 42 items, $r_{xx} = .60$

Although these values represent considerable improvement over those reported for Smith, the reliabilities of the three subtests are low, especially the value for N.

The reliability of the total test, however, compares favorably with coefficients reported by Bruni (1963) and Harris (1962).

Procedure

Subjects were given the Impressions of Similarity-Dissimilarity scale immediately prior to the showing of the films. (Appendix B) The three Cline films of D, N, and P were then shown serially in that order. When the films were over, subjects completed the scale. They were then administered the Observational Accuracy Test and instructed to record their answers on a computer answer sheet, along with their sex and student identification numbers. Course grades were obtained at the end of the term.

RESULTS

In this section, the hypotheses will be restated and the data relevant to the evaluation of each hypothesis will be presented. Levels of significance will be reported only when beyond the .10 level.

Hypothesis 1. Accuracy in observation is unrelated to academic skills.

Final grades in the course were determined from the sum of the raw scores on the midterm and final exam. The Pearson product-moment correlations between observational accuracy and total exam scores were calculated for males and females and found to be -.04 and .12, respectively. For all subjects (n=147), the correlation was -.01. These near-zero correlations support the hypothesis that grades are unrelated to skills in observation.

Hypothesis 2. Women are more accurate observers than men.

This hypothesis was tested by comparing the mean score for women on observation to the mean score for men. The differences were highly significant and in the direction predicted. ($p<.001$) This data is presented in Table 2.

Table 2.

Mean total accuracy scores for males and females.

Males		Females			
$\bar{X}$	S_x	$\bar{X}$	S_x	T	p
22.93	4.66	25.48	3.68	5.68	<.001

Hypothesis 3. People who empathize highly with a target person are more accurate observers of that person than those who are low empathizers.

To test this hypothesis involving differences across subjects, Pearson product-moment correlations were calculated between scores on observational accuracy and assumed similarity for D, N, and P. The results did not confirm the hypothesis. None of the correlations obtained were large enough to be statistically significant. Data is presented in Table 3 below.

Table 3.

Correlations between accuracy scores and empathy scores for males and females.

	D	N	P
Males	-.02	-.15	.18
Females	.11	.19	-.07
All subjects	.07	.02	.05

Hypothesis 4. A person observes more accurately those people with whom he has high empathy; he observes less accurately those people with whom he has less empathy.

In testing this hypothesis, several data problems were present. Although one might assume that within-subject differences in empathy scores reflect psychological differences in the trait, this may not be true for differences in accuracy scores. Since there was no prior testing to equate the difficulties of the three tests, differences in scores may instead reflect differences in item difficulties. For this reason, observation scores were standardized to render the different observation scores comparable. If Hypothesis 4 is true, one would expect that each subject would have the highest observation score for the interviewee he rated as being most similar to himself. The next highest standard score would be for the interviewee he rated as being second in similarity, and the lowest score would be on the woman he rated lowest on empathy. Thus, if a subject has empathy scores of 7, 5, and 3 for P, D, and N, respectively, one would expect his standardized observation scores to be greatest for P, smaller for D, and smallest for N.

Each subject's three empathy ratings were arranged from highest to lowest, and the rank order of the corresponding observation scores was recorded. Two-way ties on empathy ratings were broken in a random fashion. Given the

ordering of empathy scores, there are six possible rank orderings for the accompanying observation scores. These are: 1-2-3, which represents a perfect positive correlation between the two sets of scores; 1-3-2 and 2-1-3, which represent a correlation of +.5; 2-3-1 and 3-1-2, which are correlated -.5 with the standard; and 3-2-1, which represents a perfect negative correlation. If within-subject differences in empathy and observation are unrelated, one would expect the rank order of observation scores to be equally distributed among the above six orderings. If one found the proportion of the first three orderings to be greater than .5, he would have evidence in support of Hypothesis 4. This is, in fact, the case, as Table 4 indicates. The proportion of positively correlated orderings was significantly greater than .5, the proportion expected by chance.

Table 4.

Frequency of positively correlated and negatively correlated orderings of observation and empathy scores.+

Order	Males		Females		All Subjects	
1-2-3	11		19		30	
1-3-2	16	39 (.58)	11	48 (.62)	27	87 (.60)
2-1-3	12		18		30	
2-3-1	10		9		19	
3-1-2	9	28 (.42)	10	29 (.38)	19	57 (.40)
3-2-1	9		10		19	
	z=1.10		z=2.11*		z=2.40**	

+ For these calculations, n=144. Three subjects who had identical empathy scores for D, N, and P were not used.
* $p<.02$.
** $p<.01$.

DISCUSSION AND SUMMARY

Female superiority in observation of people seems evident. Their scores on the accuracy test were significantly higher than the males' scores. The literature is fairly consistent in supporting this hypothesis (Witryol and Kaess, 1957) although an important distinction exists between their study and the present one. The nature of the observational task was made explicit in the above investigation, whereas it was concealed in the present one. Thus, to the explanation by Witryol and Kaess that women have developed better skills than men in observation, might be added the notion that some motivational component makes them more attentive in an informal setting.

Of the four conclusions, this is the strongest. The measurement of the independent variable, sex, should have perfect reliability and validity. The scores on the dependent variable, observational accuracy, possess an internal consistency of .60, which, although minimal, compares favorably with previous film measures. The overall effect was very significant, and improvements in the reliability of the observation instrument should enable the size of the effect to be more precisely measured. Perhaps

the weakest aspect of this conclusion is in trying to generalize observation of a filmed interview to observation of people in real life. The interviewees were three adult females, and therefore not representative of people in general. In using a filmed presentation instead of live people, external validity suffers. However, what is lost in reduced ability to generalize to real life situations seems to be compensated for by an increase in the ability to compare the data of different researchers using a standardized presentation.

The zero correlation between course grade and score on observation presents evidence contradictory to Harris's results, but supportive of those reported by Witryol and Kaess, and Bruni. Harris, using the same three filmed interviews as this study, found a correlation of .31 for all students between course grade and accuracy in observation. A year later, Bruni obtained a correlation of -.01 between the same two variables using a slightly modified measuring instrument. Witryol and Kaess reported that scores on a social memory task correlated zero with scores on a Modified Alpha Exam. Even if a slight relationship does exist, a general intelligence factor as measured by current tests of achievement and I.Q. probably accounts for only a very small proportion of the systematic variance within observation scores. As mentioned above, the effects may be differential in a natural versus formal setting.

The ability to memorize assigned class material may aid in memorization of visual details in the formal situation. It may not play a role in motivating a person to attend to these stimuli in an informal situation, however.

Since Hypotheses 3 and 4 are so intimately related to each other, they shall be discussed jointly. At first, it might seem contradictory that one was supported and the other not. Both involve a relationship between observational accuracy and assumed similarity involving identical measuring instruments, but the first postulate concerns differences across subjects while the second deals with differences within subjects. The first is a statement about the differential relationships of several observers to one target person, and the second concerns an individual's different observations for different target people. Because of the analysis used, one would expect differences across subjects to be more susceptible to measurement error than differences within subjects.

The measure of assumed similarity is intraordinal at best. That is, for a particular subject, two identical scores on this scale would represent equal amounts of the trait, and one score higher than another would indicate that the former trait is stronger. Thus, comparisons of the relative strengths of degrees of empathy for D, N, and P are possible within a subject.

Comparison across subjects poses a problem. One cannot say that two subjects having equal scale values possess equal amounts of the trait, nor can we be sure that if person X has a higher score than Y, he is more empathic than Y. The problem arises from the somewhat ambiguous choice of responses on the empathy scale. Although all scales of this nature possess some degree of semantic ambiguity, this writer feels that the present scale, possessing no estimates of reliability or validity, is probably a very crude measure of empathy. The failure of the experimenter to obtain estimates of the reliability of the scale is a major drawback in evaluating the results, and represents a strong recommendation for future research in the area.

The test of observational accuracy is another contributor to error variance. Both Hypotheses 3 and 4 must deal with the accuracy scores on D, N, and P, which possess reliabilities of .37, .16, and .54, respectively. These estimates are so low that it is surprising that a significant difference was obtained for Hypothesis 4. In light of this, one must be suspicious of the stability of these results and await an experimental replication before drawing any confident conclusions. As mentioned earlier, the effects of differences in item difficulty from one subtest to another are partly eliminated by standardizing scores, but only partly. Future research should attempt

to equate for difficulty of items and try to improve the reliability of the observation test.

In general, the results of this study indicate that females are superior to males in observing people. As suggested by Witryol and Kaess, their superiority may be due to the greater development of social skills demanded by the female role. The ability to observe others accurately seems to be fairly independent of general intelligence factors. Both of these conclusions are strongly supported in the literature and in the present study. The relationship of empathy to observational accuracy is not clearcut. The present study suggests that empathic differences within observers for various target people are associated with somewhat predictable observational differences. Those people we find ourselves more similar to, we remember more about. Those we find less similar do not receive the same degree of attention. There was no support given to the contention that observers who assume more similarity to a target person observe him better than observers who assume less similarity to that person.

BIBLIOGRAPHY

BIBLIOGRAPHY

1. Bakan, P. and Leckart, B. Attention, extraversion, and stimulus-personality congruence. Perception and Psychophysics, 1966, 1, 355-357.

2. Bruni, E. A film test of accuracy in observing people and its correlates. Unpublished Master's thesis, Michigan State University, 1963.

3. Chance, J. and Meaders, W. Needs and interpersonal perception. Journal of Personality, 1960, 28, 200-210.

4. Christiansen, K. C. Response duration as a measurement of ambivalent response tendencies. Journal of Personality, 1961, 29, 115-123.

5. Cline, V. and Richards, J. Accuracy of interpersonal perception: A general trait? Journal of Abnormal and Social Psychology, 1960, 60, 1-7.

6. Fiedler, F. Assumed similarity measures as predictors of team effectiveness. Journal of Abnormal and Social Psychology, 1954, 49, 381-388.

7. Grossman, B. A. The measurement and determinants of interpersonal sensitivity. Unpublished Master's thesis, Michigan State University, 1963.

8. Guilford, J. Fundamental Statistics in Psychology and Education. New York, McGraw-Hill, 1956.

9. Harris, W. A. The relation of observational to inferential accuracy in judging people. Unpublished Master's thesis, Michigan State University, 1962.

10. Hershey, G. College grades in relation to inventory measures of personality. Unpublished Master's thesis, Michigan State University, 1963.

11. Kagan, J. On the need for relativism. American Psychologist, 1967, 22, 131-141.

12. Levine, J. M. and Murphy, G. The learning and forgetting of controversial material. Journal of Abnormal and Social Psychology, 1943, 38, 507-517.

13. Mietus, J. R. Evaluation of a sensitivity training program with a component criterion. Unpublished Master's thesis, Michigan State University, 1969.

14. Mullin, J. Empathic drive and its influence on accuracy in judging people. Unpublished Master's thesis, Michigan State University, 1962.

15. Silkiner, D. A cross-cultural study of the measurement, determinants, and effects of stereotype accuracy. Unpublished Master's thesis, Michigan State University, 1962.

16. Smith, H. C. Bill Wilkins as a model for sensitivity training. Paper from the Symposium on Interpersonal Perception, American Psychological Association Convention, 1968.

17. Smith, H. Sensitivity to People. New York: McGraw-Hill, 1966.

18. Taft, R. Selective recall and memory distortion of favorable and unfavorable material. Journal of Abnormal and Social Psychology, 1954, 49, 23-28.

19. Waly, P. and Cook, S. W. Attitude as a determinant of learning and memory: A failure to confirm. Journal of Personality and Social Psychology, 1966, 4, 280-288.

20. Witryol, S. and Kaess, W. Sex differences in social memory tasks. Journal of Abnormal and Social Psychology, 1957, 54, 343-346.

APPENDIX A

H. C. Smith
November, 1968

A STUDY OF EMPATHY

This is an exercise in empathy i.e., your ability to assess your general similarity to other people. You will see five minute interviews with Mrs. D, Mrs. N, and Mrs. P.

Put your name and student number on the answer sheet. Then, in the lower half of the answer sheet write the names of the women as follows with a space after each:

RANK

Mrs. D. _____

Mrs. N. _____

Mrs. P. _____

After you have seen the film, put a "1" beside the woman whom you think would be most like you in her personality traits and in her religious, social, and political behavior and attitudes. Rank "2" the woman who would be next like you and "3" opposite the woman who you think would be <u>least</u> like you.

H. C. Smith
November, 1968

OBSERVATIONS OF MRS. D., MRS. N., AND MRS. P.

You have just seen film interviews with the three women. Below are statements about the appearance and behavior of each of them. Some are true and some false.

Record your answers both in the appropriate blanks below and on the separate answer sheet.

Mark "T" in the space below and "1" on the separate answer sheet if you think the statement is true.

Mark "F" in the space below and "2" on the separate answer sheet if you think the statement is false or if it cannot be answered from information in the film.

Answer all the questions for Mrs. D. before continuing with those for Mrs. N, and all those for Mrs. N before continuing with Mrs. P. Please answer all the questions even if you have to guess.

When you have finished, the answer sheets will be collected and the correct answers read and the significance of the task explained.

Mrs. D.

(T) 1. She had her purse on her lap.
(F) 2. She kept her hands in her lap.
(T) 3. She has a long thin neck.
(T) 4. She has high cheek bones.

(F) 5. She had a double chin.
(F) 6. She had thick lips.
(F) 7. She wore a wide leather belt.
(F) 8. She was wearing a black shirt.

(T) 9. She frowned often.
(F) 10. She wore a ribbon in her hair.
(T) 11. She moved her hands when talking.
(F) 12. She had pale cheeks.

(F) 13. Her hair was parted in the middle.
(T) 14. She wasn't wearing a wedding ring.
(T) 15. She had her hair cut in bangs in front.
(F) 16. She had a brooch on the left side of her sweater.

Mrs. N.

(F) 17. She had false teeth.
(F) 18. She didn't wear any lipstick.
(F) 19. She had a grey streak in her hair.
(T) 20. She blinked her left eye often.

(T) 21. She moved her hands often.
(F) 22. She wore heavy rouge on her cheeks.
(T) 23. She had on a black coat.
(T) 24. She had waves in her hair.

(T) 25. She carried a purse in her left hand.
(F) 26. She didn't smile throughout the interview.
(T) 27. The collar of her dress was unbuttoned.
(F) 28. She stuttered at times.

(T) 29. She wore earrings.
(T) 30. She held something in her left hand.
(F) 31. She wore a gold watch on her left wrist.
(F) 32. She had her ears pierced.

Mrs. P

(T) 33. Her hair was parted in the middle.
(F) 34. She wore a gold wedding ring.
(T) 35. The first and third buttons of her coat were undone.
(F) 36. She ran her finger along the edge of the table.

(T) 37. She wore earrings.
(F) 38. Her hands came into view three times.
(T) 39. She had black hair.
(F) 40. Her watch had a gold strap on it.

(F) 41. She had heavy eyebrows.
(T) 42. She giggled sometimes.
(F) 43. Her fingernails were polished red.
(T) 44. She had her collar turned down.

(F) 45. She had a rather low forehead.
(T) 46. Her hair was turned under on the ends.
(T) 47. She started many sentences with "oh" or "well."
(F) 48. She had gloves on.

APPENDIX B

February, 1969
RF/HCS

Student Number _____
Your initials _____
Sex _____

Impressions of Similarity-Dissimilarity

The purpose of the following exercise is to give you a chance to evaluate how similar you are to three different people. You will see filmed interviews of three women, Mrs. D, Mrs. N, and Mrs. P, in that order. During her interview, each woman will answer certain questions put to her by the interviewer. From these answers, your task will be to evaluate each woman as to how similar she is to you in her beliefs, attitudes, values, and personality.

Complete the following three statements by placing an "X" in the blank provided in front of the appropriate answer. Do not begin until all three interviews are over.

1. Mrs. D and I are
   _ like each other in most ways.
   _ like each other in many ways.
   _ like each other in more ways than we are unlike each other.
   _ like each other in as many ways as we are unlike each other.
   _ unlike each other in more ways than we are like each other.
   _ unlike each other in many ways.
   _ unlike each other in most ways.

2. Mrs. N and I are
   _ like each other in most ways.
   _ like each other in many ways.
   _ like each other in more ways than we are unlike each other.
   _ like each other in as many ways as we are unlike each other.
   _ unlike each other in more ways than we are like each other.
   _ unlike each other in many ways.
   _ unlike each other in most ways.

3. Mrs. P and I are
_ like each other in most ways.
_ like each other in many ways.
_ like each other in more ways than we are unlike each other.
_ like each other in as many ways as we are unlike each other.
_ unlike each other in more ways then we are like each other.
_ unlike each other in many ways.
_ unlike each other in most ways.

APPENDIX C

February, 1969
RF/HCS

A Study in Informal Observation

Please write your student number on the IBM answer sheet, and indicate your sex.

You have just finished viewing three filmed interviews of Mrs. D, Mrs. N, and Mrs. P, and have indicated how similar you are to each. Now, you will be given questions on what you saw during the film. There are 14 questions about Mrs. D, Mrs. N, and Mrs. P. Each question has three possible answers, but only one is correct. From memory of what you saw, mark on the IBM answer sheet the number of the correct answer for each question. If you are unsure of the correct choice, please guess.

Please turn the page and answer questions about Mrs. D first; then go on to Mrs. N; and finally Mrs. P.

MRS. D

1. She had (1) her purse on her lap.
   (2) a book on her lap.
   (3) neither a purse nor a book on her lap.
2. She had (1) a long thin neck.
   (2) a prominent Adam's apple.
   (3) a scar on her neck.
3. She had (1) a mole on her cheek.
   (2) full cheeks.
   (3) high cheek bones.
4. She (1) wore a wide leather belt.
   (2) wore a narrow brown belt.
   (3) may have worn a belt but her sweater concealed it.

5. She wore (1) red fingernail polish.
   (2) no fingernail polish.
   (3) gloves and her fingernails weren't visible.
6. She wore (1) a black sweater.
   (2) a red sweater.
   (3) a blue sweater.
7. She wore (1) a red ribbon in her hair.
   (2) a black ribbon in her hair.
   (3) no ribbon in her hair.
8. She wore (1) a gold wrist watch.
   (2) a chain bracelet on her wrist.
   (3) neither a watch nor a bracelet.

9. She wore (1) a small pin on the left side of her sweater.
   (2) a small pin on the right side of her sweater.
   (3) no pin on her sweater.
10. Her hair (1) had no visible part in it.
    (2) was parted in the middle.
    (3) was parted on the left.
11. She had (1) black hair.
    (2) dark brown hair.
    (3) light brown hair.
12. Her hair was (1) cut in bangs in front.
    (2) combed to one side in front
    (3) combed straight back from her forehead.

13. She (1) kept her hands in her lap during the interview.
    (2) occasionally moved her hands when talking.
    (3) used her hands frequently when talking.
14. She wore (1) a white pearl necklace.
    (2) a gold chain necklace.
    (3) no necklace.

MRS. N

15. She had
    (1) a grey streak in her hair.
    (2) predominantly grey hair.
    (3) no grey visible in her hair.
16. She wore
    (1) heavy rouge on her cheeks.
    (2) no rouge on her cheeks.
    (3) rouge and her left cheek seemed more pink than her right.

17. Her hair
    (1) was straight.
    (2) had small waves in it.
    (3) had large waves in it.

18. She wore
    (1) a hair net.
    (2) a brown hat.
    (3) neither a hair net nor a brown hat.
19. She
    (1) did not smile during the interview.
    (2) smiled, but only occasionally.
    (3) smiled most of the time.
20. She
    (1) wore a ring on her right hand.
    (2) wore a wedding ring on her left hand.
    (3) wore no rings.
21. She held
    (1) a purse in her hand.
    (2) gloves in her hand.
    (3) a purse and gloves in her hand.

22. She wore
    (1) a watch on her left wrist.
    (2) no watch on her left wrist.
    (3) a coat which obscured her left wrist.
23. She wore
    (1) earrings in the shape of a gold button.
    (2) earrings in the shape of three little balls.
    (3) no earrings.
24. She wore glasses
    (1) with reddish brown frames.
    (2) with gold frames.
    (3) with silver frames.
25. She had
    (1) turned up the collar of her coat in back.
    (2) turned up the collar of her coat in front.
    (3) buttoned her coat in front and the collar lay flat.

26. She had
    (1) a mole on her face to the left of her nose.
    (2) a mole at the right corner of her mouth.
    (3) no moles on her face.
27. She
    (1) adjusted her glasses once.
    (2) touched her left ear once.
    (3) touched neither her glasses nor her ear.
28. She had
    (1) deep wrinkles in her forehead
    (2) deep lines from her nose to the corners of her mouth
    (3) a relatively unwrinkled complexion.

MRS. P

29. Her hair (1) was parted in the middle.
(2) was parted on one side.
(3) had no part visible.

30. The (1) first and third buttons of her coat were undone.
(2) top button was the only one undone.
(3) top two buttons were undone.

31. She (1) ran her finger along the edge of the table.
(2) placed her fingers on the table occasionally.
(3) had her hands concealed from the camera.

32. She wore (1) gold earrings.
(2) no earrings.
(3) her hair so that it completely concealed her ears.

33. Her (1) watch had a gold band on it.
(2) watch had a leather strap on it.
(3) coat covered her wrist so that it was not visible.

34. She wore (1) a pearl necklace.
(2) a gold chain necklace.
(3) no necklace.

35. She had (1) very heavy eyebrows.
(2) thin dark eyebrows.
(3) a slight irregularity in her left eyebrow.

36. She had (1) a mole on her chin.
(2) a cleft in her chin.
(3) neither a mole nor a cleft in her chin.

37. Her hair (1) was turned under on the ends.
(2) hung straight at the ends.
(3) was flipped out on the ends.

38. She had (1) her collar turned down in front.
(2) her collar turned up in back.
(3) buttoned her coat at the collar.

39. She had (1) her hands clasped together most of the interview.
(2) raised her hands once.
(3) her hands out of the view of the camera.

40. The buttons on her coat were (1) red.
(2) gold.
(3) white.

41. She had (1) dark brown hair.
(2) black hair.
(3) light brown hair.

42. Her fingernails were (1) polished red.
(2) not polished.
(3) hidden from the camera.