1967, 72, 271-280

# ILLUSORY CORRELATION AS AN OBSTACLE TO THE USE OF VALID PSYCHODIAGNOSTIC SIGNS[1]

LOREN J. CHAPMAN AND JEAN P. CHAPMAN

*University of Wisconsin*

Practicing psychodiagnosticians ($N = 32$), when surveyed, failed to report observing Wheeler-Rorschach Signs 7 and 8 as accompanying male homosexuality although research evidence indicates that these are valid signs. They instead reported observing Wheeler Signs 4, 5, 16, 19, and 20, which research literature indicates are invalid. These popular invalid signs were found to have much stronger rated, verbal associative connections to male homosexuality than the unpopular valid signs. Six hundred and ninety-three undergraduates (divided among 13 conditions) viewed 30 Rorschach cards on each of which was arbitrarily designated a patient's response and his two symptoms. The *Ss* "rediscovered" the same invalid Rorschach content signs of homosexuality as the clinicians reported observing in their clinical practice, although these relationships were absent in the experimental materials. They did so regardless of the degree to which the clinically valid signs were valid in the contrived task materials.

The psychodiagnostician who uses psychological tests relies usually on his clinical experience for interpreting test responses. He accumulates observations of the different responses that occur as correlates of various personality characteristics of patients. He later uses this information to infer, at least tentatively, the presence of similar characteristics in patients giving similar responses. To do this, the diagnostician must assume, of course, that he is able to observe and remember which characteristics of test performance occur as correlates of each characteristic of personality. The inference of characteristics of personality from the occurrence of isolated characteristics of performance, instead of interpretation in terms of more complex patterns of test performance, is often termed the "sign" approach. This frequently deprecated, yet widely used, approach to projective test interpretation is the focus of the present study.

A large body of research literature demonstrates that many diagnostic test signs lack the psychological meanings that clinical observers have claimed for them. It is not surprising, then, that other studies have demonstrated that psychodiagnosticians are usually less able, than they believe, to make valid statements about patients on the basis of tests. Little and Shneidman (1959) made this point very clearly. In an especially well-controlled study, they found that eminent clinicians performed only slightly above chance.

The enormous discrepancy between the reports of clinical observers and the research evidence concerning the meanings of test responses has long been a puzzle. The conflict of evidence has been especially disquieting because clinicians usually show substantial consensus as to specific meanings of various test responses, although these same meanings may have been discredited by research evidence.

The present writers (Chapman & Chapman, 1967) have recently suggested a possible resolution of this enigma by demonstrating a source of massive systematic error in observations of correlations between symptom statements and features of projective test protocols. The projective test studied was the Draw-a-Person Test, which has been largely discredited as a measure of personality by the research evidence (Swensen, 1957). Naive undergraduates viewed a series of 45 Draw-a-Person Test drawings paired with contrived

[1] This study was supported by Research Grant MH-07987 from the National Institute of Mental Health, United States Public Health Service. The authors are indebted to Fred Cosentino for his assistance in the gathering and analysis of the data, and to Richard McFall for his critique of the manuscript.

[2] Requests for reprints should be directed to Loren J. Chapman, Department of Psychology, University of Wisconsin, Madison, Wisconsin.

TABLE 1

CRITICAL RATIO (z) VALUES FOR WHEELER SIGNS THAT DISTINGUISHED MALE HOMOSEXUALS FROM HETEROSEXUALS IN THREE STUDIES

| Study | z value | | | | | |
|---|---|---|---|---|---|---|
| | Sign 7 | Sign 8 | Sign 10 | Sign 17 | Sign 19 | Sign 20 |
| Wheeler | 1.64 | | | 1.79 | | |
| Davids, Joelson, & McArthur | 1.70 | 2.17 | 1.65 | | 1.80 | |
| Hooker | | 1.94 | | | | 1.90 |

symptom statements about the alleged patients who drew them. In these contrived materials there was no correlation between the occurrence of any symptom statement and any drawing characteristic. The undergraduate *Ss* "rediscovered" the same correlations between drawing characteristics and symptoms that a group of clinicians most often reported observing in their clinical practice. These relationships are called "illusory correlations" because the naive *Ss* reported observing them, although they were actually absent in the experimental materials. Moreover, the popularity of the various illusory correlations corresponded to strength of rated, verbal associative connection between symptom and drawing characteristic. These findings suggest that the popular meanings of many test signs, as reported by clinicians, are illusory correlations based on verbal associative connection of the test sign to the symptom, rather than on valid observations.

Not all tests, however, are as completely lacking in validity as the Draw-a-Person Test. Performance on most psychodiagnostic tests has some correlation with personality characteristics. In light of the research evidence indicating that clinicians are only moderately successful in the interpretation of psychodiagnostic test performance, one must wonder whether clinicians' observations of valid signs, when such signs are present, is impeded by their proclivity to observe, instead, associatively based illusory correlations. This question is the central interest of the present paper.

To investigate this issue, it was first necessary to choose a symptom and a test for which both valid and invalid signs have been reported by clinical observers. Male homosexuality and Rorschach content analysis were chosen because they appear to fulfill these criteria. Wheeler (1949) offered 20 Rorschach signs of male homosexuality. Clinicians commonly report several of these signs as substantiated by their own clinical experience, but research evidence strongly supports only two of the signs. Three studies by different investigators (Davids, Joelson, & McArthur, 1956; Hooker, 1958; Wheeler, 1949) have reported statistically interpretable evidence on the validity of all 20 Wheeler-Rorschach content signs, and these studies show some agreement. Table 1 presents the *z* values for those Wheeler signs that distinguished homosexual and heterosexual groups in one or more of the three studies at the .05 level using a one-tailed test. For Wheeler's study, the *z* values were computed by the present writers from Wheeler's published data, using the formula for the critical ratio between uncorrelated proportions. Hooker obtained her values using the comparable formula for correlated proportions, after matching pairs of *Ss* by total number of responses. Davids et al. did not indicate which formula they used.

As seen in Table 1, Wheeler Signs 7 and 8 were both found to distinguish homosexual from heterosexual groups in two of the three studies. Wheeler Sign 7 is a response on Card IV of "human or animal—contorted, monstrous, or threatening," and Wheeler Sign 8 is a response on Card V, W or Center D, of a "human, or humanized animal." Wheeler's example of "humans" is "woman dressed as a bat" which indicates that the humans are animalized humans. Signs 10, 17, 19, and 20 were each found to distinguish the groups in one study but were tested and not found valid in the other two studies. A finding by chance alone of a significant difference for one sign out of 20 is not unexpected. Therefore, for

purposes of this study, Signs 7 and 8 were considered the only clinically valid signs. This conclusion tends to be supported by Reitzell (1949) who reported Signs 7, 8, and 16 as the most discriminating of the 20 Wheeler signs. Unfortunately, her data cannot be analyzed statistically because she reported them in terms of number of responses of each type given by homosexual and nonhomosexual groups, rather than the number of *Ss* in each group who gave each sign.

The hypotheses of this study are:

1. The "popularity" of signs among practicing clinicians has little relationship to the objective clinical validity of the signs, as indicated by research evidence.

2. The most popular signs among practicing clinicians are the ones that have the strongest verbal associative connection to male homosexuality.

3. Naive observers, when presented with contrived Rorschach responses arbitrarily paired with statements of symptoms of the patient who gave each response, erroneously report observing that these same associatively based invalid signs occur as correlates of homosexuality.

4. The naive observers report these associatively based illusory correlations even when the materials are contrived so that other (clinically) valid correlations are present.

## PART I

### *Experiment I: Clinical Observations by Practicing Psychodiagnosticians*

It was first necessary to learn the kinds of Rorschach content that psychodiagnosticians believe they have observed most often in the protocols of male homosexuals. A questionnaire was prepared for circulation among practicing clinicians. The questionnaire was anonymous, but it asked the clinician to list his academic degrees and the years he obtained them, as well as the number of years of his psychodiagnostic experience.

#### *Instructions*

Although the more traditional clinical use of the Rorschach is the interpretation of Rorschach determinants, many practicing clinicians in recent years have observed that the content of Rorschach responses is also related to the patient's emotional problems. Some of these observed relationships have been discussed in published reports, while others have not. We wish to ask about your observations concerning the content of Rorschach responses by two kinds of patients.

1. Have you seen the Rorschach protocols of a number of men who have problems concerning homosexual impulses, either overt or covert?

2. What kinds of Rorschach content have you observed to be prominent in the Rorschach protocols of men with problems concerning homosexual impulses?

3. If possible, would you list some examples of the kinds of responses made by such men.

Altogether 76 copies of this questionnaire were sent to clinicians at internship training centers, academic psychology departments, and other leading clinical installations. Each of 11 recipients received several copies with the request that he circulate them to his colleagues.

#### *Results*

Questionnaires were returned by 42 clinicians 32 of whom said that they had seen the Rorschach protocols of a number of men with homosexual problems and were willing to list the kinds of Rorschach content they had observed. Of these 32 clinicians, 22 reported having the PhD degree and 10 the master's degree. They reported 2–29-yr. psychodiagnostic experience, with a mean of 9.1.

The clinicians' most frequent responses are listed in Table 2. Only two clinicians listed one of the two valid Wheeler signs. Both of these were Wheeler Sign 7. Most of the clinicians listed one or more of five invalid Wheeler signs. Two of these five signs have,

TABLE 2

THE FIVE WHEELER-RORSCHACH CONTENT SIGNS OF MALE HOMOSEXUALITY MOST FREQUENTLY REPORTED BY THE CLINICIANS

| Wheeler sign | Percentage of clinicians reporting the sign |
|---|---|
| 16 Human or animal anal content | 44 |
| 20 Feminine clothing | 38 |
| 4 Humans with sex confused | 28[a] |
| 5 Humans with sex uncertain | 16[a] |
| 19 Male or female genitalia | 38 |

[a] The percentages listed here are for the clinicians' reports of this content regardless of the card on which they observed it. Wheeler limited Signs 4 and 5 to Card III, but few of the clinicians specified a card. Wheeler described sexual confusion on Card I as Sign 2, but did not describe sexual uncertainty on that card despite the conceptual similarity of these two classes of percepts.

as previously discussed, each received support in only one of three studies, and the other three were supported by none of them.

In order to test the hypothesis that the popularity of the invalid signs of homosexuality is based on high-strength, verbal associative connection with homosexuality, ratings of strength of associative connection were obtained. The following rating item, used for "anality," illustrates the format that was used.

The tendency for "homosexuality" to call to mind "rectum" and "buttocks" is
a. Very strong.
b. Strong.
c. Moderate.
d. Slight.
e. Very slight.
f. No tendency at all.

Rating items were used for both of the two unpopular valid signs as well as for the five popular invalid signs that are listed in Table 2. The wordings of the signs were in some cases rephrased, as in the example above, to facilitate understanding by the raters. These wordings are listed in Table 3. Because of their conceptual similarity, "sexual uncertainty" and "sexual confusion" were represented in the ratings by a single item of "part man—part woman." In addition, there was one item each for "food" and "maps," which were used as filler items in the studies that follow.

The eight items concerned with homosexuality were intermixed with 80 other items in the same format, most of which were not concerned with sexual material. The questionnaire was given to a group of 34 undergraduate students who did not participate in the other studies reported in this paper.

The six associative ratings from a to f were assigned values from 6 to 1 and a mean was computed for each item. Table 3 reports the mean rated associative strength between homosexuality and each of the eight categories.

As seen in Table 3, the popular invalid signs have a much stronger associative connection to homosexuality than either the unpopular clinically valid signs or the two categories of filler items. There is no overlap between the mean rated strength of the popular invalid signs and that of the other signs and

TABLE 3

MEAN RATED STRENGTH OF ASSOCIATIVE CONNECTION BETWEEN HOMOSEXUALITY AND CONTENT AREAS

| Content area | Mean rated strength |
|---|---|
| Popular invalid signs | |
| Rectum and buttocks | 4.38 |
| Part man—part woman | 3.53 |
| Feminine clothing | 3.12 |
| Sexual organs | 4.47 |
| Unpopular valid signs | |
| Part animal—part human | 1.93 |
| Monsters | 1.68 |
| Filler items | |
| Food | 1.09 |
| Maps | 1.09 |

categories of filler items. The rated associative strength to homosexuality of the lowest rated popular invalid sign (feminine clothing) was compared with that of the highest rated unpopular valid sign (part animal—part human). The difference was significant as indicated by a two-tailed $t$ test ($t = 3.52$, $p < .001$). These data suggest that whether or not a clinician reports a given category of percepts as a correlate of homosexuality is determined primarily by the strength of its verbal associative connection to the symptom, rather than by objective reality. This inference is congruent with earlier findings (Chapman & Chapman, 1967) concerning the Draw-a-Person Test.

## PART II

Laboratory studies were designed to determine whether naive observers, presented with contrived statements of patients' symptoms and their Rorschach responses, would make the same errors of observation that the clinicians appear to have made in their observational reports. The demonstration in the laboratory of these systematic errors of observation based on verbal associative connection would lend strong additional support to the contention that the clinicians' reports reflect illusory correlation based on associative connection.

### *Subjects*

The *Ss* in the laboratory studies were tested under 13 conditions which, for convenience of exposition, were divided into three experiments (Experiments II, III, and IV). All *Ss* were students in an intro-

ductory psychology course. The number of *Ss* in Conditions 1–13 were 60, 43, 53, 37, 39, 52, 49, 59, 54, 61, 60, 60, and 66, respectively. No *S* served in more than one condition. All *Ss* were asked on a questionnaire if they had any familiarity or experience with principles of Rorschach interpretation. The occasional *S* who indicated any such familiarity was not included in the sample.

### *Experiment II (Conditions 1–5): Illusory Correlation in the Absence of Valid Relationships*

Experiment II was designed to determine if the invalid signs that were found to be popular with the clinicians would also be reported by naive observers when no valid relationship is present between any category of percepts and the symptom of male homosexuality. This study contained five conditions, each of which was designed to explain, on the basis of illusory correlation, the popularity of the report of one of the five popular invalid signs listed in Table 2.

#### *Method*

Clinical materials were fabricated to be shown to naive observers. The materials consisted of 30 Rorschach cards, on each of which one percept (or response) was paired with two statements of the emotional problems of the purported patient who was alleged to have given the response. The cards were covered with transparent plastic. Rorschach percepts were indicated by circling an area of the card and pasting on it a typed statement of the verbalization. For example, for one of the 30 Rorschach responses, the center area of Card V—or area D-7 (Beck, 1961)—was circled and labeled "Bugs Bunny." In a corner of the card appeared the statement:

The man who said this

1. has sexual feelings toward other men.
2. feels sad and depressed much of the time.

The 30 percepts were chosen so that six fell into each of five categories, which were:

A. One popular invalid sign
B. Wheeler Sign 7: Human or animal—contorted, monstrous, or threatening on Card IV (a clinically valid sign)
C. Wheeler Sign 8: Humanized animal or an animalized human on Card V (a clinically valid sign)
D. Geographic features (a filler category)
E. Food (a filler category)

There were five conditions in Experiment II, which differed only as to which invalid sign was used. In all other respects, both the procedure and materials in all five conditions were identical. The invalid signs for the five conditions were: Condition 1—human or animal with anal content; Condition 2—feminine clothing; Condition 3—humans on Cards I or III with sex confused; Condition 4—humans on Cards I or III with sex uncertain; and Condition 5—male or female genitalia.

The percepts used for each category are listed below. Locations of percepts are indicated using Beck's system. The percepts chosen for the seven Wheeler signs were Wheeler's (1949) examples, and others modeled after his examples.

*Wheeler Sign 7.* Human or animal, contorted monstrous or threatening. All were on Card IV, W: (*a*) a horrid beast, (*b*) Frankenstein, (*c*) a headless monster about to step on me, (*d*) man looking back through legs, (*e*) a giant with shrunken arms, and (*f*) a deformed man doing a back bend.

*Wheeler Sign 8.* Humanized animal or animalized human, on Card V. Four were W responses: (*a*) woman with butterfly wings, (*b*) an alligator in a fur coat, (*c*) man dressed like a bat, (*d*) a pigeon wearing mittens. Two were D-7: (*e*) a dog wearing clothes, and (*f*) Bugs Bunny.

*Human or animal anal content.* (*a*) Card I, Dd-22, rectum, (*b*) Card VII (inverted), Dd-25, anus, (*c*) Card IV, W, anal opening of man bending over, (*d*) Card VI, W (without D-3), horse's rear end, (*e*) Card IX, both D-1, woman's buttocks, and (*f*) Card II, D-4, anal opening.

*Feminine clothing.* (*a*) Card III, D-10, woman's high-heeled shoe, (*b*) Card II, D-2, woman's hat, (*c*) Card X, D-9, fur stole, (*d*) Card VIII, D-5, woman's laced corset, (*e*) Card III, D-7, woman's bra, and (*f*) Card VI, D-1 (inverted), woman's fur cape.

*Sexual confusion.* Three of these percepts were on Card I, D-4: (*a*) it looks like a woman but with broad shoulders, (*b*) upper part is male and the bottom is female, (*c*) this has shoulders like a man but breasts and hips like a woman. Three were on Card III, D-9: (*d*) looks like a man below the waist but like a woman above, (*e*) a man but with breasts, (*f*) a woman, here are her breasts, but has the feet of a man.

*Sexual uncertainty.* Three of these percepts were on Card I, D-4: (*a*) I guess this is a man or maybe it's a woman, (*b*) a person can't tell if it's a man or a woman, (*c*) a human, could be masculine or feminine. Three were on Card III, D-9: (*d*) a person, might be male or it might be female, (*e*) could either be a man or a woman, and (*f*) two people, but I can't tell what sex they are.

*Male or female genitalia.* (*a*) Card I, Dd-22, testicles, (*b*) Card II, D-3, female genital organ, (*c*) Card III, Dd-26, penis, (*d*) Card X, D-11, male genitalia, (*e*) Card VI, D-2, penis, and (*f*) Card III, D-8, vagina.

*Geographic features.* (*a*) Card VII (inverted), D-2, map of North and South America, (*b*) Card VII, D-3, map of Spain, (*c*) Card X, D-9, map of California, (*d*) Card II (inverted), D-2, map of South America, (*e*) Card X, D-13, New Zealand, and (*f*) Card VI (inverted), Dd-25, map of California.

*Food.* (*a*) Card VII, D-4, a loaf of rye bread split in half, (*b*) Card IX, D-6, raspberry sherbet,

(*c*) Card VIII, D-7, jello, (*d*) Card VI, D-6, asparagus tips, (*e*) Card X, D-2, scrambled eggs, and (*f*) Card III, D-1, bowl of fruit.

The two statements of emotional problems or symptoms listed on the cards were drawn from a pool of four such statements. These were:

1. He has sexual feelings towards other men.
2. He believes other people are plotting against him.
3. He feels sad and depressed much of the time.
4. He has strong feelings of inferiority.

The statements of symptoms and Rorschach percepts were paired on the 30 cards so that each of the four symptom statements appeared 15 times. Each symptom statement was paired with three of the six percepts from each of the five categories of percepts. Thus there was no relationship between the occurrence of any one of the four symptoms and any one of the five categories of response.

The prediction was that despite this lack of true correlation in the experimental materials, *S*s who viewed the series of 30 cards in each of the five conditions would believe that they observed that the symptom statement, "He has sexual feelings toward other men," had appeared more often with the associatively based invalid sign than with any of the other four categories of content.

The testing was done in groups of 30 or fewer. The *S*s were given some brief introductory information as to the nature of the Rorschach. They were told that Rorschach responses indicate personality functioning. However, they were given no information about categories of either content or determinants. They were then told:

> I am going to show you a series of inkblots, one at a time. On each inkblot you will find a typed statement of what one patient saw on this blot and also what his two chief emotional problems are. Each of these 30 cards represents a different patient. You will see what 30 different patients said they saw on a card. Now let me tell you what I want you to do. Please carefully study each inkblot and the statement of what the patient said that he saw in it. Also study the statement of the patient's two severe emotional problems. When everyone has looked at all of the cards, I'm going to give you a questionnaire in which I will ask you about the kinds of things seen by patients with each kind of problem.

The cards were circulated in a prearranged pattern so that each *S* saw each of the 30 cards for 60 sec. After *S* had seen the 30 cards, he was given a questionnaire which presented four items (one for each of the four symptom statements) in the following format.

> Some of the things in the inkblots were seen by men who have the following problem:
>
> *He has sexual feelings toward other men.*
>
> Did you notice any general kind of thing that was seen most often by men with this problem? Yes ___ No ___. If your answer is yes, name that kind of thing, and give one example of that kind of thing.
>
> Kind of thing _____
> Example _____

Two forms of the questionnaire were used with different orders of the four items.

## *Results*

Since the responses to the homosexual item were of primary interest in the present paper, the analysis of the results focused on this item. Table 4 shows the correlates reported for homosexuality by *S*s in each of the five conditions. As seen there, very few *S*s indicated that they could find no relationship between the percepts and the symptoms. In each condition, *S*s, as predicted, reported that they observed the hypothesized illusory correlate as accompanying homosexual problems more

TABLE 4

PERCENTAGE OF NAIVE OBSERVERS REPORTING EACH OF FIVE CATEGORIES OF PERCEPT AS MOST OFTEN ACCOMPANYING THE PROBLEM OF HOMOSEXUALITY FOR EACH OF THE FIVE CONDITIONS IN EXPERIMENT II

| Correlate reported | Condition | | | | |
|---|---|---|---|---|---|
| | 1 Anality | 2 Feminine clothing | 3 Sexual confusion | 4 Sexual uncertainty | 5 Genitalia |
| Predicted illusory correlate | 58 | 40 | 45 | 24 | 51 |
| Clinically valid signs | | | | | |
| Human or animal, monstrous | 8 | 14 | 13 | 10 | 3 |
| Part animal—part human | 3 | 14 | 6 | 8 | 0 |
| Filler categories | | | | | |
| Geographic features | 12 | 5 | 13 | 16 | 13 |
| Food | 8 | 16 | 15 | 18 | 13 |
| Other correlates | 0 | 0 | 2 | 10 | 8 |
| No correlate reported | 10 | 12 | 6 | 13 | 13 |

often than any other category of percept. Chi-square analysis indicated that the distributions of responses among the five categories departed from chance ($p < .01$) for each condition except Condition 4, in which the hypothesized illusory correlate was sexual uncertainty.

The frequent report of the invalid Wheeler signs as illusory correlates of homosexuality cannot be attributed to a higher frequency of report for all symptom statements. The mean percentage report of the five signs for the other three symptoms were as follows: anality, 4%; feminine clothing, 13%; sexual confusion, 9%; sexual uncertainty, 8%; and genitalia, 10%. In none of the five conditions was the illusory correlate of homosexuality the most frequently reported correlate of any one of the other three symptoms.

One may conclude from the data of Table 4 that the observation of popular invalid signs of male homosexuality were reproduced in the laboratory as illusory correlates. On the other hand, the clinically valid signs of homosexuality were not reported as illusory correlates of that symptom any more often than were the two filler categories.

### *Experiment III (Conditions 6–11): Illusory Correlation in the Presence of Valid Relationships*

Experiment III was undertaken to determine if the associatively based illusory correlates are reported by naive observers even if the clinically valid signs have contrived validity in the experimental task materials. To build in contrived validity, the clinically valid Wheeler Signs 7 and 8 were paired with the symptom statement of homosexuality more often than with the other classes of percepts.

#### *Method*

Only two of the five popular invalid signs were chosen to be used in Experiment III. They were anality and sexual confusion.

The task materials were almost identical to those used in Experiment II. The instructions and answer sheets were unchanged. The percepts, as well as the pool of symptom statements were the same as those previously used in Conditions 2 and 4 in which anality and sexual confusion were used as popular invalid signs. As in the earlier conditions, each *S* saw 30 cards, six presenting a percept of a popular invalid sign, six for each of the two clinically valid signs (Wheeler Signs 7 and 8), and six for each of the two filler categories (geography and food). The task materials differed from Experiment II only in the percentage of clinically valid percepts that were accompanied by the symptom statement of homosexuality. Three degrees of contrived validity were investigated: one in which the symptom statement of homosexuality occurred with two-thirds of the percepts of each of the two valid signs, one in which it occurred with five-sixths of them, and one in which it occurred with all of them. After being rounded off to the nearest percentage point the values of the three degrees of contrived validity were 67%, 83%, and 100%, respectively. The combination of three degrees of contrived validity with each of two popular invalid signs yielded six conditions altogether. In Conditions 6, 7, and 8, the popular invalid sign was sexual confusion, with contrived validity for Signs 7 and 8 of 67%, 83%, and 100%, respectively. In Conditions 9, 10, and 11, the popular invalid sign was anality with the same three degrees of contrived validity for Signs 7 and 8. As in Experiment II, 50% of each of the other three categories of percepts were accompanied by the symptom statement of homosexuality, as well as by the other three symptom statements.

#### *Results*

Table 5 shows the percentage of *S*s in each condition who reported the co-occurrence with homosexuality of each of the five categories of percepts. The most striking feature of these data is the degree to which the illusory correlates based on associative connection are impervious to the contrary influence of valid correlations. In all six conditions, the clinically popular invalid sign was the most frequently reported correlate of the symptom statement.

For each of the six conditions, the report of the homosexual symptom statement was distributed among the various categories of percepts significantly different from chance, as shown by chi-square analyses ($p < .01$ in every case). The number of *S*s who chose the invalid sign as a correlate was compared with the number who chose one or the other of the two valid signs. A goodness-of-fit chi-square was used with the expected values being one-third of the *S*s for the invalid sign, and two-thirds of the *S*s for the two valid signs combined. In five of the six comparisons the difference was significant ($p < .01$ in each case). The one nonsignificant difference was for anality when the clinically valid signs had 83% contrived validity. The difference for this condition fell short of significance ($\chi^2 = 3.03$, $df = 1$, $p < .10$). These findings

TABLE 5

PERCENTAGE OF *Ss* IN EXPERIMENT III REPORTING THE OBSERVATION OF EACH CATEGORY OF PERCEPT AS A CORRELATE OF HOMOSEXUALITY

| Category | Percentage contrived validity for Signs 7 and 8 | | |
|---|---|---|---|
| | 67% | 83% | 100% |
| **Conditions 6, 7, and 8** | | | |
| Invalid signs | | | |
| Sexual confusion | 50 | 41 | 46 |
| Valid signs | | | |
| Human or animal—monstrous | 12 | 27 | 20 |
| Part animal—part human | 10 | 6 | 14 |
| Filler categories | | | |
| Geographic features | 2 | 12 | 5 |
| Food | 8 | 0 | 2 |
| Other correlates | 10 | 6 | 8 |
| No correlate reported | 10 | 8 | 5 |
| **Conditions 9, 10, and 11** | | | |
| Invalid sign | | | |
| Anality | 54 | 34 | 55 |
| Valid signs | | | |
| Human or animal—monstrous | 16 | 16 | 15 |
| Part animal—part human | 13 | 23 | 13 |
| Filler categories | | | |
| Geographic features | 6 | 13 | 5 |
| Food | 2 | 3 | 2 |
| Other correlates | 0 | 3 | 2 |
| No correlate reported | 9 | 7 | 8 |

indicate that the invalid signs were reported as correlates of homosexuality more often than were the valid signs, which had objectively true correlations with the symptom.

Another way to view these data is in terms of a comparison of the number of *Ss* who reported the popular invalid signs, and the number who reported the valid signs, as correlates of homosexuality at each of the four levels of contrived validity of the two valid signs. (The four levels are the 50% co-occurrence of the clinically valid sign with the symptom statement—from Experiment II—plus the 67%, 83%, and 100% levels of contrived validity of Experiment III.) When the popular invalid sign was sexual confusion, these four conditions did not differ on frequency of report of the invalid sign ($\chi^2 = .84$, $df = 3$, $.80 < p < .90$). They also did not differ on number of *Ss* reporting one or the other of the two valid signs ($\chi^2 = 4.92$, $df = 3$, $.10 < p < .20$). When the popular invalid sign was anality, the four conditions of contrived validity differed in both the number of *Ss* reporting the popular invalid sign ($\chi^2 = 8.47$, $df = 3$, $p < .05$) and in the number reporting one or the other of the two valid signs ($\chi^2 = 13.05$, $df = 3$, $p < .01$). Inspection of the data for the three anality conditions indicates that *Ss* achieved highest accuracy by both criteria in the 83% validity condition. These findings indicate that increasing the contrived validity had a small effect in reducing the report of one popular invalid sign as a correlate of homosexuality, but not the other. The validity was accurately perceived to a slight degree.

Further analyses were performed to seek evidence on the accuracy of perception of the contrived validity. For the three conditions for which sexual confusion was the popular invalid sign, more *Ss* chose the valid sign than the filler items for the 100% validity condition ($\chi^2 = 9.38$, $df = 1$, $p < .01$) but this difference did not emerge on either the 67% or 83% validity conditions. For the three conditions for which anality was the popular invalid sign, *Ss* reported the valid signs more often than the filler categories for all three levels of contrived validity ($p < .05$ in all three cases). These findings again show that the contrived validity was to some degree accurately perceived, despite the fact that its presence did little to reduce the frequency of report of the illusory correlates.

### *Experiment IV: Accuracy of Report in the Absence of Popular Invalid Signs—Condition 12*

The most striking findings of Experiment III were the frequency of report of the illusory correlates, and the low accuracy of *Ss* observational reports of the valid signs. These findings might lead one to wonder if the infrequent detection of valid signs on these tasks is due to the distracting influence of illusory correlates or whether the valid signs are inherently difficult to discover. Further, one might wonder whether associatively based illusory correlation occurs only when the detection of valid signs is very difficult.

In defense of the tasks used in Experiment III, one might point out that they surely present the observer with a less difficult job

of information processing than does conventional clinical practice. In clinical practice, the symptoms are more numerous and more ambiguous, and many percepts are given by each patient. Also, the clinician encounters the patients over a long period of time, so that retrospective falsification of observations should occur more often. Nevertheless, it is of interest to determine if accuracy is higher on a task that is comparable to those of Experiment III, but in which associatively based popular invalid correlates are absent. The first part of Experiment IV (Condition 12) was designed to give evidence on this question.

*Method*

The stimulus materials were almost identical to those of the two 83% validity conditions (Conditions 7 and 10) of Experiment III. The symptom statement of homosexuality accompanied 83% of the percepts of the two clinically valid signs (Signs 7 and 8) and 50% of the other categories of percepts. The only change from Conditions 7 and 10 was that in Condition 12, nonsexual body parts were substituted for a popular invalid sign. The percepts for the category of nonsexual body parts were: (*a*) Card II, D-4, hands, (*b*) Card VII, D-1, gray hair, (*c*) Card IX, Dd-31, nose, (*d*) Card I, Dds-30, eyes, (*e*) Card II, D-2, foot, and (*f*) Card VI, Dd-25, toe. Thus, each *S* saw five classes of percepts: nonsexual body parts, humanized animals, monsters, food, and geographical features. The mode of administration of the task was identical to that of Experiment III.

*Results*

Table 6 shows the percentage of *S*s who reported each category of percept as a correlate of homosexuality. The distribution of *S*s among the five categories differed from chance ($\chi^2 = 26.23$, $df = 4$, $p < .001$). As seen in Table 6, the two clinically valid signs were the most frequently reported categories, and 65% of the *S*s reported one or the other of them. This value is almost double the percentage reported in the comparable conditions of Experiment III, and the rise in accuracy was significant ($\chi^2 = 11.6$, $df = 1$, $p < .001$).

*Condition 13*

The second part of Experiment IV was designed to eliminate one possible source of doubt concerning the meaning of the findings in Condition 12. One might speculate that the greater accuracy in Condition 12 than in Conditions 7 and 10 is not attributable to accurate observation of the valid correlates, but to illusory correlation. Table 3 showed that the clinically valid signs have a slightly stronger verbal associative connection to homosexuality than do the two filler categories. This associative connection might produce illusory correlation when stronger verbal associates are absent.

TABLE 6

PERCENTAGE OF *S*s REPORTING EACH CATEGORY OF PERCEPT AS A CORRELATE OF HOMOSEXUALITY IN THE TWO CONDITIONS OF EXPERIMENT IV

| Category | Percentage of *S*s reporting each category of percept | |
|---|---|---|
| | Condition 12[a] | Condition 13[b] |
| Clinically valid signs | | |
| Part animal—part human | 27 | 17 |
| Human or animal—monstrous | 38 | 18 |
| Filler categories | | |
| Body parts | 15 | 23 |
| Food | 3 | 24 |
| Geographic features | 8 | 9 |
| Other correlates | 2 | 4 |
| No correlate reported | 7 | 4 |

[a] Eighty-three percent contrived validity for Signs 7 and 8, 50% for the other three categories.
[b] Fifty per cent contrived validity for all five categories.

*Method*

In Condition 13 the percepts were identical to those used in Condition 12, but each class of percept appeared equally often with each symptom statement. If the increased accuracy of report of valid correlates in Condition 12 was an artifact of illusory correlation, it should remain in this condition. If the increased accuracy of report reflected increased accuracy of observation, it should disappear.

*Results*

The results appear in Table 6. The distribution of people reporting the five categories of percepts did not differ from chance ($\chi^2 = 5.17$, $df = 4$, $p > .05$), and the two clinically valid signs were reported at almost exactly the mean frequencies of the other three categories of correlates. This finding demonstrates that the increased accuracy of report in Condition 12 was due to increased accuracy of observation, and was not attributable to illusory correlation.

## DISCUSSION

There was a marked congruence between the reports of the clinicians as to their observations in clinical practice and the reports of the naive observers in the contrived experimental situations. Almost none of the 32 practicing psychodiagnosticians reported observing either of the two signs of male homosexuality that research findings indicate are clinically valid. Instead, they tended to agree with one another in reporting several signs that appear invalid in published research, but which have a high strength verbal associative connection to the symptom of homosexuality.

The naive observers presented with contrived clinical materials reported similar erroneous observations. They reported the associatively based signs as illusory correlates of homosexuality both when there were no valid relationships present and when the two clinically valid signs had a contrived validity in the task materials.

One of the most striking findings of these studies is the persistence of illusory correlation in the face of contradictory reality. Even in the two conditions in which all of the percepts of the two valid signs were paired with homosexuality, these two valid signs were reported as correlates of the symptom less often than the associatively based invalid signs. In addition, the low accuracy of report of the valid signs cannot be attributed entirely to there being too much information for the observer to process. Instead, the low accuracy is, to a considerable degree, a result of *Ss'* susceptibility to the illusory correlation. Experiment IV showed that, when the percepts that have a high associative connection to the symptom were not presented, the observers almost doubled their accuracy of report of valid correlates. The illusory correlates blind the observer to the presence of valid correlates of the symptom.

Associatively based illusory correlation is a powerful bias in the observational report of correlations between classes of events. Yet its influence is so unapparent that many practicing psychodiagnosticians have overlooked it, and have substituted illusory correlates for valid correlates in their diagnostic practice. One possible solution to this problem might be to demonstrate to graduate students, as part of their training, their own propensities toward illusory correlation.

## REFERENCES

BECK, S. *Rorschach's test.* Vol. 1. *Basic processes.* (3rd ed.) New York: Grune & Stratton, 1961.

CHAPMAN, L. J., & CHAPMAN, J. P. The genesis of popular but erroneous psychodiagnostic observations. *Journal of Abnormal Psychology,* 1967, 72, 193–204.

DAVIDS, A., JOELSON, M., & MCARTHUR, C. Rorschach and TAT indices of homosexuality in overt homosexuals, neurotics, and normal males. *Journal of Abnormal and Social Psychology,* 1956, 53, 161–172.

HOOKER, E. Male homosexuality in the Rorschach. *Journal of Projective Techniques,* 1958, 22, 33–54.

LITTLE, K. B., & SHNEIDMAN, E. S. Congruencies among interpretations of psychological test and anamnestic data. *Psychological Monographs,* 1959, 73(6, Whole No. 476).

REITZELL, J. M. A comparative study of hysterics, homosexuals and alcoholics using content analysis of Rorschach responses. *Rorschach Research Exchange,* 1949, 13, 127–141.

SWENSEN, C. H. Empirical evaluations of human figure drawings. *Psychological Bulletin,* 1957, 54, 431–466.

WHEELER, W. M. An analysis of Rorschach indices of male homosexuality. *Rorschach Research Exchange,* 1949, 13, 97–126.

(Received February 29, 1968)