Parapsychological Monographs No. 2

ESP in Relation to Rorschach Test Evaluation

GERTRUDE SCHMEIDLER
The City College, New York, N. Y.

PARAPSYCHOLOGY FOUNDATION, INC.
29 West 57th Street, New York 19, N. Y.

Parapsychological Monographs are published, unscheduled, by the Parapsychology Foundation, Inc., 29 West 57th Street, New York 19, N. Y. (USA), Eileen J. Garrett, President. The Parapsychology Foundation is a non-profit educational organization, established in 1951 to support impartial scientific inquiry into the total nature and working of the human mind, and to make available the results of such inquiry. The Foundation's area of interest includes laboratory experiments such as those concerned with extra sensory perception and psychokinesis, as well as so-called spontaneous phenomena such as telepathy, clairvoyance and precognition.

The Parapsychology Foundation's activities include (a) the encouragement of research by university groups and individuals, (b) the organization and/or sponsorship of conferences with the aim of encouraging the exchange of data and views among parapsychologists and authorities in other fields of science, and (c) a publications program which includes the quarterly *International Journal of Parapsychology*, a bi-monthly *Newsletter*, books and proceedings of various conferences, as well as the *Parapsychological Monographs*.

ESP in Relation to Rorschach Test Evaluation

Gertrude Schmeidler
The City College, New York, N. Y.

Parapsychology Foundation, Inc.
29 West 57th Street,
New York 19, N. Y.

ACKNOWLEDGMENTS

I am indebted to many for help in working out this project. In approximate chronological order they are:

The Hodgson Fund of Harvard University, which provided the greater part of the research funds for gathering the data;

Dr. Gardner Murphy, who as Hodgson Fellow arranged for the research to be performed, and who has continually guided me with his wise and patient advice;

The American Society for Psychical Research, Inc., for which I was working as Research Fellow during part of the period of gathering and analyzing records;

The following instructors who permitted me to obtain data from their classes: Ruth Berenda, Genevieve Chase, Rudolph Ekstein, Eugene Hartley, Rose Kushner, Gardner Murphy, Burke Smith, Virginia Staudt, William Triebel, Edith Wladkowsky and Bohdan Zawadski;

Mrs. Adeline Roberts, who independently analyzed many of the same Rorschach records as I, thus providing information on the communicability of the method;

Dr. R. A. McConnell, whose analysis of a part of the data uncovered an unanticipated pattern, and who thereby prevented publication of the material in an incomplete form;

Mrs. Eileen J. Garrett and Mr. Martin Ebon, who were extraordinarily patient and encouraging, and also most helpful, both as editors of the Parapsychology Monograph series and as executives of the Parapsychology Foundation, Inc.;

Dr. Karlis Osis, who suggested one of the analyses, and whose editorial advice has been very helpful;

Mr. J. Fraser Nicol and Dr. Betty Nicol, who gave generously of their time for editorial advice and statistical treatment, and who both suggested and worked out some new and useful ways of analyzing the records;

Mr. Edward Turner, who performed two of the analyses of variance reported herein,

Dr. Jacob Cohen, who acted as statistical consultant for the latter part of the work, including the formal tests of the two major hypotheses.

TABLE OF CONTENTS

ESP in Relation to Rorschach Test Evaluation

ERRATA

Page 22, Table 6. 27 well adjusted sheep scored at chance; 130 scored below chance.
Page 24, Table 8. Number of runs for sheep with no shock was 224.

I. BACKGROUND OF THE EXPERIMENTS

Preliminary Considerations and Statement of Sheep-Goat Hypothesis

The data to be reported here were gathered as part of a larger investigation of ESP and personality dynamics. My working hypothesis was that everyone or almost everyone has some ESP ability. The fact that confronted this hypothesis was that some of my subjects made high ESP scores, some middling and some low; their average was near mean chance expectation. My problem, then, was to find whether some psychological factors were associated with the different kinds of scores: whether, in Boring's phrase (1924) these statistical frequencies represented dynamic equilibria. If so, it would be the near-random sampling of psychological traits or attitudes that resulted in the near-random distribution of ESP totals. Other major findings from this larger investigation have already been described in considerable detail (Schmeidler and McConnell, 1958) but will be summarized here because they constitute the background of the present report.

Preliminary tests had indicated that subjects who were friendly to the ESP research made higher ESP scores than subjects who disapproved of it. This is, of course, only what many previous workers have reported. But in my intellectually oriented preliminary group, consisting entirely of psychologists at Harvard, it was striking that friendliness was expressed as acceptance of the possibility of ESP and disapproval as rejection of this possibility. Formal investigations were therefore begun to follow up this lead. Subjects were asked about their attitude toward the possibility of paranormal success in the experimental situation. Their answers "separated the sheep from the goats." To avoid repetition of the cumbersome accurate designations of the two groups, the term "sheep" has been used to designate subjects who state that they accept the possibility of paranormal success in the given experimental situation (even if they think it a very unlikely possibility and even if they think that they themselves cannot succeed) and the term "goats" to designate subjects who state that they believe there is no possibility of paranormal success in the given experimental situation (even if they believe that under other conditions paranormal success might occur). The hypothesis was stated that sheep would tend to have higher ESP scores than goats.

ESP Procedure

The experimental procedure will be sketched only briefly here. (See Schmeidler and McConnell, 1958 and Schmeidler, 1959 for a fuller account.) Stimulus material consisted of concealed lists of randomized targets. There were five possible targets for each response; thus the

probability of chance success in any response was 1/5. Each list consisted of 25 items, and the sequence of 25 responses was called a run; thus the mean chance expectation for each run was 5 correct responses or "hits." Subjects' responses were written. All responses received at least two independent checks against the target. Subjects responded to the questions which determined whether they would be classed as sheep or goats before they knew their ESP scores.

Within this general framework, however, there were many changes of procedure during the seven and a half years of the investigation. Some were dictated by convenience: in the first three series, for example, two offices were available to me, and targets were kept in a separate room from the subjects. When I moved to another building where only one office was available, targets were concealed in a closed closet or drawer. Some were introduced to save time: earlier subjects were tested individually, and later subjects in groups; in the first three series subjects were required to make one run as a unit, and in later series they were required to make two or three runs without interruption. Some were made at the subjects' request: though most of the runs were of the clairvoyance type, that is, hidden from everyone at the time the responses were being made, there were some GESP runs where an agent looked at each target. Some were introduced to make the procedure more interesting to the subjects: earlier targets consisted only of lists of the standard five ESP symbols, but for the last year and a half of the research the targets consisted of lists of five colors paired randomly with the five ESP symbols.

After each unit of two or three runs, a different task was usually, but not always assigned. There were many other variations, of which I will mention only a few: most of the group tests were conducted by myself on my own students, but some were conducted by others on their students (using my instructions and procedure) and some by myself on students in others' classes. The interpolated tasks differed for different groups. Some of the subjects were given only a few words of introduction to the experiment and others discussed it for an hour or more before they began their responses. Some of the classrooms were overheated. Some tests were administered soon after examination papers had been returned. Some were given early in the semester and others near the end.

Sheep-Goat Differences and Their Implications

Seven series with individually tested subjects gave over-all results in conformity with the sheep-goat hypothesis, although the distribution of scores always showed considerable overlap between sheep and goats, and in many of the series the mean difference between groups was small. The later group experiments also supported the hypothesis, though here there were occasional reversals within a class or within a semester, and the average scores of both sheep and goats were nearer chance than in the individual series. The data are summarized in Table 1. Other investigators, using the same or similar methods, have tended on the whole to obtain somewhat similar results. The hypothesis that sheep will on the average score higher than goats seems therefore to be rather well supported for the samples (consisting almost entirely of college students) that have been tested.

An affirmative conclusion suggested by these data is that ESP success, like success in other activities, is affected by the subject's motivation.

Sheep, who in general were interested, friendly and cooperative, could be expected to be motivated toward high scores. Goats, who in general rejected the basic plan of the research ("This is ridiculous!" was a not infrequent comment) could be expected to be motivated against high scores, since failure to hit the targets seemed, to many of them, to prove their thesis that the assigned task was an impossible one.

Table 1

Mean ESP Scores of Subjects Who Accepted the Possibility of Paranormal Success under the Conditions of the Experiment (Sheep) and of Subjects Who Rejected This Possibility (Goats)

Subjects	Number of Subjects	Number of Runs (25 guesses)	Deviation from Chance Expectation	Mean Hits per Run	P
Sheep tested individually	111	1055	+242	5.23	.00019
Sheep tested in groups	692	5985	+614	5.10	.00007
Total sheep	803	7040	+856	5.12	.0000006
Goats tested individually	40	853	−116	4.86	.047
Goats tested in groups	465	4050	−301	4.93	.018
Total goats	505	4903	−417	4.91	.004

Difference between mean scores of sheep and goats tested individually is .37; P = .00006.

Difference between mean scores of sheep and goats tested in groups is .17; P = .00003.

(But here an apparent paradox arises. Almost all goats who discussed the procedure with me reported that while they were making their responses they were consciously trying to succeed, though they knew they could not. But even with this conscious attitude of cooperation on the part of most, the over-all average of the goats was significantly below chance expectation. The paradox can of course be resolved by the inference that the goats' tendency to reject the task was associated with an *unconscious* negativism. We can make an analogy between them and a resentful acquaintance who makes tactless remarks or brings up hurtful conversational topics and then is surprised at his own ineptness because he had no conscious intention of causing pain. Thus our first conclusion should be read as a statement that unconscious motivation as well as consciously held attitudes may influence ESP scoring.)

A second, more negative conclusion must also be stated. It is that the sheep-goat division of subjects is not in itself an important factor in ESP success. A large minority of the sheep had ESP scores below mean chance expectation; and a large minority of the goats had ESP scores above this level. Further, there was no indication that the sheep who most completely accepted the possibility of ESP success had the highest ESP scores (Table 2) nor was there evidence from intro-

spective reports that the goats who were most vehement in their rejection of the research had the lowest scores. (One other point should be noted in passing, in connection with Table 2—a point which is perhaps irrelevant here, but to which we shall return later. It is that the question about attitude which gave psychologically meaningful results when the answers were divided into the sheep-goat dichotomy

Table 2

Attempt at Quantitative Division of Sheep Attitudes
(Subjects tested 1948-1951)

Subjects' Attitudes	Number of Subjects	Number of Runs (25 guesses)	Deviation from Chance Expectation	Mean Hits per Run
Sheep (+): think it probable that there can be ESP under these conditions	61	519	+46	5.09
Sheep (?): undecided	175	1426	+88	5.06
Sheep (−): think it unlikely that there can be ESP under these conditions	121	1034	+114	5.11
Goats: think it impossible that there can be ESP under these conditions	163	1329	−128	4.90

did not lend itself to further *quantitative* meaningful subdivisions. There is evidence from other work (see Schmeidler and McConnell, 1958) that *qualitative* differences in the reasons for adopting the sheep or goat attitude toward the ESP task are meaningful. The implication is that here a yes-no, all-or-none division is useful, though finer quantitative gradations are not.)

What should be our interpretations of this second conclusion? One surely is that the responses are determined in part by such non-ESP factors as card preferences and sequence patterns, which will show only a chance relation to the targets. Like static in radio reception, this cuts randomly across what is of interest to us. But another interpretation may well be that the sheep-goat question, which asks only about intellectual attitudes, taps only a minor or fringe part of the motivational pattern. A sheep may fear ESP, or may resent taking part in a task which is tedious to him; and a subject may give lip service to the materialist credo by calling himself a goat but retain delight in childhood fantasies of omniscience because of his secret ESP ability.

In one sense this latter interpretation is a corollary of the first of our conclusions. Both converge on the same research directive: to learn more about the subjects' attitudes and motives, including unconscious motives. The simplest way to attempt this is to ask the subject directly. But preliminary questions along these lines indicated that, even with subjects who liked to talk about themselves, the answers had little or no clear relation to ESP scores. And indeed if unconscious motives are important, no other result could be expected. Direct answers to straightforward questions, taken literally, will describe only conscious factors.

What indirect questions could be put to the subjects? There are many questionnaire-type psychological tests which could be employed; and one of these, the Allport-Vernon Study of Values, indicated that it was the subjects most keenly concerned with theoretical, intellectual values who gave the clearest sheep-goat separation (see Schmeidler and McConnell, 1958). But the techniques on which I depended most heavily, techniques strongly advocated by clinical psychologists, were projective tests. In these tests the subject must respond to a task in which there is no single "correct" answer. He may be shown a picture and asked to tell an imaginative story about it (Thematic Apperception Test); he may be asked to complete a sentence which begins as vaguely as "John likes . . ." (Sentence Completion Test); he may be shown a meaningless ink blot and asked to say what it looks like or what it might be (Rorschach Test). His answers can be influenced only to a small extent by the stimuli, when the stimuli are so vague; to some extent they must be influenced by what he himself brings to the task—by his past experience and his response tendencies, including the unconscious tendencies.

Several such projective methods were explored. One, a study of responses to frustration, gave results which were marginally significant and seemed meaningful: a low positive correlation between ESP scores and the tendency to respond to frustration without aggression, and a low negative correlation between ESP scores and the tendency to respond to frustration with outwardly directed aggression. But my greatest emphasis has been placed on the Rorschach test, where more than one thousand protocols were obtained. The data on frustration and a small part of the Rorschach data have already been described (Schmeidler and McConnell, 1958). The following chapters give a short description of the Rorschach method used, and report the rest of the Rorschach findings.

II. THE RORSCHACH TEST AND THE INSPECTION TECHNIQUE

The Rorschach Test

The most ambitious personality test, probably the most widely used and perhaps the most controversial is the Rorschach. Its proponents claim that it shows the basic structure of personality, and also that its flexibility allows it to demonstrate many of the unique patterns of each individual who has taken it. It is used to diagnose neurosis, psychosis and even brain tumors, to predict the likelihood of successful psychoanalysis, to show whether a workman will cooperate better with one or with another foreman, and for almost the whole intermediate range of possible problems. Many clinicians feel that if only one personality test was available to them, this would be their choice—however all clinicians, I think, agree that the Rorschach should if possible be supplemented by other tests. All agree that it is unreliable: that results vary with the personality of the examiner and the skill and interests of the interpreter. Research results have run through the range of showing very high validity to none. Some psychologists consider it useless. In this welter of conflicting opinions and data, my own estimate is that the test is excellent when wisely handled. Negative results on validity seem to me to be due either to inadequate conditions of administration or else to inadequate preparation for interpretation. (If we do not know the personality characteristics of a good combat pilot, we cannot legitimately use a personality test to judge whether a man will with training become a good combat pilot. If we try to make this judgment and fail, it has not been demonstrated that the *test* was deficient.)

The Rorschach consists of ten standard ink blots, administered serially. The subject is asked to state what they look like or what they might be. The thesis behind the test is that the blots do not in fact represent anything. Therefore any answer which the subject gives must in large part be a result of what he has brought to the situation and of what he is ready to see in ambiguous stimuli: a projection of his own tendencies for perception and response. Scoring norms have been found for the location of the response (whole, part, small part, etc.), for the "determinants" of the response (form, shading, color, impression of movement, etc.), for the content, and for various other factors such as the number of blots to which the subject refuses to respond. Interpretation is made in terms of these norms and also in terms of such qualitative patterns as content analysis and serial analysis of what the subject says and how he acts while he is saying it. The average time for administration of the Rorschach is about one hour; the average time for interpreting a record varies widely, but probably averages three to five hours.

The Munroe Check List: a Test of Social Adjustment

A useful method for quick interpretation of the major adjustment patterns in the Rorschach was proposed by Munroe (1945). It consists, essentially, of listing twenty-odd scoring categories and defining the normal or "safe" limits for each. When a subject's responses fall outside of those limits he is assigned one, two or three entries on the check list, the number of entries being dependent on the extent of his deviation from the previously stated limits. The total number of checks is his final score and represents an approximation of his general social adjustment. With this measure, then, the higher scores should correspond to less adequate social adjustment. Scoring of a protocol can be completed in ten minutes to half an hour.

This method could be made objective enough for routine coding by clerks and machine scoring. Munroe has however been careful to state that it should not be so used. She writes:

> The percentages have purposely been left blurred at the edges. For example F% (the percentage of responses determined by form of the blot, without reference to shading, color or impression of movement) receives no entry between 15 and 50%, one check between 50 and 75%. Decision about entering a check for an F% at or very near to 50% should be made on qualitative grounds rather than on the grounds of strictly arithmetical calculation. If many of the F (form-determined responses) are somewhat doubtfully scored and seem to verge on Fc (responses determined primarily by form and secondarily by shading), FM (responses determined by an impression of movement of an animal), etc., or if the F score is frequently enriched by additional elements, the check should not be entered. If on the other hand the F scores are mostly quite clear and rather bare, if they seem to represent a positive search for formal accuracy as in frequent meticulous detail-responses, if they seem to indicate formal evasiveness as in frequent mention of maps, bones and the like, the balance is swung toward the plus entry. Decisions of this nature are made only at the borderline, but there they seem to serve the intention of the check list more nearly than an arbitrary break between 49 and 50%. The check list should be objective, but not pedantic.

Determination of check list entries should therefore depend partly on routine coding, but for borderline cases should depend largely on the intuitive impression made by the protocol as a whole on an experienced examiner.

Munroe describes a validation of her technique in research on students at Sarah Lawrence College. She found that more satisfactory college records were achieved by students whose general social adjustment was good (as gauged by check list entries and certain other criteria) than by those whose adjustment was poor, even when intelligence was equated for the two groups. This implies that the relatively well adjusted students were more able than the poorly adjusted to use their ability to achieve their goals. In a later study of students at Barnard College, Schmeidler, Nelson and Bristol (1957) used a similar approach and obtained similar findings.

Munroe's adjustment categories were A, B, C and D; of 348 students, 53% fell into the A and B categories. The letters corresponded roughly to the number of entries on the check list. No detailed analysis of college records and check list entries has been published; but Dr. Munroe was generous enough to make available to me a listing of

all the data which she retained that bear on this question. They are summarized in Table 3.

Table 3

Total Number of Check List Entries and College Adjustment
(Subjects tested by Munroe)

Subjects	Total Number of Check List Entries													
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Number of subjects with adequate college records	1	3	4	2	9	15	13	11	10	1	4	0	0	2
Number of subjects with evidence of college maladjustment	0	1	3	1	2	4	6	4	5	5	7	0	0	4
Per cent adequately adjusted	100	75	57	67	82	79	68	73	67	17	36	—	—	33

It will be noted that there is a crude correspondence between ratio of well adjusted students and number of check list entries. This is shown by Munroe to be significant when the data are divided into the groups listed in her original monograph: "6 entries or less, 7 to 9 entries and 10 entries or more." A curious fact is the sharp break in ratios between the categories of 9 or less, where in all cases there is a larger number of well adjusted than of poorly adjusted students, and the categories of 10 or more, where the pattern is reversed. Perhaps the important distinction for college adjustment is between the crude classes of good or poor social adjustment, while finer distinctions within the two classes are relatively unimportant? The question will be discussed at greater length, with reference to ESP scores as well as to college performance, in Chapters III and V.

Procedure for Administering and Scoring the Rorschach

In the research to be reported below, the group method of administering the Rorschach was used for most subjects. Slides representing the Rorschach cards were exposed in a darkened room and subjects were instructed to write their responses to those slides. Details of the procedure for administration are given in full in another publication (Schmeidler, Nelson and Bristol, 1957) and some comments about differences between group and individual protocols are added. For a few subjects in the preliminary group who had recently taken individual Rorschachs, the individual records were used. For a few other subjects who were absent on the day of the group Rorschach, individual tests were given. These were usually self-administered, that is, the subjects were given instructions similar to those for the group records, were permitted to examine the Rorschach cards without time restriction, and wrote their own responses to them.

One disadvantage of using group instead of individual Rorschachs is that the subjects do not know what information is needed for scoring their responses, and therefore what they write is often inadequate. It is considered inadvisable to tell them explicitly what is of concern to the examiner, since this may suggest that they ought to see what they are asked if they saw. Techniques for handling the problem in individual administration consist first of pointed but non-leading questions (inquiry) and secondly of leading questions (testing the limits) where the answers are used only in carefully delimited ways. In early series an attempt was made to combine these advantages of an individual test with the time-saving of the group test by giving all records a preliminary score and returning to the students an inquiry, in writing, about ambiguous responses. Answers from the students were occasionally helpful, but left a great deal to be desired. In all later series, students were asked to come to my office to discuss their Rorschachs. A brief interpretation of the major features was promised, and this was attractive enough to bring most students to the appointment. During the session I inquired about the ambiguous responses, then rescored the record, then discussed its interpretation with the student. Though this was time-consuming, it had the advantages of permitting scores to be entered with about the same level of confidence as in the conventional individual administration, and also of establishing friendly relations with the students. Most of the Rorschach scores reported below represent the results of such interviews, as supplements to the original protocols. In the few cases where no appointment could be arranged, I made an effort to inquire about doubtful responses before the beginning or after the end of the class period. Where no inquiry was possible, my best guess as to the correct scoring was utilized.

Scoring and revision of the Rorschach was in the test series completed without knowledge of ESP scores, or of whether the subject was a sheep or a goat. (Occasionally a student's spontaneous and unsolicited comments would indicate a favorable or unfavorable attitude to ESP. But this is not enough to give sheep-goat classification, since a subject whose attitude is unfavorable may still retain enough doubt, or caution, to keep from complete rejection of the possibility of paranormal success; and a subject extremely interested in psychic research will often disapprove of the impersonal experimental procedure so strongly as to call himself a goat for those particular conditions.) For the last few years of the project, Rorschach scoring and revision of the scoring were both completed before the ESP tests were administered.

All Rorschachs were scored by me according to the instructions given by Munroe for the check list; and the total number of check list entries is the only criterion of social adjustment used in this report. Some supplementary Rorschach scores not used in check list entries have been noted for all records, and the reader will observe that certain of the signs described later do not represent items on the check list.

Two points in connection with the check list entries probably need explicit statement here. One is that Munroe's instructions deliberately leave some margin of doubtful scores, where decisions are to be made on the basis of an experienced examiner's evaluation of the context of the protocol rather than a mechanical or pedantic count of items. As a result, the only correct way to follow her method is to determine some entries on the basis of rules which are stated in general rather than specific terms. The second point is that my interview method of

conducting an inquiry produced the awkward social situation of having a student sit twiddling his thumbs and waiting for an interpretation of his record while I revised my original, tentative scores in the light of the information he had just given me. Naturally I did not like to delay him unduly, and so my scoring was more hurried, and therefore probably in some cases less accurate, than if he were not there. In some instances I returned to the record after the student left, and made a more careful second rescoring. This was especially likely to be done for marginal cases where (as will be described later) the total number of check list entries fell just above or just below a preset boundary between the categories of good and poor adjustment. In most cases, however, because my own schedule was heavy, the scoring made while the student was present was retained without further examination.

III. THE HYPOTHESIS OF SOCIAL ADJUSTMENT IN RELATION TO ESP SCORES

Preliminary Data and Their Implications

In the spring of 1945, group ESP tests were first used in my research project, and I continued to use group tests until 1951, when the project ended because financial support for it had terminated. All the data reported hereafter come from subjects whose ESP tests were conducted in groups. All these group tests were held in college classrooms.

At the same time that the group ESP tests were begun, I began to administer the Rorschach to all subjects who were willing to take it. With the exception of a few who reported that they were color blind, and who therefore could not have their records scored in the conventional way, all ESP subjects tested in college classes by me, or tested for me by a colleague, between 1945 and 1951, for whom Rorschachs were available, are described in the following tables.

The total number of such subjects is 1,062. Of these, the first 58 constitute the preliminary group from whom the hypothesis about adjustment was derived; the remaining 1,004 constitute the test group for the hypothesis about adjustment. The first 250 of these latter subjects constitute the preliminary group from whom the hypothesis about Rorschach signs was derived; the remaining 754 constitute the test group for this hypothesis.

For the first 58 subjects, ESP scores were known to me when the Rorschachs were scored. For all later subjects, I performed the Rorschach scoring and made the entries on the Munroe check list and on my own list of signs before I knew the subjects' ESP scores. In the later years of the research, ESP tests were not administered until after all Rorschach scores and check list entries had been completed.

After gathering the first 58 Rorschachs I examined the results in the hope of finding Rorschach patterns that related to ESP scores. One lead seemed promising. The sheep with relatively few check list entries (better social adjustment) tended to have higher ESP scores than other sheep; and the goats with relatively few check list entries tended to have lower ESP scores than other goats. This makes good sense in terms of our concept of social adjustment; and since there was the happy combination of a marked trend of the data and a good rationale, it was decided to perform formal investigations of whether the trend would persist in later series.

Let us explore the rationale. Social adjustment (as distinguished from personal adjustment) means essentially handling oneself effectively according to the situation one is in. Now the question is: What situation were our subjects in? Presumably the sheep were in a situation where they had been assigned an unfamiliar but not impossible

task, hitting ESP targets. Successful handling of themselves would therefore be represented by high scores in ESP.

But what situation were the goats in? Presumably one where a misguided experimenter was wasting her time and theirs because she had not yet realised that it was impossible to hit ESP targets (or in a more sophisticated version, that it was impossible to hit ESP targets with better than chance scores). Effective handling of oneself here would probably be the demonstration of the impossibility, that is, the obtaining of low scores for the statistically naive, or the obtaining of approximately one hit out of five responses for the statistically wise. ESP scores of chance or less would thus be consistent with good social adjustment for goats.

(This now seems to me an oversimplified statement. People with good social adjustment who have alien demands made upon them by an authority, might set aside their own preconceptions, temporarily, to conform to those demands. This would be especially likely if their personal concern with the rights and wrongs of the situation was slight, and if the pressure from the authority was strong. Thus perhaps high scores should be predicted for some of the well adjusted goats, and low scores for others, depending on the intensity of their feeling about ESP theory and on their interpretation of the pressure I was exerting. Unfortunately we do not have information about either of these questions, since the point did not occur to me until after the data had been gathered.)

With this reasoning, no assumption is made about the relation between social adjustment and ESP *ability*. The only expectation is that, whatever the level of ESP ability, its effective *utilization* will tend to be related to adjustment patterns, as is the utilization of other abilities. This should be considered only a special case of the general rule. When allowance is made for differences in intelligence, both Munroe and I have found that students with better social adjustment tend to have better college grades. Within the limits of their physical endowment, people with good social adjustment are likely to present a pleasant appearance. If they are given a modicum of encouragement, individuals with good social adjustment are likely to be friendly and cooperative. We could multiply examples; our assumption is that the general principle will apply to ESP as it does to other types of response.

What of the students with poor social adjustment? Many of them fail at college, even if they are highly intelligent; but many have an especially strong drive for intellectual achievement and make outstanding college records. Some wear ill-fitting, dirty or inappropriate clothes and present a far less attractive appearance than their physical make-up and clothing budget justify; but some are compulsively neat, and some put such exaggerated effort into making the best of their appearance that they are conspicuous for good grooming and dress. Some are seclusive and disliked; some are so eager to be helpful or friendly or elected to college office that they are well and favorably known. As a group they do not represent failures so much as unpredictables—although of course any individual's pattern may well be predictable on the basis of his personal history.

Applying the same sort of expectation to ESP performance as to other kinds of behavior, we could anticipate that some of the best and some of the worst ESP scores would occur among the subjects with

poor social adjustment, but that for the group taken as a whole no single clear pattern would appear. In other words, because—by definition—individuals with poor social adjustment give idiosyncratic responses rather than the responses that are expected of people in general, an experimenter would be well advised to refrain from making predictions about them unless there is detailed knowledge of individual idiosyncracies. It is not even safe to expect (as I initially did) that the group of poorly adjusted subjects in a college population will be more variable than the group of well adjusted subjects; for if a substantial number of them aim at a safe middle course of extreme conformity, they could balance the deviant behavior of the others.

Once this reasoning had been worked out, after the general relation between social adjustment and ESP scores was observed in the first sample of 58 cases, the next research problem was to put a generalization about it into testable form. My solution, which later work proves to have been unfortunate, was to state my hypothesis in terms of a dichotomy of good and poor adjustment. This is clearly inadequate from one point of view: any quantitative scoring scale, such as the check list, will show degrees of adjustment from very good to very poor. Another inadequacy is that any single statement of a person's social adjustment level is an oversimplification. Someone may habitually show good adjustment in situations where expectations are clear and lines of authority are well drawn, but poor adjustment in others; another person may be flexible and effective in a wide range of situations where he feels himself trusted and liked, but be unable to act or even think effectively if he feels himself among suspicious strangers; and so on. The useful statement of social adjustment, as of most other psychological tendencies, is rather a profile showing adjustment in different types of situations than an average of all the adjustment levels which make up the profile.

But the issue is a delicate and complex one. As Murphy (1947) has brilliantly demonstrated in his discussion of discontinuities in personality, social factors can create discontinuities which, once created, have pervasive and lasting personality effects. Two students of equivalent ability, interests and outlook may, for example, take the same entrance examinations to medical school. Because of random factors, one may score just at the level which permits his acceptance and may some years later become a doctor; the other may score just below this level, be denied admittance, become embittered and resentful of academic intellectual standards, and engage in a different career. The social attitudes and the personality traits of the two may diverge markedly in later years, if one has identified himself with the dedicated members of a demanding profession and the other with a cynical and self-indulgent group.

There might, similarly, be qualitative differences in outlook between those students whose social adjustment is just good enough to make them able to conform to a reasonable number of college demands and those whose adjustment is just enough poorer to make them fail many college demands. The former might feel sufficiently at home in the classroom so that they tend to identify with the instructor and to gain reassurance and satisfaction from conforming behavior; the latter might find satisfaction in frequent criticisms and show markedly more withdrawal and ambivalence. (I do not argue that there is such an effect; there are no clear data, though Munroe's, cited in Table 3,

are not inconsistent with it. I argue only that there might be one; that this is an issue to be decided by research, not dismissed out of hand.) It *might,* therefore, be meaningful to divide subjects into the two distinct groups of those who seem adequately adjusted socially when tested in a particular situation and those who do not. If this is ever justified it is likely to be so in research like the present one, where the (Rorschach) test from which adjustment was judged and the dependent variable (of ESP tests) were administered by the same experimenter in the same setting.

Whether or not an adjustment dichotomy is sound in theory, it has worked out badly in practice. Since I knew the records would be dichotomized, and since I was under considerable pressure for economy in time, I tended to be hasty in entering check marks for subjects who would clearly fall on one or the other side of the good-poor boundary, and to give far more careful attention to the records near the boundary line. Thus to some extent, though staying within the broad directives set down by Munroe, I often used more careless methods of examination for records where adjustment was really good or really bad than for records where it was dubious. Increased care took the form not only of more precision in scoring specific responses, but also—as Munroe recommends—of making more decisions about individual entries with reference to the protocol as a whole, that is, of utilizing responses remotely relevant to the entry under consideration as well as responses directly relevant to it. There are two consequences of this. One is that there are probably substantial random irregularities at the two extremes of the adjustment scale. The other is that qualitative interpretations of the protocols as a whole entered into the scoring of the borderline cases to a greater extent than into scoring of the extreme cases.

Table 4

Preliminary Study of ESP Scores of Sheep and Goats with Good or Poor Adjustment (Subjects Tested Spring, 1945)

Subjects	Number of Subjects: ESP Scores in Relation to Chance Expectation — Above	At	Below	Number of Runs	Deviation from Chance Expectation	Mean Hits per Run
Well adjusted (0-10 check list entries)						
Sheep	13	0	5	162	+95	5.59
Goats	4	0	11	137	−39	4.72
Poorly adjusted (11+ check list entries)						
Sheep	7	0	7	129	−26	4.80
Goats	6	1	4	99	+8	5.08

To return to a chronological description of the procedure: in preparation for stating the formal hypothesis about adjustment in terms of two distinct groups, I reexamined the available data to find

where the line of separation should be drawn. The normative population was the 348 Sarah Lawrence students described by Munroe; and I assumed that the distribution of adjustment there would be approximately the same as in the college groups I tested. In Munroe's sample 53% of the subjects were classed as well adjusted. In my sample of 58 subjects, drawing the dividing line at 10 checks or less for good adjustment gave the separation closest to 53%. This was therefore adopted as the criterion. The same criterion was used throughout the period of gathering and scoring data; and at final count about 54% of my subjects had 10 check list entries or fewer and were therefore classed as well adjusted.

The data of the preliminary group are summarized in Table 4.

Statement of Formal Hypothesis

The formal statement of the hypothesis about adjustment was first published in the following words (Schmeidler, 1947): "Sheep who are well-adjusted will, on the average, make higher ESP scores than sheep who are not; and goats who are well-adjusted will have lower ESP scores than the other goats." Good adjustment was operationally defined as ten or fewer entries on the Munroe check list; poor adjustment was operationally defined as eleven or more entries.

IV. THE HYPOTHESIS OF RORSCHACH SIGNS IN RELATION TO ESP SCORES

Given a test like the Rorschach, which measures many aspects of personality, it seems wasteful to limit its use to a single score—unless that score gives nearly a one-to-one correspondence with the process being studied. Since it has been obvious from the first that there was no one-to-one correspondence between ESP scores of sheep and goats and their total number of check list entries, further analyses of the Rorschachs was undertaken.

Procedure

For this portion of the project, the first 250 records of the subjects in the test group for the adjustment hypothesis constituted the preliminary group. The method of study was almost mechanical. A list of 104 Rorschach scoring items was made, following and enlarging upon the list of the Inspection Technique. The 250 ESP scores were tabulated in these categories. Where a substantial number of sheep with a given item had low ESP scores, and the goats with that item did not include a large number with low ESP scores, the presence of the item was taken tentatively as a "sign" that subjects with this characteristic would not have good ESP scores (i.e., would not show the postulated sheep-goat difference). Six signs were pulled out of the records by this method; and a seventh (C+) was added because according to my interpretation of the Rorschach at that time, it seemed to be tied with the sixth (CF+). The records of the 754 subjects studied thereafter were taken as a test group for the signs. A description of them follows.

List of Signs

(1) R+, a large number of responses. This number was set at more than 30 for the group Rorschach and at more than 60 for the individual Rorschach.

(2) F%+. This category was taken from Munroe's Inspection List, and has been defined above.

(3) Mr, or rigid human movement. This category, also taken from the Inspection List, is defined thus by Munroe: "r (rigid, restricted). . . . If over half of the M (human movement) responses are described as the action of statues, marionettes, drawings, etc. (Probably should not be scored M at all unless lively action is described or other M are present and there is evidence of inhibition.) Or if M responses are extremely passive or rigid (e.g., sleeping figures in V, figure standing with legs together in I).

(4) Total Movement++. This category is also taken from the Inspection List. It is assigned when 60% or more of the Rorschach responses deserve a movement score. A movement score is given if the subject gives evidence of having seen the blot in terms of human movement or facial expression (as, "a man scowling") or of mood (as, "It gives me a feeling of calm.") or of animal movement or of inanimate movement (such as drifting clouds or shooting flames).

(5) No shading or color shock. "Shock" has been scored extremely lightly, that is, an entry for shading shock was assigned whenever the response to Figure IV seemed to express a somewhat different feeling from the responses to the preceding cards; and an entry for color shock was assigned whenever responses to either Figure II or Figure VIII seemed to express a somewhat different feeling from those to the preceding card (in the case of Figure II) or cards (in the case of Figure VIII). When these low standards for defining shock are employed, less than five per cent of the records show neither color nor shading shock.

(6) CF+. This entry is also taken from the Inspection List. The CF score is given to a response which is primarily determined by chromatic color, but secondarily by form. It is often compared with the FC score, assigned when the subject's emphasis is reversed. The entry CF+ is made by Munroe under two conditions, which are: "(a) FC (−) (i.e., only one FC or the equivalent), 2 or 3 strong CF or whole CF responses (e.g., fire, blood, sunset, vague anatomical drawings, colored fountain, etc.) Count ½ for additional CF (unless strong), CF verging on FC, and doubtful scorings. (b) CF:FC = 2:1 . . . if FC is more than 1." (The half scores are assigned to Rorschach responses. Check list entries are never given in fractions.)

(7) C+. This entry is also derived from the Inspection List. The C score is assigned to responses made entirely on the basis of chromatic color. The + entry is made for one or more of these responses, or for two responses that give color description or color symbolism. Where these latter are present but of secondary importance, or where scoring is doubtful, half credit is assigned.

It will be noted that there is considerable overlap between the signs and the adjustment scores. All signs except R+ and no shock constitute entries on the check list. On the other hand, the no shock category means that entries on the check list have been withheld.

Preliminary Data and Their Implications

A summary of the ESP scores for the 250 records from which the sign hypothesis was derived is given in Table 5. The data have been presented in more detail elsewhere (Schmeidler, 1947).

A question that must have occurred to any reader unfamiliar with the Rorschach is: "How do you translate these symbols into English?" or put more formally, "What are the psychological implications of each of these signs?" The first answer which must be given is that no single item on the Rorschach has much meaning when taken out of context—that any statement about the interpretation of a particular sign is only the crudest average approximation, and may be sadly mistaken in the individual instance. With this disclaimer which should properly be repeated again and again, I shall rough out some broad interpretations.

R+ has been described as indicating "quantity ambition," a desire for showy achievement. In a college population, this usually is more

Table 5

Preliminary Study of ESP Scores of Subjects Whose Rorschach Protocols Showed Certain Signs
(Subjects Tested Fall, 1945—Summer, 1946)

Signs	Number of Subjects: ESP Scores in Relation to Chance Expectation: Above	At	Below	Number of Runs	Deviation from Chance Expectation	Mean Hits per Run
A. Sheep						
R+ (Many responses)	7	1	13	188	−33	4.82
F%+ (Many form responses)	6	1	14	189	−24	4.87
Mr (Rigid human movement)	3	0	7	90	−21	4.77
Total Mvt++ (Many active responses)	1	0	7	72	−27	4.62
No shock (No color or shading shock)	2	0	4	54	−8	4.85
CF+ (Many responses with color dominant over form)	5	1	11	150	−42	4.72
C+ (Many pure color responses)	1	0	2	27	+8	5.30
All sheep with one or more signs*	22	2	42	590	−93	4.84
All sheep without signs	38	2	11	459	+204	5.44
B. Goats						
R+ (Many responses)	9	0	4	117	+10	5.09
F%+ (Many form responses)	9	0	8	153	+7	5.05
Mr (Rigid human movement)	7	1	2	90	+24	5.27
Total Movement++ (Many active responses)	6	1	7	127	−1	4.99
No shock (No color or shading shock)	7	1	2	91	+14	5.15
CF+ (Many responses with color dominant over form)	5	0	5	91	−13	4.86
C+ (Many pure color responses)	4	0	2	54	+20	5.37
All goats with one or more signs*	35	2	25	559	+48	5.09
All goats without signs	22	1	48	638	−175	4.73

* Since many subjects have more than one sign, this row is less than the total of the separate signs.

specifically a desire for showy intellectual achievement. When R+ is taken from the group Rorschach, with its limited time intervals, it also shows quickness of perceptual responses.

F%+ implies, as could be inferred from the passage by Munroe quoted in Chapter II, a desire to be factual and correct, even if this rationality and safety are attained at the expense of warmth, speculation and subtleties.

Mr carries the connotation of lack of freedom in developing one's feelings or ideas.

Total Movement++ indicates overconcern with introspection or intellectual activity. Such an extremely strong drive in one direction often—though of course not always—implies a turning away from other possibilities. It may be associated with introversive or intellectual overemphasis that does not permit free awareness of the potentialities of the environment. Like R+ it is often, in a college population, associated with a need for intellectual achievement and recognition.

Absence of either shading or color shock, when shock is scored as lightly as has been my practice, indicates an absence of ready responsiveness to changes in the stimulus field, a tendency to follow one's predetermined plan without quick sensitivity to new factors.

CF+ indicates lively, impulsive responsiveness to the more colorful or emotional aspects of the world.

C+ indicates that on some occasions a subject will respond in an extremely impulsive or emotional way, neglecting the cold facts that are before him.

According to this list of approximate meanings, the seven signs cluster around three personality patterns, which thus are hypothesized to be associated with poor sheep-goat differentiation in these classroom experiments. One is an extremely strong push toward intellectual achievement and marked inner or intellectual activity (R+ or Total Movement++); a second is rigidity, constraint or withdrawal (F%+, Mr or no shock); a third is over-impulsiveness (CF+ or C+).

These terms are admittedly loose, and not easy to define operationally. For one (rigidity) there is good evidence that different kinds of rigid behavior are independent of each other; this may well be true for others as well. Thus the lack of precision mentioned earlier is compounded: not only may one of these signs, in a particular Rorschach, not deserve the general interpretation just cited; but the general interpretation may also denote one form of behavior in one person and a different form in another. No high level of validity can be expected for them. There may however be some low level of validity, and the general behavioral syndrome that they connote may be identifiable. Thus if enough cases are gathered, a trend supporting these interpretations may be shown within a group in spite of the individuals who are exceptions.

Statement of Formal Hypothesis

Following the line of thought suggested by these interpretations, the formal statement of the hypothesis about signs was first published in the following words (Schmeidler, 1947): "Sheep who are not constricted, not inclined to impose rigid barriers on their creative thinking, responsive to change, not needing to impress others with their intellectual prowess, not too overactive in their inner life and not overimpulsive, will have higher average ESP scores than other sheep; and goats with these characteristics will tend to have lower ESP scores than other goats." "These characteristics" were of course operationally defined as the absence in the Rorschach protocol of any of the seven signs: R+, F%+, Mr, Total Movement++, no shock, CF+ and C+.

V. INTERIM ANALYSES OF DATA ON ADJUSTMENT AND SIGNS

Summary of Data on Social Adjustment

The first examination of the data on ESP scores and social adjustment took the form prescribed by the original hypothesis. The results are summarized in Table 6, and are consistent with the prediction that sheep who are adjudged well adjusted socially (that is, who were assigned 10 check list entries or fewer) will tend to have higher ESP scores than sheep who are adjudged poorly adjusted socially (that is, who were assigned 11 check list entries or more) and that goats who are adjudged well adjusted socially will tend to have lower ESP scores than goats who are adjudged poorly adjusted socially.

Table 6

Interim Summary of ESP Scores Collected to Test the Hypothesis that Well Adjusted Sheep Will Have Higher ESP Scores than Other Sheep and that Well Adjusted Goats Will Have Lower ESP Scores than Other Goats
(Subjects Tested Fall, 1945-1951)

Subjects	Number of Subjects: ESP Scores in Relation to Chance Expectation: Above	At	Below	Number of Runs	Deviation from Chance Expectation	Mean Hits per Run
Well adjusted (0-10 check list entries)						
Sheep	213	26	131	3189	+610	5.19
Goats	85	13	126	1962	−312	4.84
Poorly adjusted (11+ check list entries)						
Sheep	101	16	128	2112	−66	4.97
Goats	89	4	72	1421	+114	5.08

Number of well vs. poorly adjusted subjects scoring above vs. at or below chance:

Sheep – Chi square = 15.77, 1 df; $P < .001$
Goats – Chi square = 9.84, 1 df; $P < .01$

Difference between mean scores of well and poorly adjusted sheep is .22; $P < .001$.
Difference between mean scores of well and poorly adjusted goats is .24; $P < .001$.
Interaction between sheep-goat and adjustment by analysis of variance: $F = 25.75$; d.f. = 1:1000; $P < .001$.

Analysis by chi square and by *t* indicate that the hypothesis as it was stated has been confirmed at a high level of statistical significance.

The question has been raised as to whether this method of judging social adjustment is communicable, and as to how reliable it is. The only information on these points that was collected as part of our project came from the generous cooperation of Mrs. Adeline Roberts, an experienced and skillful Rorschach analyst who volunteered to help with the research. She had been unfamiliar with the method of scoring which I used (there are several methods of scoring Rorschach responses, which resemble each other about as closely as various languages of the same family) and was therefore of course unfamiliar with the check list technique, which depends upon this scoring method. Nevertheless, after reading Munroe's description of the check list and discussing it with me, she began to use the check list. In conformity with my procedure, she did not know at the time of scoring what the subjects' ESP scores were, nor whether they classified themselves as sheep or goats. During the period that we worked together, each of us scored independently the Rorschachs of all ESP subjects. Each of us conducted about half of the student interviews. The results of our scores are given in Table 7.

Table 7

Extent of Agreement between Two Examiners on Whether Subjects Were to Be Categorized as Well or Poorly Adjusted

Scoring	Number of Subjects: ESP Scores in Relation to Chance Expectation: Above	At	Below	Number of Runs	Deviation from Chance Expectation	Mean Hits per Run
Examiners agreed on good adjustment category (0-10 check list entries)						
Sheep	26	1	8	315	+57	5.18
Goats	1	0	4	45	−30	4.33
Examiners agreed on poor adjustment category (11 or more check list entries)						
Sheep	6	2	20	252	−63	4.75
Goats	2	0	4	54	−5	4.91
Examiners disagreed on whether a protocol had more than 10 check list entries						
Sheep	4	1	4	81	−6	4.93
Goats	0	0	0	—	—	—

The over-all pattern of ESP scores for subjects classed as well or poorly adjusted is similar for both our scores, and both are consistent with the general trend of the data. It will be noted that in nine out of the 83 cases our check list totals differed in placing a subject in the

well or poorly adjusted category; and we discussed each of these cases. It was interesting to us that we always agreed on the general qualitative picture, though there were minor differences between us as to the relative strength of the different factors.

The correlation between our total check list entries was +.88. This figure is based on only 81 subjects, since for two cases Mrs. Roberts entered a "+" indicating that the total was higher than 10, but did not complete the detailed entries.

Disagreements in entries depended largely on our using different standards for giving main as opposed to additional scores, and for counting variants of the same response as a single response or as separate ones. This small study therefore indicates, insofar as a single case can, that the general method is communicable to someone experienced in Rorschach interpretation, and that its reliability, for two examiners who have not worked closely together, is fair. All in all, this is about what was to be expected from the reliability coefficient of +.65 reported by Munroe, a figure obtained from the data of eleven examiners, each of whom scored eleven Rorschach records.

Summary of Data on Signs

Summary data on signs and ESP scores are given in Table 8. These results, taken as a whole, are consistent with the hypothesis that sheep whose Rorschach protocols are free of all signs tend to have higher ESP

Table 8 (continued on p.25)

Summary of ESP Scores Collected to Test the Formal Hypothesis that Sheep Whose Rorschach Protocols Are Free of Certain Signs Will Have Higher ESP Scores than Other Sheep and that Goats Whose Protocols Are Free of Those Signs Will Have Lower ESP Scores than Other Goats
(Subjects Tested Fall, 1946-1951)

Signs	Number of Subjects: ESP Scores in Relation to Chance Expectation: Above	At	Below	Number of Runs	Deviation from Chance Expectation	Mean Hits per Run
A. Sheep						
R+ (Many responses)	46	6	40	802	+79	5.10
F%+ (Many form responses)	52	9	61	1040	−2	5.00
Mr (Rigid human movement)	37	4	30	621	+83	5.13
Total Mvt++ (Many active responses)	39	6	30	650	+40	5.06
No shock (No color or shading shock)	12	2	12	217	+3	5.01
CF+ (Many responses with color dominant over form)	24	7	29	514	+30	5.06
C+ (Many pure color responses)	8	0	4	103	+22	5.21
All sheep with one or more signs*	149	29	152	2842	+86	5.03
All sheep without signs	105	10	53	1410	+347	5.25

Table 8 (continued)

B. Goats						
R+ (Many responses)	16	1	14	274	+26	5.09
F%+ (Many form responses)	20	2	22	380	−21	4.94
Mr (Rigid human movement)	21	0	16	319	+27	5.08
Total Mvt++ (Many active responses)	31	1	19	437	+84	5.19
No shock (No color or shading shock)	6	1	5	103	+7	5.07
CF+ (Many responses with color dominant over form)	17	2	6	216	+48	5.22
C+ (Many pure color responses)	1	0	5	50	−17	4.66
All goats with one or more signs*	86	5	62	1317	+149	5.11
All goats without signs	31	9	63	869	−220	4.75

* Since many subjects have more than one sign, this row is less than the total of the separate signs.

scores than other sheep, and that goats whose protocols are free of all signs tend to have lower ESP scores than other goats. A formal analysis, described in Chapter VI, shows that the hypothesis was confirmed at the level of $P < .001$. It is important to note, however, that no one of the signs shows a significant separation of high from low scoring subjects. This argues that more exploration is necessary before we can interpret the results with confidence.

A comparison of Mrs. Roberts' results with mine is summarized in Table 9. We agreed, out of the total of 83 subjects, that 25 were free

Table 9

Extent of Agreement between Two Examiners on Presence or Absence of Signs in the Rorschach Protocols

Scoring	Number of Subjects: ESP Scores in Relation to Chance Expectation: Above	At	Below	Number of Runs	Deviation from Chance Expectation	Mean Hits per Run
Examiners agreed that protocols were free of signs						
Sheep	13	0	7	180	+30	5.17
Goats	1	0	4	45	−16	4.64
Examiners agreed that protocols had one or more signs						
Sheep	14	4	19	333	−41	4.88
Goats	2	0	2	36	0	5.00
Examiners disagreed on whether signs were present						
Sheep	9	0	6	135	−1	4.99
Goats	0	0	2	18	−19	3.94

of signs; and we agreed that signs were present in 41 of the records. There were 53 instances in which we agreed on the assignment of specific signs. There were 21 signs which I entered but which she did not; and there were 15 signs which she entered but which I did not. Our subsequent discussion of these differences showed, as might be expected, that we had in some cases been following slightly different interpretations of Munroe's rules; having pinpointed these differences, both of us felt confident that if we were to continue with the project, our results would in future be much more similar to each other.

Probably the close agreement between us in evaluating the qualitative picture, the fairly good agreement on the over-all summary given by check list totals and the many instances of disagreement on detail, would have been anticipated by anyone familiar with the Rorschach. It is often described as more valid than reliable. An analogy might be drawn between the notes of two students who attended the same lecture: they would probably show greater similarity of content than of wording. Wording here corresponds to a particular Rorschach notation, and content to qualitative interpretation; and it is clear that many different wordings could convey the same content. It would take considerable drill for two individuals to adopt the same verbal patterns for conveying their ideas.

Further Analysis of Check List Entries and ESP Scores, and Its Implications

Tables 6 and 8 are not to be taken as the last word on the postulated adjustment-ESP relationship, nor, consequently, on the sign hypothesis which was a part of the same broad project. In the spring of 1957, Dr. R. A. McConnell had made for him a scatter diagram of the ESP scores of sheep and another of the ESP scores of goats, charted against the number of check list entries for each subject. He sent me a summary of the results; and they suggested a number of new questions about the data. The tables which follow do not represent his calculations; but I wish to acknowledge here that it was he who by this analysis opened up many of the problems and new research possibilities to be discussed below.

Table 10 summarizes the ESP scores for each of the check list totals. Three peculiarities of the data deserve attention. The first is that ESP scores are significantly higher for sheep with a total of 10 entries than for sheep with a total of 11, and the converse is true of the goats. When the dichotomy between good and poor adjustment was set at the line between 10 and 11, no such sharp break was expected. We must therefore inquire whether the abrupt change is a consequence of my having previously set an arbitrary line of demarcation there, and also whether a slight increase in the adjustment scores of some sheep near the 10-11 division-point (to give them "poor" adjustment) and a slight reduction in the adjustment scores of other sheep near the 10-11 division-point (to give them "good" adjustment) could have produced the significant difference between poorly and well adjusted sheep; and the same question must be raised for the goats. A second irregularity is the relatively (but not significantly) low scores of the sheep in the midrange of good adjustment, in contrast to the higher ESP scores of the sheep at the two ends of the good adjustment category. A third is the suggestively high deviation (if we do not correct for selection) of the goats with a very low number of check list en-

Table 10

Check List Entries and ESP Scores
(Subjects Tested Fall, 1945-1951)

	Number of Check List Entries	Number of Subjects: ESP Scores in Relation to Chance Expectation: Above	At	Below	Number of Runs	Deviation from Chance Expectation	Mean Hits per Run
A. Sheep							
	2	4	0	1	45	+13	5.29
	3	6	0	5	95	+14	5.15
	4	13	3	9	213	+44	5.21
	5	19	4	12	299	+24	5.08
	6	22	6	20	418	+13	5.03
	7	25	3	25	449	+8	5.02
	8	44	2	22	597	+134	5.22
	9	38	6	18	536	+183	5.34
	10	42	3	18	537	+177	5.33
	11	20	7	36	535	−72	4.87
	12	17	0	19	313	+10	5.03
	13	18	2	15	304	+18	5.06
	14	12	1	13	222	+6	5.03
	15	9	0	10	164	+11	5.07
	16	8	2	9	165	−5	4.97
	17	5	2	7	119	−10	4.92
	18	3	0	8	97	−32	4.67
	19-27	9	2	11	193	+8	5.04
	Total	314	43	258	5301	+544	5.10
B. Goats							
	2	1	0	5	54	−27	4.50
	3	5	0	3	72	+20	5.28
	4	8	1	7	142	−21	4.85
	5	13	2	7	186	+54	5.09
	6	14	2	14	262	−43	4.84
	7	10	3	16	252	−57	4.77
	8	11	3	20	298	−53	4.82
	9	10	1	27	336	−105	4.69
	10	13	1	27	360	−80	4.78
	11	28	0	13	352	+69	5.20
	12	19	0	18	317	+25	5.08
	13	10	1	9	173	+15	5.09
	14	5	0	10	130	−23	4.82
	15	8	1	10	160	+11	5.07
	16	9	0	2	98	+23	5.23
	17	4	2	4	87	−13	4.85
	18	3	0	2	45	+12	5.27
	19-28	3	0	4	59	−5	4.92
	Total	174	17	198	3383	−198	4.94

tries. Other minor irregularities and reversals also occur, but they appear from this table to be random.

Our first and major problem may be put as a question: Why was there such a marked discontinuity in ESP scores of subjects with 10 check list entries and subjects with 11? I shall discuss what seem to be the most plausible possible answers, giving what evidence is available to support or argue against them.

This abrupt change in the dependent variable may be characteristic of results obtained with the check list technique. It will be remembered that Munroe's own data showed a similar large shift between adjacent midrange categories (Table 3). Of her subjects with nine check list entries, 67% had adequate college adjustment; of her subjects with ten entries, only 17% had adequate college adjustment. The number of cases is small and the difference is only suggestive ($P = .05$); but the trend is even more marked than that of my data. When the difference between her pattern and mind is evaluated by chi square, $P = .36$; thus we need not consider them different.

These two cases are hardly enough to demonstrate that such a marked change in the dependent variable is characteristic of the midrange of check list totals, but I know of no other sources of data that bear on the question. Dr. Munroe was good enough to suggest one other research worker who might possess relevant material; but when I telephoned her and explained the problem she told me that none of her records would bear on the point. The data of Table 3, therefore, seem the only ones available for comparison with mine.

If, then, an abrupt change at the midrange is characteristic of these scores—and we cannot be sure that it is—we may wonder why it should be so. In any event, we may wonder why it appeared in these two sets of records. I shall suggest two possibilities. The first is the one discussed in Chapter II: that there may be a certain level of conformity, of fitting in with what other people do, that permits a person to feel at ease within his environment, and that a fair-sized number of deviations from the social norm may still be consistent with such feeling at ease. People who feel this way would presumably be likely to behave as they are expected to: to have satisfactory college records, to score well if they are sheep, etc. But perhaps there is a critical area where even a slight increase in the number of deviations from the norm will swing the balance to the other side. A person's feelings about himself, once the critical point has been passed, may be that he is different from his group rather than that he belongs in it. People with such feelings may be more likely to behave inconsistently with the demands made on them: to have inadequate college records, to score low if they are sheep, etc. If this is so, then Dr. Munroe's division at 53% and mine at 54% suggest that in our eastern colleges a small majority of undergraduates tend to feel at home in the college setting while a large minority tend to feel that they are in it but not of it. This is supported by Hartley's (1956) finding that only 55% of City College students reported that they planned to keep in touch with the college after leaving it. One would expect the proportion to vary for different types of groups.

Another explanation for the discontinuity lies in the attitude of the Rorschach examiner. (The two, of course, are not mutually exclusive.) Certainly records were scored with more care when they were in the neighborhood of the previously set dividing line between 10 and 11

than when they were unequivocally above or below this line. My feeling for the latter case was that a check more or less did not matter, since it would not change the classification; if I was slightly inaccurate there would be no harm done in testing the adjustment hypothesis as it was stated. But where I could see that it would be a close decision as to whether a subject would go in the well adjusted or the poorly adjusted group, I made a stronger, more consistent effort to come to the right decision for each individual entry. Did this additional concern result in some qualitative difference in my scoring of these borderline records? We can probably never recreate exactly what I did in the years between 1945 and 1951 while the records were scored; but I was curious enough about it to go to my files, pull out a large number of records, and see if I could at least make a good guess.

Having piled up a substantial number of protocols originally scored with either 10 or 11 check list entries, I tried to rescore them as if they were new. I have no notion of how much of the previous scoring was remembered at some level below that of recall; but there was no conscious recollection of which checks had previously been assigned. Before a dozen records were rescored it was clear that all contained some doubtful category where more than a mechanical judgment was required; whether this would be true for all college records I cannot say. The range of scores which it seemed to me might perhaps be defensible was two to five, in these first records, with a mode at three. In all but one, some argument could be made for setting the total at 10 or at 11. The Inspection Technique thus seems a flexible one; and this is how its originator described it.

But it is not illimitably elastic. In the first place, even though giving or withholding an entry could both be defended, there were many instances where one course of action had only a weak legalistic argument in its favor, and where in my opinion experienced examiners would have decided against it. In the second place, the same problems kept recurring. If the decision for one record was to assign an entry, then any examiner who made that decision would have assigned an entry in the corresponding instances for other records. Thus two examiners could differ in the absolute number of checks they gave, but show good agreement in the ranking of the records. My own conclusion from this was that Munroe's claim for the (limited) reliability and validity of the Inspection Technique are justified, though statistical extrapolations from those claims might not be. Her instructions deliberately permit the exercise of individual judgment; it is possible for two people to stay within the broad limits of those instructions and still to disagree on some points with each other. It is also possible, of course, for one person's opinions about the clinical significance of the protocol to change over the years; I would not have scored all those records this past winter just as I scored them a decade ago.

My next step was to sort the records into those which had been originally scored at 10 and those which had been scored at 11, and read over a number from each batch to see if any clear pattern emerged. My impression here was that the records which showed lively, impulsive vitality—sometimes misdirected, often associated with anxiety, aggression or guilt—had been scored at 10; and that the records indicating compulsive constriction and constraint had been scored at 11. This might be the key to the problem of why the pooled scores of 10 contrasted so markedly with the pooled scores of 11. In pondering

over the doubtful entries of the protocols near the 10-11 boundary line, my decisions may have been guided by the theory that more liveliness implied more flexibility and therefore better potential adjustment than constraint; and entries may have been assigned or withheld (within the framework of the Inspection Technique rules) according to this theory. If—as many previous studies have indicated—liveliness is associated with better ESP scoring than is constraint, the problem of why ESP scores at the 10 level were so much better than those at 11 has at least a possible answer. (Whether similar considerations about flexibility in relation to college adjustment guided Dr. Munroe in her assignment of 9 or 10 entries is a matter of which I am ignorant.)

Another suggestion to explain the discontinuity between ESP scores of 10 and 11 can probably be dismissed as ingenious but inapplicable. This is the proposition that the assigned check list totals were largely determined by my casual observation of the subjects (and my guesses about their ESP scores). In support of the proposition are the points that the subjects were in most cases my own students, and that the experiments were performed late in the semester, after we had had an opportunity to know each other. Thus if I were able to distinguish subjects likely to have psi success from those likely to show psi-missing (this of course is a big "if") there might have been some unconscious juggling of the check list entries to make them correspond with these unconscious predictions.

Let me add parenthetically here that no such mis-scoring of the records could have occurred consciously. Not only because it would be an absurd way to conduct a research project, though this is surely reason enough, but also because I have too healthy a respect for the Rorschach to tamper with it. Again and again, when students have come for interviews, I have prefaced some interpretation with, "This isn't at all what I thought about you, but the Rorschach says that . . ."—and almost always the student has told me that the test was right and my impression was wrong. The Rorschach seems to me to be an extraordinarily accurate diagnostic tool when it is carefully interpreted. Though it frequently does not give information of interest, it is almost always correct in what information it gives.

Dr. Karlis Osis has been kind enough to suggest that one objective check of this possibility that I unconsciously permitted my judgment of students to influence the entries, was to segregate the students who were not in my own classes, and examine their records to see if they also show a discontinuity of ESP success between check list entries of 10 and 11. Table 11 lists the relevant data. For both sheep and goats there is, between 10 and 11, a marked shift in ratio of subjects scoring above chance to subjects scoring below chance. Mean ESP scores of sheep decline abruptly between 9 and 10 entries, and continue low thereafter; mean ESP scores of goats have the same general pattern as those of the students in my classes. The two sets of data thus show a pronounced similarity, though they are not identical. It will be noted also that the odd rise of mean ESP scores of goats with 5 check list entries is present here, and also the U-shaped curve of ESP means of sheep in the range of 2-9 check list entries. On the whole, the data of subjects known and unknown to me seem too similar to warrant the assertion that my acquaintance with some of the students was a factor in the scoring of their Rorschach protocols.

Table 11

Check List Entries and ESP Scores of Subjects Who Were Not in the Author's Classes

	Number of Check List Entries	Number of Subjects: ESP Scores in Relation to Chance Expectation: Above	At	Below	Number of Runs	Deviation from Chance Expectation	Mean Hits per Run
A. Sheep							
	2	3	0	0	27	+18	5.67
	3	2	0	1	27	+2	5.07
	4	3	0	2	45	+17	5.38
	5	2	1	6	80	−16	4.80
	6	4	1	11	144	−38	4.74
	7	4	1	5	90	+6	5.07
	8	11	0	5	144	+47	5.33
	9	7	2	3	108	+43	5.40
	10	7	1	4	108	−3	4.97
	11	5	1	10	141	+1	5.01
	12	4	0	3	63	+3	5.05
	13	4	0	6	90	−12	4.87
	14	3	0	3	53	+5	5.09
	15	4	0	0	36	+26	5.72
	16	2	1	2	45	−8	4.82
	17	0	0	1	9	−10	3.89
	18	0	0	2	18	−11	4.39
	19-27	4	0	2	54	+23	5.43
	Total	69	8	66	1282	+93	5.07
B. Goats							
	2	0	0	3	27	−26	4.04
	3	5	0	1	54	+26	5.48
	4	6	0	5	99	−10	4.90
	5	6	1	5	108	+24	5.22
	6	6	1	8	135	−39	4.71
	7	4	0	8	108	−46	4.57
	8	4	0	8	108	−21	4.81
	9	7	0	17	215	−66	4.69
	10	4	1	6	99	−3	4.97
	11	13	0	2	135	+54	5.40
	12	4	0	10	126	−20	4.84
	13	4	1	4	81	+15	5.19
	14	2	0	5	63	−13	4.79
	15	3	0	3	54	−6	4.89
	16	5	0	0	45	+15	5.33
	17	1	0	0	9	+4	5.44
	18	1	0	0	9	+12	6.33
	19	1	0	0	9	+3	5.33
	Total	76	4	85	1484	−97	4.93

Our most pressing problem is now to consider whether the entire difference between ESP scores of subjects with 10 entries or fewer and subjects with 11 entries or more, results from the striking difference noted just at the line of demarcation. If so, the most probable interpretation is that the ESP difference is not due to factors of adjustment, but rather to the distinction between records with more liveliness and records with more constraint. If not, the stated hypothesis about adjustment is supported by the data.

As will be described in the following chapter, the method employed for the formal test of the adjustment hypothesis was essentially to compare (approximately) the quarter of the subjects with the lowest number of check list entries, the half with a moderate number and the quarter with most, thus throwing check list entries of 10 and 11 into the same middle category. Another method of examining the pattern of the data is shown in Figure 1. Here it is apparent that for the bulk of subjects with many check list entries there was no superiority of sheep scores over goat scores, so that the sheep-goat hypothesis does not hold true, under our experimental conditions, for subjects whose social adjustment is poor.

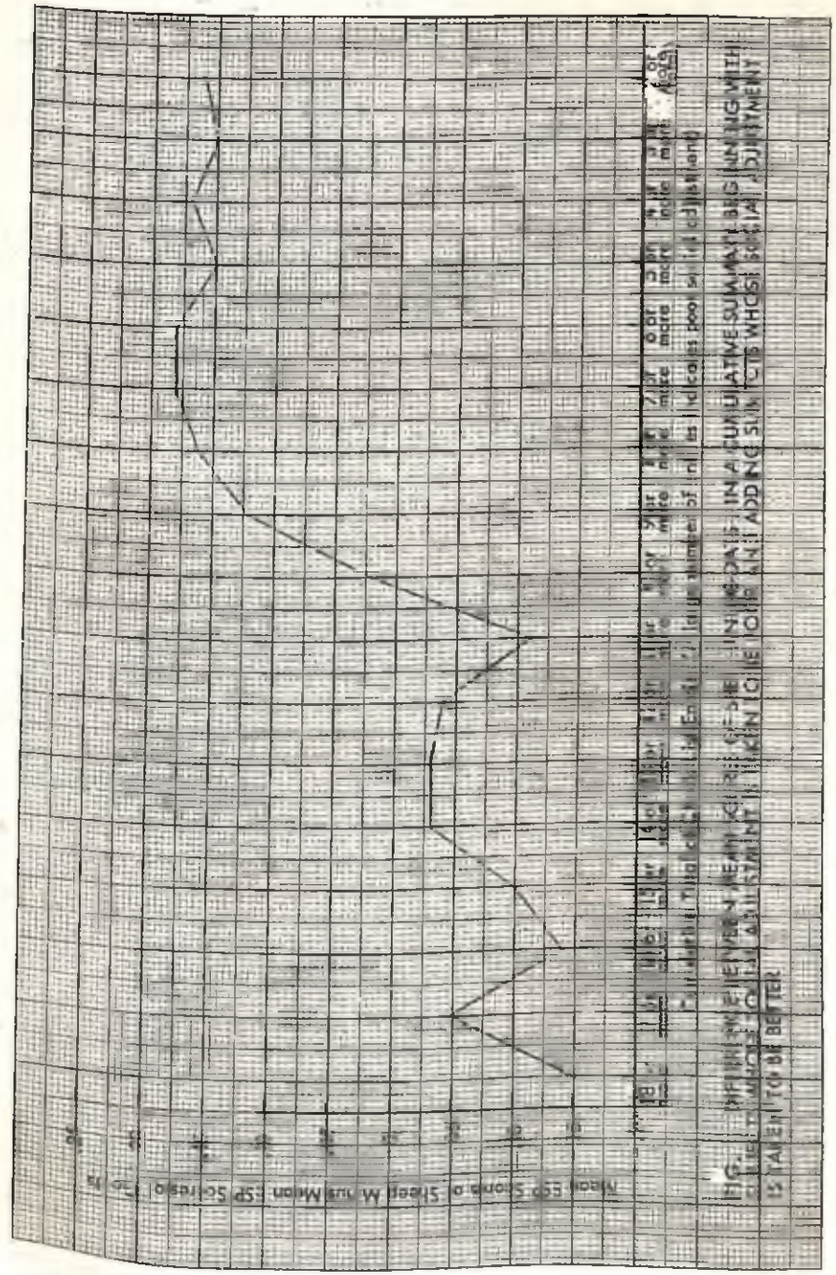
The second question with which Table 10 confronted us was to explain why the mean ESP scores of the sheep should be lower for those with 5-7 check list entries than for those with 2-4 or for those with 8-10. No clear answer has occurred to me. These differences may be random ones; or they may reflect specific patterns of good, better or only fair adjustment which need a more careful qualitative analysis than it has been possible for me to make. Each of the check list entries is, after all, qualitatively different from the others; perhaps important insights have been sacrificed with my emphasis on totals rather than syndromes.

The third of the problems presented by Table 10 is the relatively high ESP scores of the goats with few check list entries. This also may be only a random difference, for the scores of these subjects are not significantly higher than those of the goats with a somewhat larger number of entries when correction is made for selection. A suggestion to account for these differences has however already been proposed: that for a good many of the goats whose social adjustment is *extremely* good, cooperation with the instructor (which has yielded so many rewards of praise or high grades in the past) is prepotent over distaste for the particular task which the instructor assigns.

Differences between Male and Female Subjects

After the analysis of check list totals and ESP scores had shown that important variables not previously specified were influencing the scores, Mr. J. Fraser Nicol and Dr. Betty Nicol, with the support of the Parapsychology Foundation, Inc., were good enough to examine the data in detail. Among their many helpful suggestions were two to be described below: that the sex of the subjects, and the number of runs made by the subjects, be investigated in relation to the other variables. The Nicols demonstrated, for the former of these, a striking difference significant at the level of $P = .0015$, between the mean number of check list totals for male and female sheep, among the 754 subjects who constituted the test group for the sign hypothesis. The females had fewer check list entries than the males.

It could be argued that some complicated factor of sex in relation



to sheep-goat choice was operative here: that it was the better adjusted females who chose to be sheep, and the better adjusted males who chose to be goats, so that the difference between male and female sheep would not correspond to differences between males and females in the entire sheep-goat sample. I therefore calculated the difference between all males and females in the group of 754 subjects, and also in the entire experimental population (Table 12). It is apparent that

Table 12

Total Number of Entries on the Check List for Males and Females
(Subjects Tested Fall, 1945-1951)

	Subjects Tested Fall, 1946-1951		All Subjects	
	Males	Females	Males	Females
Number of Subjects	569	185	708	296
Mean of Check List Entries	10.33	9.63	10.02	9.73
Difference	.70		.29	
P_{diff}	.03		.28	

the results for the group of 754 subjects were similar to those found by the Nicols. The trend for the entire group was not significant.

If we continue to assume that check list entries are a measure of social adjustment, these results can probably be explained fairly readily in terms of the populations from which our subjects were drawn. Let us consider the latter 754 subjects. All except 14 of the 185 girls were students at uptown City College, during the years that girls were admitted to the School of Education at this branch of the college, but not to the division of liberal arts. This means that almost all the girls were preparing to teach in elementary or high schools. The men represented a more typical cross section of college undergraduates. It is to be expected, I think, that high school graduates who choose to be school teachers and are willing to take the prescribed curriculum for such teaching, should be a group that conforms more readily to social demands than most other undergraduates; thus the better social adjustment that was found for the females in this sample is not surprising.

Of the 111 girls tested earlier, 35 were enrolled at uptown City College, and thus presumably in the School of Education. Others were enrolled at the School of Business and Civic Administration at City College, or in evening classes (some as matriculated and others as non-matriculated students) at Hunter College or City College. Their age range was wide, and their interests varied even more widely.

Among the men in this earlier sample, a few were non-matriculated evening session students, and the rest were undergraduates at City College. A very large number were veterans, who had elected to continue their formal education. It was the consensus among those of us who taught such veterans that they tended to be far more mature and competent than most undergraduates; one would therefore expect them to have a lower number of check list entries. It looks as if the

high proportion of male veterans in the earlier sample balanced out the high proportion of female prospective school teachers in the later sample, and that this is why there is no significant difference in mean number of check list entries for the males and females taken as a whole.

These findings, and these considerations, raise some interesting questions about the uncontrolled social variables in the research. Would we have different adjustmental patterns, different reactions to the ESP task, and different patterns of ESP scores for the girls who were well pleased with their choice of a teaching career as opposed to those who regretted it? Were the veterans' reactions to the ESP task as different from those of the younger men and women as their reactions to the Rorschach seem to have been? What of other differences in status or interest or past experience that cut across sex lines? Questions could be multiplied; and though a search of the college records might be able to give data with which to answer some few of them, I do not now have available the necessary information about which men received an honorable discharge from the armed forces, and which girls went on to teach, which would permit making a beginning along these lines.

But leaving all these problems aside, the demonstrated difference between total number of check list entries for males and females in the test group for the sign hypothesis makes it questionable whether it is legitimate statistically to pool the scores of the two sexes. A strong argument could be adduced for separating their data. This has accordingly been done in the formal analysis which will be reported in the following chapter.

Differences between Subjects with 8-Run and with 9-Run Procedures

Another interesting point brought forth by the Nicols was that ESP scores tended to be higher for the subjects whose targets were colored cards. Let me describe in a little more detail the two major procedures for the ESP tests.

The method that was used for the last four series of individual tests, and for the group tests from their beginning through 1949, consisted of asking subjects to respond to three runs of 25 items each, where the items were the conventional ESP symbols. After each such unit of 75 responses there was a pause, usually filled with some psychological test or questionnaire. Three of these units, comprising nine runs and thus responses to 225 serial items, constituted a session.

Many subjects complained that the long series of guesses was dull. In the last year and a half of the project a procedural change was introduced which, it was hoped, would make the session less tedious. Target items now represented colored ESP cards. Five colors were used. The order of colors was randomized independently of the order of the ESP symbols, so that pairings of color and symbol were random. Subjects were instructed to guess at the color and symbol of each item; thus they made fifty responses for each list of 25 double targets. Hits on colors were scored just as hits on symbols were and no difference between tendency to succeed on one or the other has been observed. The standard number of runs was reduced from nine to eight. With this procedure subjects seemed to feel as if they were responding to 100 units (instead of 225); but the number of scorable responses was

reduced only to 200 (instead of 225). A different task was given the subjects after each set of 25 double items.

When the Nicols studied the data, they observed that ESP scores tended to be higher for subjects with this 8-run procedure than for subjects with the previous 9-run procedure. Omitting all subjects who did not complete the specified number of runs, and pooling scores of sheep and goats, and utilizing only the records of the latter 754 subjects, the Nicols found the mean of the 9-run subjects was 5.020 and the mean of the 8-run subjects was 5.165. The difference between these means was significant at the level of P = .024. Table 13 presents the

Table 13

ESP Scores of Subjects with Nine-Run Group Procedure and with Eight-Run Group Procedure
(Subjects Tested Spring, 1945-1951)

	Eight-Run Procedure			Nine-Run Procedure		
Subjects	Number of Subjects	Number of Runs	Mean Hits per Run	Number of Subjects	Number of Runs	Mean Hits per Run
Sheep	110	824	5.26	537	4768	5.08
Goats	83	638	5.01	332	2981	4.92
Total	193	1462	5.15	869	7749	5.02

data of all subjects in the current project, and shows that adding the extra cases to the Nicols' computations does not alter the pattern they uncovered. The implication of this pattern is, of course, that the subjects tend to make higher ESP scores when the procedure is more interesting.

Clearly another cross-cut of the data would now be desirable, in terms of either the procedure (colored cards vs. standard symbols) or else the number of runs completed by the subjects. But there are technical disadvantages to introducing this. A major one is that we have already incurred the obligation to analyze the data of the latter 754 subjects according to sheep-goat classification, sex, some measure of number of check list items, and the presence or absence of signs. Adding a fifth classification would reduce to a seriously low level the number of subjects in certain of the cells.

Fortunately a preliminary analysis indicated that separate treatment of 8-run and 9-run subjects was not obligatory. This analysis was performed by Mr. Edward Turner, through the good offices of Dr. Gardner Murphy. After preliminary discussions with me (which later turned out to have been incomplete) Mr. Turner performed two analyses of variance. The first was made for the 157 subjects of the 8-run procedure who made eight runs, and examined sheep-goat classification, sex, adjustment (with subjects divided into those who had 2-8 check list entries, 9-11 check list entries and 12 or more check list entries) and protocols with signs contrasted with protocols free from signs. The second was made for the 511 subjects in the test group for signs who had completed nine ESP runs. It used the same classifications

except that check list entries were divided into four classes: 2-7, 8-9, 10-12 and 13 or more. The two analyses gave similar results, of which the only significant values were the sheep-goat division (P = .005) and the sheep-goat vs. signs or no signs interaction P = .001). Because the detailed patterns shown in these two analyses were so much alike, it seemed unnecessary to continue separating subjects on the basis of the number of runs they had performed, while testing for either the adjustment or the sign hypothesis.

VI. FORMAL ANALYSES OF DATA ON SOCIAL ADJUSTMENT AND SIGNS

It is usually considered desirable to test a hypothesis according to the procedure determined before the data were collected. When this was done for our hypothesis about social adjustment and ESP scores (Table 6) significant differences in the predicted direction were found. However it later became clear that there were strong reasons for making two changes in the statistical treatment originally planned. These changes and the reasons for them were discussed in the preceding chapter, but I shall repeat them briefly here, and then describe the final formal statistical test.

The original hypothesis (that sheep with better social adjustment would have higher ESP scores than other sheep and that goats with better social adjustment would have lower ESP scores than other goats) was stated in terms of a dichotomy of adjustment. Rorschach protocols were scored and check list entries from the Rorschach (our measure of social adjustment) were made with this dichotomy in mind. Re-examination of the scoring indicates that for many records which were immediately above or below the pre-set dichotomizing line, and for which there were some doubtful entries, the doubt was usually resolved in favor of good adjustment scores for the records which gave an over-all impression of liveliness, and in favor of poor adjustment for the records which gave an over-all impression of constraint. There is no reason to believe that such impressions influenced adjustment scores markedly, elsewhere. Examination of ESP scores immediately above and below the pre-set dichotomizing line shows a marked and significant shift of scores there. These two facts taken in conjunction suggest that in testing the presence of a relation between ESP and adjustment scores, some division other than the original one be made, in order that decisions about liveliness and constraint, made just above and just below the pre-set boundary, should not be confounded with some marginal decisions about social adjustment.

The second change in the original design was made because post hoc analysis showed a highly significant difference between adjustment scores of males and females. It is questionable, in view of this difference, whether scores of males and females should be pooled. We have therefore treated them separately.

Alternative methods for solving the first problem are to omit the scores near the pre-set boundary or to utilize all records but group them into different categories so that the original division may be effaced. At the advice of Dr. Jacob Cohen, the latter was selected. Subjects were divided as nearly as possible in the ratio of 1:2:1, according to the number of their check list entries. Dr. Cohen had originally proposed this division because it is a common one in psychological research. It corresponds in general to dividing subjects into extreme and middle groups, or into superior, average and inferior, without reducing to an unreasonably small number the groups that are more interesting. A further advantage specific to this particular project is that the 1:2:1 boundaries fall between 7 and 8 check list entries and between 12 and 13. This probably ensures that all or almost all the subjects where my judgment in terms of the pre-set boundary influenced the allocation of check list entries would fall in the middle group of 8-12 check list entries. (It seems unlikely that there would be many subjects to whom I assigned 10 checks because of the liveliness of their records who would have been given as many as 13 checks by a competent examiner; and it is even less likely that many subjects to whom I assigned 11 checks because their records were inhibited would have been given as few as 7 checks by someone familiar with the Inspection Technique.) Chance variations would undoubtedly on rescoring change some subjects from one of these three groups to an adjacent one; but the systematic variation I suspect to have been present is presumably nullified when Dr. Cohen's advice is followed.

The results of the formal test of the adjustment hypothesis are given in Tables 14 and 15. They show a highly significant difference between sheep and goats, with which we are already familiar, and an interaction between number of check list entries and sheep-goat categories, significant at the level of $P < .001$. It is clear from Table 15 that the

Table 14

Analysis of Variance of ESP Scores for Sheep-Goat Category, Sex and Social Adjustment (Low, Average or High Totals of Check List Entries)*
(Subjects Tested Fall, 1945-1951)

Source	Degrees of Freedom	Sum of Squares	Mean Sum of Squares	Corrected Mean Sum of Squares**	F	P
Sheep-Goat	1	5.7049	5.7049	345.0352	27.079	< .001
Male-Female	1	0.4548	0.4548	27.5065	2.159	not signif.
Check List Totals 2-7, 8-12, 13+	2	0.3437	0.1718	10.3905	.815	not signif.
Sheep-Goat × Male-Female	1	1.9168	1.9168	115.9290	9.098	<.005
Sheep-Goat × Check List Totals	2	3.4236	1.7118	103.5305	8.125	<.001
Male-Female × Check List Totals	2	1.2822	0.6411	38.7740	3.043	<.05
Sheep-Goat × Male-Female × Check List Totals	2	0.2295	0.1148	6.9432	.545	not signif.
Within Cells	992	12640.0097	12.7419			

* Snedecor's method of unweighted means was used for this analysis (Snedecor, G. W. Statistical Methods, 5th ed. Ames: Iowa State Coll. Press, 1956).
** Harmonic mean of subclass numbers is 60.4805.

Table 15

Mean ESP Scores (Prorated for Nine Runs) of Sheep-Goat, Male-Female and Social Adjustment Subgroups
(Subjects Tested Fall, 1945-1951)

A. Sheep × Male-Female

Subjects	Males	Females
Sheep	45.85	46.24
Goats	44.80	43.98
Sheep Minus Goats	1.05	2.26

B. Sheep-Goat × Social Adjustment (Low, Average or High Totals of Check List Entries)

Subjects	2-7 Entries	8-12 Entries	13+ Entries
Sheep	45.72	46.56	45.05
Goats	44.41	44.29	45.24
Sheep Minus Goats	1.31	2.27	−0.19

C. Male-Female × Social Adjustment (Low, Average or High Totals of Check List Entries)

Subjects	2-7 Entries	8-12 Entries	13+ Entries
Males	45.10	45.70	45.43
Females	45.51	45.57	44.36
Males Minus Females	−0.41	−0.13	1.07

subjects with poorer social adjustment do not contribute to the higher mean scores of the sheep compared to the goats. This was foreshadowed in the visual presentation of Figure 1.

The Nicols' suggestion of sex differences in scoring patterns is also strongly supported. The difference between sheep and goats is greater for females than for males, and the interaction is significant at the level of $P < .005$. There is also a suggestive tendency for the females with fewer check list entries to score higher, and those with more check list entries to score lower than the males ($P < .05$).

The formal test of the sign hypothesis was made by computing the analysis of variance, on the 754 test subjects, for sheep-goat category, sex, check list entries 2-7, 8-12 and 13 or more, and for presence or absence of any of the signs (Table 16). Here the difference between sheep and goats was significant at the level of $P < .005$, and the interaction between the sheep-goat category and the signs category was significant at the level of $P < .001$. This hypothesis is therefore also strongly supported. The mean ESP scores of the relevant subgroups are given in Table 17. No other significant differences were present.

Both of the formal hypotheses about Rorschach scores in relation to ESP scores have thus been confirmed at a satisfactory level of significance ($P < .001$) when the data of all subjects in each test group

Table 16

Analysis of Variance of ESP Scores for Sheep-Goat Category, Sex, Social Adjustment (Low, Average or High Totals of Check List Entries) and Presence or Absence of Any of Seven Specified Rorschach Signs*
(Subjects Tested Fall, 1946-1951)

Source	Degrees of Freedom	Sum of Squares	Mean Sum of Squares	Corrected Mean Sum of Squares**	F	P
Sheep-Goat	1	20.9104	20.9104	336.6240	8.381	<.005
Male-Female	1	2.8608	2.8608	46.0543	1.147	not signif.
Check List Totals 2-7, 8-12, 13+	2	10.5236	5.2618	84.7066	2.109	not signif.
Presence or Absence of any of seven signs	1	1.9064	1.9064	30.6900	0.764	not signif.
Sheep-Goat × Sex	1	0.3128	0.3128	5.0356	0.125	not signif.
Sheep-Goat × Check List	2	9.9271	4.9636	79.9060	1.989	not signif.
Sheep-Goat × Signs	1	40.2227	40.2227	647.5211	16.122	<.001
Sex × Check List	2	11.5537	5.7768	92.9972	2.315	not signif.
Sex × Signs	1	0.1370	0.1370	2.2055	0.055	not signif.
Check List × Signs	2	2.5818	1.2909	20.7814	0.517	not signif.
Sheep-Goat × Sex × Check List	2	0.1043	0.0522	0.8403	0.021	not signif.
Sheep-Goat × Sex × Signs	1	2.0091	2.0091	32.3433	0.805	not signif.
Sheep-Goat × Check List × Signs	2	1.0953	0.5476	8.8155	0.219	not signif.
Sex × Check List × Signs	2	1.9315	0.9658	15.5478	0.387	not signif.
Sheep-Goat × Sex × Check List × Signs	2	9.4138	4.7069	75.7736	1.887	not signif.
Within Cells	730	29319.3295	40.1635			

* Snedecor's method of unweighted means was used for this analysis (Snedecor, G. W. Statistical Methods, 5th ed. Ames: Iowa State Coll. Press, 1956).

** Harmonic mean of subclass numbers is 16.0984.

Table 17

Mean ESP Scores (Prorated for Nine Runs) of Sheep-Goat and Rorschach Sign Subgroups
(Subjects Tested Fall, 1946-1951)

Subjects	Protocols without Any Sign	Protocols with One or More Signs
Sheep	47.19	45.34
Goats	42.89	46.07
Sheep Minus Goats	4.30	−0.73

are examined. How far does this advance us toward our goal of understanding the pattern of ESP success and failure? Only a short step, for there is still so much overlap between groups of subjects, and such wide divergences within each group, that it is impossible to predict with confidence how any individual will score, even if he is tested under conditions like those of our experimental groups. The confirmation implies, however, that the work is heading in something of the right direction. It justifies further inferences along the same lines as these; and it gives us a base from which we can try with some hope to develop better theories and testing procedures. The following chapter attempts to take further steps in the same direction.

VII. SUGGESTED INTERPRETATIONS OF THE RELATION OF RORSCHACH SIGNS TO ESP SCORES

Let us recapitulate. From a mechanical listing of Rorschach scores and ESP scores, a half dozen Rorschach items were found in my preliminary group to be associated with relatively low ESP scores for sheep and with relatively high ones for goats. The formal hypothesis was stated that these and a seventh which I believed to be logically related to one of them would continue to be associated with relatively low ESP scores for sheep and high ones for goats. This hypothesis was confirmed in the experimental group at such a high level of significance that it deserves attention. What can we make of it?

The starting point for explanations should be the usual interpretation of these signs in Rorschach protocols, brought into congruence with what understanding we have of the dynamics of ESP success. I shall rough out the suggestions that seem most reasonable. But before doing so, must give a blanket warning—or apology—for the chapter as a whole. With one exception, theories in it derive from the data and therefore are not confirmed by the data. If my wording sometimes suggests the contrary, I trust the reader will mentally revise it for me, instead of taking it as deliberate.

Analysis of Subjects with R+ or Total Movement++

The first on the list of seven signs was R+, entered when the subject gave an unusually large number of responses. Rorschach analysts sometimes interpret it as "quantity ambition"; in a college population it probably indicates that the student who shows it has a need for conspicuous intellectual achievement. In the group Rorschach, where available time is short, it also implies intellectual liveliness. Another of the seven signs, Total Movement++, is an indicator of marked mental liveliness or what is sometimes called an "active inner life"; and this also, in a college population, seems often associated with intellectual interests and a concern for intellectual achievement. The two signs have somewhat different implications and neither can properly be interpreted out of the context of the record as a whole; but since they converge on the same area and the same interpretation, for undergraduate subjects, we probably can pair them without impropriety in our search for their meaning in relation to group ESP scores.

Years after the data had been gathered, the work of two other investigators suggested to me a hypothesis about these two signs. Waldron (1958) has developed the thesis that where ESP tests are administered to college students by a college instructor (who, for students while they are in a college atmosphere, is associated with the outward signs of achievement) need for achievement might be

intimately related with attitude toward the instructor, hence with attitude toward the task the instructor assigns them, and hence with ESP scores. Anderson (1959) has described data showing higher ESP scores for subjects with higher school grades in the fifth and sixth grades and in high school. Putting these two points together, a natural conclusion is that, all other things being equal, college students tested by their own instructor who have a high need for achievement will make higher ESP scores if their grades in the course have been high than if their grades in the course have been low. The causal pattern might be that students who succeed in one task (getting good grades) can so mobilize their abilities as to be able to succeed in another (getting good ESP scores); or it may be that the low morale resulting from low grades—to those for whom grades are important—will carry over to low morale, and therefore low hitting rate, in a classroom ESP task. (For students without the marked need for intellectual achievement, grades would be a poorer indicator of either morale or utilization of ability. Such factors as the intrinsic interest of the material, its personal relevance, or reactions to the instructor might well be far more important.)

This led directly to a testable formulation. Leaving out the goats, where special factors of resistance or ambivalence are probably present, and taking R+ or Total Movement++ as indicators of a high need for achievement in this population, and considering only the students in my own classes, the new proposition is: Of the sheep with R+ or Total Movement++, those whose grades in my class have been high will have higher ESP scores than those whose grades have been low.

To test this new proposition I went back to my roll books, and found that a record of grades received during the semester had been retained from 1948 through 1951. There were 513 ESP subjects who had taken the Rorschach in this period, 101 of whom were sheep with R+ or Total Movement++. These 101 subjects therefore constituted the test group.

Two procedural questions immediately presented themselves. What were the grades at the time of the ESP test? And what are "high" and "low" grades? I could not answer either with precision, the first because my roll book did not date each grade, and the second because these value judgments vary with the individual. But approximate answers could be found. The ESP tests were usually given near the end of the semester, partly to allow me time to complete the Rorschach interviews and scoring, and partly because the classes seemed to be friendlier and more at ease after we had had more time together. Therefore the best estimate of grades at the time of the ESP test would be the average of grades before the final examination. As to when grades are "high" or "low," a reasonable pattern seemed to be to make a 1:2:1 division.

This sets the boundary of high grades just above the B level, and includes among high grades both an average of A and an average intermediate between A and B. The boundary for low grades falls between an average of C and an average intermediate between B and C. These check with common sense estimates of student reaction. Most would not be dissatisfied if they had, in their own phrase, "a chance of an A in the course"; and few with a strong need for achievement would be well pleased with a C or less in psychology.

Table 18 summarizes the data of the group with which we are primarily concerned, and includes for comparison the data of the other groups. The pattern seems by inspection to be consistent with the proposition that is being tested. Analysis of variance shows that the hypothesis is supported at the significance level of P = .05 (Table 19).

Table 18

ESP Scores of Sheep and Goats Divided According to Presence of Signs Indicating a Strong Need for Intellectual Achievement (R+ or Total Movement++) and Grades in the Class Where They Took the ESP Tests
(Subjects Tested 1948-1951)

Subjects	Number of Subjects: ESP Scores in Relation to Chance Expectation — Above	At	Below	Number of Runs	Deviation from Chance Expectation	Mean Hits per Run
A. Sheep with Evidence of a Strong Need for Intellectual Achievement (R+ or Total Movement++)						
High Grades	16	2	8	216	+24	5.11
Average Grades	29	5	25	509	+23	5.05
Low Grades	5	0	11	132	−35	4.73
B. Sheep without R+ or Total Movement++						
High Grades	29	4	29	511	−25	4.95
Average Grades	75	14	50	1151	+203	5.18
Low Grades	25	4	23	440	+63	5.14
C. Goats with Evidence of a Strong Need for Intellectual Achievement (R+ or Total Movement++)						
High Grades	4	1	4	76	−5	4.93
Average Grades	14	1	10	209	+11	5.05
Low Grades	4	0	3	58	+8	5.14
D. Goats without R+ or Total Movement++						
High Grades	9	4	11	195	−25	4.87
Average Grades	24	3	36	526	−101	4.81
Low Grades	13	2	12	213	−2	4.99

No such pattern appears for either sheep or goats without R+ or Total Movement++ in their Rorschachs. The results of the handful of goats with R+ or Total Movement++ and high or low grades are hard to interpret. Tests by chi square show no relation between grades and the presence or absence of R+ or Total Movement++, nor between frequency of sheep and goats with these signs for high and low grades.

It is interesting to make a further division by sex, for the subjects

Table 19

Analysis of Variance of ESP Scores of Subjects with High Grades and with Low Grades, among Those Sheep Who Had Signs Indicating a Strong Need for Intellectual Achievement
(Subjects Tested 1948-1951)

Source	Degrees of Freedom	Sum of Squares	Variance Estimate
Between	1	132.37	132.37
Within	40	1109.77	27.74
Total	41	1242.14	

$F = 4.77$; $P < .05$

Table 20

ESP Scores of Sheep and Goats with Rorschach Signs Indicating a Strong Need for Intellectual Achievement (R+ or Total Movement++) Divided According to Sex and to Grades in the Class Where They Took the ESP Tests
(Subjects Tested 1948-1951)

Subjects	Number of Subjects: ESP Scores in Relation to Chance Expectation — Above	At	Below	Number of Runs	Deviation from Chance Expectation	Mean Hits per Run
A. Male Sheep						
High Grades	11	2	7	164	+7	5.04
Average Grades	25	2	19	399	+25	5.06
Low Grades	4	0	7	93	−20	4.78
B. Female Sheep						
High Grades	5	0	1	52	+17	5.33
Average Grades	4	3	6	110	−2	4.98
Low Grades	1	0	4	39	−15	4.62
C. Male Goats						
High Grades	4	1	2	59	+9	5.15
Average Grades	9	1	5	125	+37	5.30
Low Grades	2	0	1	25	+5	5.20
D. Female Goats						
High Grades	0	0	2	17	−14	4.18
Average Grades	5	0	5	84	−26	4.69
Low Grades	2	0	2	33	+3	5.09

with evidence of the strong need for achievement (Table 20). The mean scores hint that the postulated effect was more strongly present among the girls than among the men. This may be meaningful in the special conditions at City College, where there were very few women on the teaching staff. It seems not unlikely that girls with a strong need for intellectual achievement, taking a course with a woman instructor, would tend to have a more personal reaction to their good or poor grades than in other courses; they might be more ready to identify with her if they were doing well in her course, and more resentful if their grades were low. But the data are too slight to force us to such an interpretation.

We can conclude that the suggestion which was derived from Anderson's and Waldron's work is supported at a moderate level of confidence by these results. This, then, very tentatively, could suggest one reason why R+ and Total Movement++ have a place in the list of signs: each implies that factors other than the sheep-goat decision will be influential in deciding the attitude toward ESP success.

Table 21 presents a summary of all test subjects with R+ or Total Movement++. The mean of the sheep, 5.08, is almost the same as 5.10, the mean of sheep without these signs; however the mean of the goats, 5.17, is substantially higher than 4.97, the mean of the goats without these signs. Could there be a tendency for these subjects, presumed to have a strong need for achievement, to try to achieve even when they do not sympathize with the imposed task? Could there be some relation between the intellectual liveliness implied by these signs and ESP success? The latter seems a promising possibility to me (especially since it does not rule out the former); and I shall return to it in the final section of this chapter.

Analysis of Subjects with F%+, Mr or No Shock

Three of the seven signs (a high proportion of form responses, rigid human movement and absence of either color or shading shock when both are scored very lightly) all imply a lack of free responsiveness, although there are subtle differences between them and though—as has been said before—no one of them can properly be interpreted out of context of the record as a whole. Let us therefore group these three together, in order to try to interpret their contribution to the discriminative list of signs.

Assuming that the Rorschach is a valid personality test, and that we are justified for most subjects in interpreting any of our three signs as evidence of some lack of free responsiveness, what inferences should we make about these unresponsive subjects? The most obvious would be that they are intellectually lethargic, emotionally deadened, apathetic, functioning subnormally—but though such a literal conclusion might apply to some groups, such as the senile or the feeble-minded or deteriorated schizophrenics, it must be ruled out for a college population. If then the lack of responsiveness is not a direct expression of dulled function, the next possibility is that it represents a marked inhibition of response tendencies. This inhibition would hold in check the liveliness which is potentially present, but which for some reason the subject is fearful of expressing. (Everyone, of course, must sometimes inhibit his response tendencies if he is to live effectively in the world; the implication of our three signs, in a normal population, is

Table 21

ESP Scores of Sheep and Goats Whose Rorschach Protocols Have R+ or Total Movement++
(Subjects Tested Fall, 1946-1951)

Subjects	Number of Subjects: ESP Scores in Relation to Chance Expectation: Above	At	Below	Number of Runs	Deviation from Chance Expectation	Mean Hits per Run
A. Sheep						
R+ or Total Mvt++, no other sign						
Only R+	18	4	18	348	+20	5.06
Only Total Mvt++	23	5	21	424	−5	4.99
Both R+ and Total Mvt++	2	0	1	27	+3	5.11
R+ or Total Mvt++ and also F%+, Mr or no shock (in any combination) but without CF+ or C+	27	2	22	447	+37	5.08
R+ or Total Mvt++ and also CF+ or C+ (in any combination)	13	1	6	171	+65	5.38
All sheep with R+	46	6	40	802	+79	5.10
All sheep with Total Mvt++	39	6	30	650	+40	5.06
All sheep with R+ or Total Mvt++	83	12	68	1417	+120	5.08
B. Goats						
R+ or Total Mvt++; no other sign						
Only R+	7	0	6	114	+18	5.16
Only Total Mvt++	21	1	9	264	+85	5.32
Both R+ and Total Mvt++	2	0	2	35	−3	4.91
R+ or Total Mvt++ and also F%+, Mr or no shock (in any combination) but without CF+ or C+	8	0	13	184	−25	4.86
R+ or Total Mvt++ and also CF+ or C+ (in any combination)	7	1	1	79	+36	5.46
All goats with R+	16	1	14	274	+25	5.09
All goats with Total Mvt++	31	1	19	437	+83	5.19
All goats with R+ or Total Mvt++	45	2	31	676	+111	5.16

that the inhibition is stronger than usual.) Let us try to apply this interpretation to our data, and see if it helps put them in order.

We note initially that the earlier group of signs, R+ and Total Movement++, indicate *more* "inner" or intellectual activity than is usual. And to anticipate the following section: the third group of signs, CF+ and C+, indicate more outward-oriented or emotional activity than is usual. Perhaps with subjects who have either of these tendencies, more inhibition than usual represents something of a healthy counterbalance: a more nearly average—though precarious—picture than if either liveliness or inhibition was present alone.

In Table 22 are listed separately subjects who have F%+, Mr or no shock with no other sign, and subjects who have one or more of these along with a sign of overactivity. For comparison, let us note that the mean ESP scores of subjects without F%+, Mr or no shock are 5.15 for the sheep and 4.96 for the goats. Subjects with both F%+, Mr or no shock and a counterbalancing sign have as mean ESP scores 5.09 for the sheep and 4.83 for the goats. The difference between sheep and goats is approximately the same for them and for the comparison group. But for the subjects with F%+, Mr or no shock and *no* counterbalancing sign, the mean scores are 4.98 for the sheep and 5.09 for the goats. This looks as if subjects with some evidence of a balanced pattern of inhibition and overactivity have contributed to the sheep-goat effect, but subjects with evidence of inhibition and no evidence of overactivity have not contributed to it.

The reader will remember that, after the retrospective examination of Rorschachs with 10 or 11 check list entries, I came to the conclusion that records whose over-all impression was one of liveliness were more often assigned the score of 10; those whose over-all impression was one of constraint were more often assigned the score of 11. If this is so, the liveliness-constraint interpretation should supersede the sign interpretation, because over-all impressions of the protocol are more likely to be valid than single scores taken out of context. Therefore a record scored 10 (if it was scored 10 for this reason) but containing F%+, Mr or no shock may tentatively be considered lively but balanced; one scored 11 (for this reason) and containing one of these signs can tentatively be considered overcontrolled.

Let us turn to those records. We find that among the sheep who had F%+, Mr or no shock, 24 were given 10 check list entries and had an ESP mean of 5.46; 26 were given 11 check list entries and had an ESP mean of 4.74. Among the goats with these signs, 10 were given 10 check list entries and had an ESP mean of 4.60; 11 were given 11 check list entries and had a mean ESP score of 5.15. Thus the retrospective interpretation for the 10-11 gap ties in with the present interpretation of the effect of balanced and unbalanced patterns including inhibition on ESP scoring.

If we draw the conclusion from these data that subjects who are overcontrolled are not likely to succeed at ESP tests of the type we are describing, it has the merit of being consistent with a good deal of other work. Humphrey's expansion-compression dimension presumably relates to the same personality factor; and Humphrey (1951) reports that compressives had low ESP scores in clairvoyant picture drawing although not in GESP picture drawing (the great majority of our tests were of the clairvoyant type) and also that compressives tend to score low at ESP card tests. Nicol and Humphrey (1955) found

Table 22

ESP Scores of Sheep and Goats Whose Rorschach Protocols Have F%+, Rigid Human Movement, or No Shock
(Subjects Tested Fall, 1946-1951)

Subjects	Number of Subjects: ESP Score in Relation to Chance Expectation: Above	At	Below	Number of Runs	Deviation from Chance Expectation	Mean Hits per Run
A. Sheep						
F%+, Mr or no sh; no other sign						
Only F%+	24	7	36	564	−44	4.92
Only Mr	14	2	11	235	0	5.00
Only no shock	6	1	6	113	−2	4.98
Any pair of F%+, Mr, no shock	7	0	6	110	+29	5.26
F%+, Mr or no shock and also R + or Total Mvt++ (in any combination) but without CF+ or C+	27	2	22	447	+37	5.08
F%+, Mr or no shock and also CF+ or C+ (in any combination)	10	3	10	201	+23	5.11
All sheep with F%+	52	9	61	1040	−2	5.00
All sheep with Mr	37	4	30	621	+83	5.13
All sheep with no shock	12	2	12	217	+3	5.01
All sheep with F%+, Mr or no shock	88	15	91	1670	+43	5.03
B. Goats						
F%+, Mr or no sh; no other sign						
Only F%+	12	1	11	204	−6	4.97
Only Mr	11	0	6	148	+38	5.26
Only no shock	3	1	3	59	+2	5.03
Any pair of F%+, Mr or no shock	4	0	2	52	+8	5.15
F%+, Mr or no shock and also R+ or Total Mvt++ (in any combination) but without CF+ or C+	8	0	13	184	−25	4.86
F%+, Mr or no shock and also CF+ or C+ (in any combination)	3	1	5	77	−20	4.74
All goats with F%+	20	2	22	380	−21	4.94
All goats with Mr	21	0	16	319	+27	5.08
All goats with no shock	6	1	5	103	+7	5.07
All goats with F%+, Mr or no shock	41	3	40	724	−3	5.00

a positive correlation between ESP scores and happy-go-lucky, carefree disposition; the Nashes (1958) (to mention only the most closely relevant factors) found a positive correlation between ESP scores and general activity, sociability and friendliness; and they found a negative correlation between ESP scores and restraint. For normal perception also, strong tendencies to inhibition are likely to result in poor performance. Klein (1954), in a quotation which B. Nicol was kind enough to send me, refers to his subjects who gave slow or inaccurate perceptual responses as those who on the basis of personality tests have constricted-control, as opposed to subjects with more effective perceptual responses, who seem on the basis of personality tests to have flexible control. He analyzes the unconscious motivation of the constricted-control subjects thus:

> . . . in the detailed efforts to reduce overlap, to segregate and increase the definitiveness of objects in the field, there are involved firm standards of what is 'really out there,' involving detailed search and emphasis upon objectively verifiable anchors and cues. It does not seem far-fetched to conceive that an orientation to reality, distrust and suppression of hunch and affect as bases of judgement would be the rule . . . we may expect that members of this group would hold the reins tightly on communication and release of affect, that feelings even if admitted into consciousness would not be easily communicated.

It would be interesting to administer perceptual tasks like those of Klein along with ESP tests, to find if the subjects who showed flexible control in normal perception had more marked ESP success than the subjects with constricted control in normal perception. Such a relation seems probable.

Before concluding the discussion of subjects with F%+, Mr and no shock, I report regretfully an analysis of the data which did not come out as I anticipated. It had occurred to me that, for subjects who showed no faint trace of color or shading shock, this denial of responsiveness to Rorschach stimuli might correspond to a denial of responsiveness (in either hits or psi-missing) to ESP stimuli. My guess was therefore that the no shock subjects would have a low standard deviation for ESP hits, with few scores either very high or very low. For the 26 sheep with no shock, who made 224 ESP runs, the standard deviation of the distribution was 1.88. This is lower, though not much lower, than the theoretical value of 2 or slightly more (some of the target decks were closed (Schmeidler, 1959)); and I was encouraged to make the same analysis for the goats with no shock. Here there were 12 subjects with 103 runs, and their standard deviation was 2.16.

This is such a strong hint that I was on the wrong track that it did not seem worth while to continue with similar examinations for other signs. We could, of course, devise some tortuous explanation to reconcile the two sets of scores (that the "no shock" lack of responsiveness represents a lid clamped down tight on impulses so strong that a person fears free expression of them; that for the goats, who say in effect, "ESP can't work; I have nothing to fear here" the lid comes off and the strong potential responsiveness is released). But the data are not strong enough to support such a topheavy theory, for the critical ratio of the difference between the standard deviations for sheep and goats is only 1.60.

Analysis of Subjects with CF+ and C+

The last of the signs for which there was empirical evidence in the preliminary series was CF+, which indicates a spontaneous, impulsive, emotionally toned response to what is happening at the moment. An additional sign, C+, for which there was no empirical evidence, was included because of its similarity: C+ responses "suggest extreme lack of emotional control" (Munroe, 1945).

A summary of records in the test group which show CF+ or C+ is given in Table 23. It will be noted that the pattern is roughly parallel to that of the preliminary group: again (though with many exceptions) sheep with CF+ tended to have low ESP scores but sheep with C+ did not, and goats with CF+ tended to have ESP scores above the chance level but goats with C+ did not. We will follow the practice of the preceding sections, and without regard for statistical significance, examine the data (perhaps with more attention than they deserve) to try to formulate some theory as to why their faint tendencies appear.

Table 23 (continued on p. 53)

ESP Scores of Sheep and Goats Whose Rorschach Protocols Have CF+ or C+ (Subjects Tested Fall, 1946-1951)

Subjects	Number of Subjects: ESP Scores in Relation to Chance Expectation: Above	At	Below	Number of Runs	Deviation from Chance Expectation	Mean Hits per Run
A. Sheep						
CF+ or C+; no other sign						
Only CF+	9	3	17	245	−22	4.91
Only C+	2	0	0	17	+8	5.47
Both CF+ and C+	1	0	2	26	−3	4.88
CF+ or C+ and also R+ or Total Mvt++ (in any combination)	13	1	6	171	+65	5.38
CF+ or C+ and also R+ or Total Mvt++ (in any combination) but without F%+, Mr or no shock	7	1	2	85	+42	5.49
CF+ or C+ and also F%+, Mr or no shock (in any combination)	10	3	10	201	+23	5.11
CF+ or C+ and also F%+, Mr or no shock (in any combination) but without R+ or Total Mvt++	4	3	6	115	0	5.00
All sheep with CF+	24	7	29	514	+30	5.06
All sheep with C+	8	0	4	103	+22	5.21
All sheep with CF+ or C+	29	7	31	574	+48	5.08

Table 23 (continued)

B. Goats

CF+ or C+; no other sign						
Only CF+	9	1	2	101	+25	5.25
Only C+	0	0	1	9	−3	4.67
Both CF+ and C+	0	0	1	9	−1	4.89
CF+ or C+ and also R+ or Total Mvt++ (in any combination)	7	1	1	79	+36	5.46
CF+ or C+ and also R+ or Total Mvt++ (in any combination) but without F%+, Mr or no shock	6	0	1	61	+31	5.51
CF+ or C+ and also F%+, Mr or no shock (in any combination)	3	1	5	77	−20	4.74
CF+ or C+ and also F%+, Mr or no shock (in any combination) but without R+ or Total Mvt++	2	0	5	59	−25	4.58
All goats with CF+	17	2	6	216	+48	5.22
All goats with C+	1	0	5	50	−17	4.66
All goats with CF+ or C+	18	2	10	257	+32	5.12

We begin with a problem. It is the consensus that spontaneity is desirable for ESP success. Why then should CF+, an indicator of spontaneity, be among the signs that counterindicate the typical sheep-goat pattern?

One of two explanations that seem plausible refers to the specific situation of classroom testing for ESP, which imposes severe restraints on spontaneity. If a rather lively person is required to conform to a dull, routinized procedure, morale might suffer and spontaneity be inhibited or even (because of resentment at boredom) be unconsciously used against, rather than according to the instructions of the experimenter. It would be interesting to test this possibility by comparing a group of markedly spontaneous subjects with a control group. ESP tests would be conducted both in a lively, individualized way and in a formal, routinized one. According to this interpretation, the advantage of the lively condition should be much greater for the sheep who gave evidence of marked spontaneity than for the control group who did not give such evidence. (I do not dare predict its effect upon the goats.)

Informal evidence to support this interpretation comes from the observations of several experimenters, including myself, that subjects who report striking spontaneous experiences (i.e., real life experiences which seem paranormal) are likely to have lower ESP scores than other subjects when tested under repetitive, formal conditions. Another scrap of evidence comes from the only group test I have conducted on subjects who were not at one of the colleges of the College of the City of New York. These were students at Duke University, and in general seemed gayer and more relaxed than the New York groups. They might therefore be expected to be less tolerant of the formalized

procedure. In conducting the experiment I found myself self-conscious and tense, talking to the large number of total strangers; it was an effort for me to make my voice carry in the room; my Northern accent probably jarred on their ears; and I had planned too long a procedure to fit comfortably into the time available, so that I hurried them along in a way that may have seemed like badgering them. All in all, in contrast to their own agreeable and friendly instructor, my guess is that I seemed like a cold pedantic person to them (but that my own classes do not have this impression of me).

The data of this group show a marked difference from the New York students, though the difference is not statistically significant when allowance is made for selection. Not only did sheep have low ESP scores, but well adjusted sheep and sheep without signs had low ESP scores. Only six of the 24 sheep had a positive ESP deviation, and four of those six had the check list entry of CF−. All but one of these subjects with positive deviations showed in their Rorschachs evidence of coldness or reserve. If there is any pattern in the results, it is that ESP scores were better (or at least, less bad) for subjects with the personality traits that they probably attributed to their Northern experimenter. The data are far too weak for any firm conclusion, but the feeble suggestion they make (that the personality traits associated with high ESP scores in one social atmosphere are associated with low scores in a very different atmosphere) should not be completely ignored.

A second explanation of why CF+ belongs among the unfavorable signs is that subjects whose most marked characteristic is lively impulsiveness may not be much concerned with the details of the sheep-goat distinction. If a task seems like fun to them they will throw themselves into it; if it bores them they will not. Thus the intellectualistic question of acceptance or rejection of the ESP hypothesis in this situation may be almost irrelevant—unless there is other reason to believe that in addition to emotionally toned, outward directed responsiveness the subjects are concerned with this particular issue.

As for C+, which was added to the list of signs against the evidence, it looks as if I made a bad guess. Only three subjects showed this sign and no other; all three scored consistently with the sheep-goat hypothesis. Of the 15 subjects who had C+ and another sign, ten scored in accordance with the sheep-goat hypothesis. If anything, then, the scant data indicate that subjects with C+ are likely to score consistently with the sheep-goat prediction even though other signs are present. This argues against the explanation proposed in the preceding paragraph. An ad hoc hypothesis to account for the apparent differences in pattern between subjects with CF+ and C+ is that (as was suggested above) subjects with fairly strong tendencies to spontaneity will be unfavorably impressed by a routinized procedure like the one of these experiments, but that subjects with extremely strong spontaneity will show it even under routinized conditions.

Now let us turn to the subjects who have CF+ or C+ and one or more of the other signs. Those with counterbalancing signs of overcontrol tend to score consistently with the sheep-goat hypothesis; but this has already been discussed. Subjects who have no signs of overcontrol but CF+ or C+ and also R+ or Total Movement++, that is, who show marked *general* lively responsiveness, tended to have high ESP scores whether they were sheep or goats. The same point

has been made in another publication, describing some of these subjects and some earlier ones tested individually, who gave many movement and color responses (Schmeidler and McConnell, 1958). These "dilated ambiequal" subjects tended to have high ESP scores. The impression they gave was one of eagerness, interest and zest; the findings are therefore similar to those of Humphrey (1951) whose expansive subjects scored high where her compressives scored low. It also is consistent with the Nashes' statement (1958) in discussion of their data and those of Humphrey and Nicol (1955) that "Strong drive, energy and activity may be related to psi expression." (To this list of three motor factors I would, however, like to add some such term as "receptivity," to give emphasis also to perceptual factors.) These subjects did not contribute to the sheep-goat difference because their ESP scores tended to be high whatever their answer to the sheep-goat question.

One other analysis should be made: that of the ESP scores of subjects with CF+ or C+ who had 10 or 11 check list entries. If the gap that appeared in ESP scores between 10 and 11 was due to my putting more of the inhibited records in the 11 category, then we should not expect to find that gap among the records with CF+ or C+, since all of them show considerable liveliness. And in fact the gap does not appear: the five sheep with 10 check list entries have an ESP mean of 5.00; the four with 11 entries have a mean of 5.32. Among the goats with CF+ or C+, the two with 10 check list entries have a mean of 5.32; the five with 11 entries have a mean of 5.48. This then seems consistent with the earlier interpretation.

In summary, the seven signs indicate three types of personality imbalance: overcontrol, intellectual over-responsiveness and emotional over-responsiveness (Table 24). Both sheep and goats with overcontrol had mean ESP scores near the chance level. Sheep and goats who showed any sign of over-responsiveness had mean ESP scores slightly above chance. Sheep and goats who showed signs of both intellectual and emotional over-responsiveness had mean ESP scores markedly above chance. Subjects who had signs of both overcontrol and over-responsiveness tended to have sheep-goat differences comparable to those of subjects without signs.

Perhaps we can add up these separate statements and arrive at a shorter summary of them. In an impersonal classroom ESP test, subjects who were freely responsive to intellectual or emotional stimuli tended to score higher than subjects who were inhibited. For subjects without marked imbalance toward either responsiveness or inhibition, sheep tended to score higher than goats. Our conclusion therefore might be that certain marked personality traits affect behavior almost independently of reaction to the specifics of a non-compelling situation; but when these marked personality traits are not shown, the response to the situation (i. e., the sheep-goat question) is the important determinant of behavior.

Table 24

Interpretative Summary of ESP Scores and Signs
(Subjects Tested Fall, 1946-1951)

Signs	Number of Subjects: ESP Scores in Relation to Chance Expectation: Above	At	Below	Number of Runs	Deviation from Chance Expectation	Mean Hits per Run
A. Only one type of sign						
Overcontrol						
Sheep	51	10	59	1022	−17	4.98
Goats	30	2	22	463	+42	5.09
Intellectual over-responsiveness						
Sheep	43	9	40	799	+18	5.02
Goats	30	1	17	413	+100	5.24
Emotional over-responsiveness						
Sheep	12	3	19	288	−17	4.94
Goats	9	1	4	119	+21	5.18
B. More than one type						
Overcontrol and intellectual over-responsiveness						
Sheep	27	2	22	447	+37	5.08
Goats	8	0	13	184	−25	4.86
Overcontrol and emotional over-responsiveness						
Sheep	10	3	10	201	+23	5.11
Goats	3	1	5	77	−20	4.74
No overcontrol; both intellectual and emotional over-responsiveness						
Sheep	7	1	2	85	+42	5.49
Goats	6	0	1	61	+31	5.51

VIII. DISCUSSION AND CONCLUSIONS

Limitations of Research Plan

In broad outline, the purpose of this project has been to find relationships between personality traits and ESP scores. The method has been to organize the research into three stages: (1) exploring the data of a preliminary group to see what personality patterns seem to relate to ESP success; (2) stating as a formal hypothesis that any pattern so found will recur in a test group; then (3) gathering new data under conditions which seem similar to the preliminary ones, and with subjects who seem comparable to the preliminary ones, to test whether the postulated pattern will again be present. Two such formal hypotheses concerning Rorschach scores and ESP scores have been investigated; both were confirmed.

The question now before us is: Can this three-stage method achieve the purpose of the research? My answer is: "No. Not if we stop there." Finding a pattern and confirming its presence under certain roughly defined conditions gives a datum which is specific to those conditions; it does not describe the meaningful relation which is our concern. To achieve our purpose a fourth and a fifth stage are needed: (4) stating the general rule which should apply, mutatis mutandis, to other conditions than those which have been investigated (a rule which should ideally tie in with other facts and theories) and (5) investigating the general rule with a variety of procedures.

A few hesitant steps have been taken toward this fourth stage of refining the formal hypotheses so that they can approach a little nearer to a general rule. Two processes are involved. One is utilizing hints from what seem to be trends in the data, that is, modifying and elaborating the conclusions from the statistically significant data according to impressions from insignificant trends. This is dangerous; and though I put forth my speculations, I do not defend them with confidence. The other is re-examining the method to find its special limitations and uncontrolled conditions. This is also improper. Approved procedure would be running experiments under different conditions to find their effect, not sitting back and musing about what would have happened if. . . . But since time does not permit doing the research that should be done, let us consider as well as we can the extent to which limitations of the method have influenced the findings.

Limitations of ESP Tests

First, what are the probable consequences of my special methods for testing ESP? All the data were derived from classroom tests; and there have been some critics who argue that the group method is likely to be too impersonal to bring forth the whole-hearted involve-

ment conducive to ESP success. The number of such critics must be smaller, now that the massive results of Van Busschbach and Anderson and White have appeared; but I would tend to be in qualified agreement with their general thesis. There is probably a close analogy between group or individual ESP sessions on the one hand and teaching in a class or in private sessions on the other. The private sessions usually give faster results and are more economical in student hours; they permit the instructor to note more sensitively what conditions inhibit or facilitate success; they encourage rapport. But some learning takes place in classrooms; and under the best conditions rapport is high and social interactions among class members result in lively learning and sudden insights.

Let us pursue this line of thinking. How good was rapport in my experiments? It varied. My impression is that it was worst during the session described in the preceding chapter and in one other when certain distinguished visitors were present, and in both sessions the scoring level was low. But no objective measures of rapport were recorded; this is an uncontrolled variable. What of the possibility of paranormal interaction among class members? This is difficult to test. I made one effort to do so for a small portion of the data by listing the hits obtained on each target by the various students. Some targets were more popular than others; and it seemed to me that a scoring pattern appeared: the students whose hits came preponderantly on the popular targets were more lively, outgoing and friendly; the students whose hits came preponderantly on the unpopular targets were more seclusive. Because of the difficulty of controlling for stimulus and serial preferences, this approach was not pursued further. But it indicates that there might be class interaction in ESP just as there is in discussion of academic topics. If so, it is another uncontrolled variable in the experiments reported above.

Another factor in the ESP experiments is that no reward was ever offered other than the intrinsic interest of the task and perhaps my implied concern for high scores. This might have been the reason for the significant sheep-goat differences: in the absence of other stronger rewards, theoretical concern with the rationale of ESP would be relatively important, weak though it was for many. It would therefore not be surprising if no sheep-goat difference were found in experiments where strong rewards were offered, or if with stronger rewards, a sheep-goat difference were found only for subjects who had considerable concern with the sheep-goat problem.

A third factor in the ESP experiments is that they were repetitive and therefore soon began to seem dull, even for most of the students who started to respond with keen interest. Probably this change of mood lowered both the scoring level and the reliability of the scores; and this is the obvious explanation of the tendency for higher scores with the 8-run procedure (where there was a maximum of 100 target items) than with the 9-run procedure (where there was a maximum of 225 target items). Thus a special limitation of generalizing about personality traits as they seem related to ESP scores in these experiments is that they appeared with a preponderantly dull procedure. We have no information about personality traits related to ESP performance when a more interesting method or a more meaningful target is used.

Limitation of Personality Tests

Next we must inquire about the probable consequences of my method of personality testing. It consisted of classing subjects as sheep or goats and scoring their Rorschach protocols. Its major limitations are clear. The Rorschach is primarily a test of basic personality structure, which should therefore relate to long-term behavioral trends. But our concern in the ESP tests was with short-term performance in a single task—a task, moreover, whose meaning to the subjects was never fully explored. Tieing Rorschach data to behavior is slippery; probably no one-to-one correspondence should be expected even with the best possible Rorschach interpretations. But here one virtue of the method probably compensates for some of its defects. In most cases the Rorschach was administered by the same person and under the same classroom conditions as the ESP test. Insofar as the Rorschach responses vary according to the special situation of administration, there should therefore be more correspondence between the two groups of data than if the two tests had been administered very differently.

But this correspondence must admit of many exceptions. In the first place, most subjects took the Rorschach with a special interest because they expected it to tell secrets about themselves; many did not have this attitude toward the ESP tests. In the second place, the general personality tendencies which (I believe) were validly shown in the Rorschach protocols sometimes applied and sometimes did not apply to the specific task of ESP responses; and the procedure did not effectively show whether they applied or not. Let me give an example to make this point, of two young men, both of whom called themselves sheep, and both of whom had fair social adjustment and the sign of F%+.

The first showed in his protocol a considerable amount of hostility and warm emotional reactions. The content indicated that the hostility was directed especially at his family and was due to resentment of what he considered their rigid, stupid, old-fashioned standards; he needed the restraint implied by F%+ to keep from losing his temper too often at home. But he seemed to enjoy using his mind, to welcome friendship, and to be ready to relax those controls of his when he found himself in congenial surroundings. His check list total was 10. The question for our research is: did he identify me with the demanding mother that he resented, and identify the rigid ESP procedure with the rigid demands of his home, or did he associate the psychology class and the intellectually oriented research with the broader life that he found congenial, where his potential strong self-control was relaxed and his potential responsiveness and warmth were released? A really good research design would let us answer this question with confidence; without an answer we do not know which of two opposite predictions to make from the Rorschach. And in our design, the only other datum is that he called himself a sheep.

Calling oneself a sheep does not necessarily mean that there is good rapport. The young man of our second example showed in his Rorschach consistent critical, negativistic and perfectionistic standards, which applied both to himself and to others. His protocol indicated strong self-control and keen awareness of facts, resulting in con-

forming, realistic behavior; he would be likely to follow rules but disregard subtle or unanticipated possibilities. His check list total was 11. He also called himself a sheep, but reported that he did so only because it was wrong to say that anything was impossible; his opinion was that paranormal success in any situation was most unlikely. Here the implications of the Rorschach for ESP performance are unambiguous, but the sheep-goat inquiry puts a negativistic subject into the category which was designed to represent cooperative and friendly ones.

Thus the general limitations of the personality measures are that the sheep-goat categories were too rigid, and the Rorschach responses too general, to point with accuracy to the responses expected in the ESP test. A subject who usually (as shown by the Rorschach) is spontaneous and cooperative might feel so threatened by the mysterious supernatural implications of clairvoyance that he acts inhibited and withdrawn in our ESP tests; someone who is usually unsure of himself and therefore hesitant and inhibited may feel relaxed and free in an ESP test for which he will receive no grades, and where he thinks success for anyone is so unlikely that he has as good a chance as the next man. Our measures may be valid enough to show in group averages some correspondence with performance; but they need a supplementary measure of how each individual felt when he was making each of his responses before they can be expected to give correspondences in detail.

Two special limitations of the personality measures, which relate to the methods of testing the two formal hypotheses, need repetition here. The check list, our only measure of social adjustment, is designed to give a hasty but balanced survey of personality. Because it is hasty—and general—it will not give precise correspondence with behavior. Further, because it is flexible, it permits of some modifications in terms of the examiner's preconceptions. There are indications that I modified it (in what according to the original design had been a critical area) by assigning only 10 checks to more of the livelier records and 11 checks to more of the constrained records. This was compensated for, in later formal tests, by putting all records with 8-12 checks in the same broad adjustmental category.

The second special limitation is that the second formal hypothesis is concerned with fragmentary items derived from the Rorschach scoring. Interpretation of these fragments can only be tentative, because a full interpretation will take into account the context of the entire protocol.

Conclusions

In spite of all these test limitations and methodological shortcomings, the over-all effect of many group sessions has been to demonstrate ESP patterns consistent with three previously stated hypotheses. These were confirmed at a level of significance lower than $P = .001$; and our first substantive conclusions are a restatement of them. In these, as in the following tentative conclusions, I have tried to distinguish between findings which are statistically significant, findings which have considerable statistical support, and other generalizations. The first two are stated in terms of the specifics of the procedure; and my hope for them is that other investigators, replicating the method, will obtain similar results. The latter are stated in terms of underlying personality

dynamics; and my hope for them is that other investigators, even if they use methods markedly different from mine, will find that their inferences converge on the same generalizations.

1. Subjects who accept the possibility of paranormal success under the conditions of the experiment (sheep) will tend to have higher ESP scores than subjects who reject this possibility (goats). The assumptions behind this thesis are (a) that in a college group or any other group with intellectual orientation, more interest and cooperation and higher morale will be displayed in tasks which seem reasonable to the subject, while negativism and withdrawal are more likely to be displayed in tasks which seem unreasonable; and (b) ESP scores tend to be higher with favorable motivation than with negativism and withdrawal. We would not expect this trend to be markedly present if other motives are aroused which are stronger than those of acceptance or rejection of the experimenter's implied hopes.

2. Sheep with few entries on the Munroe check list tend to have higher ESP scores than sheep with many entries, and the converse is true of goats; in other words, sheep superiority to goats in ESP scoring is more likely to occur in well adjusted than in poorly adjusted subjects. The assumptions here are that (a) the total number of entries on the Munroe check list is negatively related to social adjustment; that (b) subjects with good social adjustment are likely to set aside their personal preoccupations and concern themselves with a task as the experimenter sets it to them, but (c) subjects with poor social adjustment are likely to interpret instructions and task situations in an idiosyncratic way and—considered as a group, rather than as individuals—respond unpredictably.

There is some conflict between this hypothesis and the first: goats with excellent social adjustment might be more likely to set aside, temporarily, their intellectual rejection of the assigned task, and make a strong effort to succeed in what their experimenter tells them to do. But since this second hypothesis also was confirmed, we can take it that there was a gross tendency in the postulated direction. The results are similar to the patterns described by Humphrey (1951) of higher ESP scores for midrange than for extreme subjects, and of higher ESP scores for subjects who have better scores on the Bell Adjustment Inventory. It also is similar to the positive relation found by Nicol and Humphrey (1955) between ESP scores and freedom from nervousness and between ESP scores and emotional stability, and to the finding of the Nashes (1958) that subjects who rate higher on emotional stability have higher ESP scores.

3. Sheep whose Rorschach protocols show any of certain specified signs tend to have lower ESP scores than other sheep; and goats who show any of these signs tend to have higher ESP scores than other goats. My assumption in stating this hypothesis was only that a pattern found in a large preliminary group would be found again in a test group, if the test situation was similar to the preliminary one. Tentative interpretations of the specific signs have been proposed, and will be summarized and discussed below.

The next two items are post hoc findings, not anticipated in the original research plan.

4. When subjects are tested by a female experimenter in a setting where most authority figures are male, female subjects who give other evidence of accepting the experimenter or the research task are likely

to have higher ESP scores than males—but females who give other evidence of rejecting the experimenter or the task are likely to have lower ESP scores than males.

This is primarily a generalization of the fact that among my subjects, female sheep had higher ESP scores than male sheep, and female goats had lower ESP scores than male goats. A still more general statement of the same proposition might well be made, if sex is taken as only a special case of the factors which make for acceptance or rejection of experimenters and of the tasks which those experimenters assign. This more general statement would then apply to the sheep-goat hypothesis as well as this one.

5. Subjects who are given shorter and more interesting ESP tasks tend to have higher ESP scores than subjects who are given longer, duller ESP tasks.

This is a generalization of the fact that both sheep and goats tended to have higher ESP scores when they were asked to respond to both color and symbol on 100 targets than when they were asked to respond to the symbol on 225 targets. It is consistent with many other findings which other investigators have demonstrated at a very high level of statistical significance.

The next items are suggested as interpretations of patterns which seemed to be indicated by the signs reported in the third conclusion. Except for the one immediately following they are post hoc, and no attempt has been made to evaluate their statistical significance.

6. College students with a strong need for intellectual achievement, given ESP tests by their own instructor, and evidencing some acceptance of the ESP task (i.e., classed as sheep) will have higher ESP scores if their marks in the course have been high than if their marks in the course have been low.

The data of sheep in my classes give suggestive support to this proposition. I offer in interpretation the following possibilities: College students with a strong need for intellectual achievement whose grades are low in a certain course are already evincing negative response patterns in that course. These patterns will be present in the ESP task as they were in the other assigned tasks. Or: College students with a strong need for intellectual achievement who show by their high grades that they can achieve their goals, will (all other things being equal) accept as a goal an ESP task which their experimenter sets, and will succeed in it as they do in their other assignments. Or: Students with a strong need for intellectual achievement who have received high grades have (other things being equal) kindlier feelings toward their instructor, and thus to any task she assigns, than those with a strong need for intellectual achievement who have received low grades. These feelings will manifest themselves in higher ESP scores, as they would in other measures of performance which are influenced by mood.

(These possibilities are, of course, not mutually exclusive. They could be checked by a similar project in which ESP scores were compared with grades in the class where the tests were administered and also with grades in other classes, and also by comparing with grades the scores on ESP tests administered early in the semester before grades had been received, and ESP tests administered later. In the latter case, different groups of subjects would have to be used, since there are many reasons for expecting students' scores on a second test

of ESP to be different from their scores on an initial test. If such research is performed, need for achievement should be examined by one of the available well validated measures of it, making a distinction between a general need of this sort and a need which seems to apply specifically to desire for high grades in college courses.)

7. Subjects who show habitual restraint in certain situations tend to have poor ESP scores when tested in similar situations. Sheep will tend to score at or below the chance level; goats will tend to score at or above it.

This is a generalization of the near chance ESP totals of subjects with F%+, Mr or no shock, of the low ESP mean of sheep with check list totals of 11, and of the high ESP mean of goats with check list totals of 11. It deliberately uses the ambiguous word "poor" (which can refer to scores near chance expectation or scores opposite to what we infer the subject would like) because some of the data which are being generalized point to one of these and some to the other. It is supported by the low clairvoyance and card scores of Humphrey's compressives (1951), of my coartated subjects (Schmeidler and McConnell, 1958), of the subjects of Nicol and Humphrey who are low on self-confidence and thinking extraversion (1955) and of the subjects of the Nashes who are high on restraint and low on general activity (1958). It includes the modifier about relevant situations because of Humphrey's finding of high GESP scores for compressives, and because of my observation that subjects who usually show restraint, but who did not seem to feel restrained during the ESP test, often had high ESP scores.

The interpretation is simple, and consistent with many psychological findings. It is that inhibitory tendencies are likely to prevent effective responses to subtle or novel stimuli, including ESP stimuli. Klein, who was quoted on this topic in the preceding chapter, is only one of many psychologists who have obtained similar results in perception. A corollary of the general thesis is that when temporary inhibitions are established, as by failure at an imposed task accompanied by feelings of general inadequacy, performance on other tasks becomes less flexible and effective; and this has been clearly established for learning and cognition as well as for perception.

The ambiguity about the pattern of ineffective performance cannot yet be resolved. Just as it is hard to predict whether stage fright will result in a person's "freezing" and being tongue-tied or in his stumbling over his own feet, so we cannot yet, I think, say with confidence whether ineffective ESP performance because of emotional factors will take the form of near-chance psi scores or of psi-missing.

8. Subjects who show lively responsiveness in most situations and who give no evidence of inhibition or withdrawal in the experimental situation, will tend to have high ESP scores.

This is a generalization of the high ESP scores of both sheep and goats who had in their Rorschachs both R+ or Total Movement++ and also CF+ or C+, that is, whose Rorschachs indicated both intellectual and emotional liveliness. It is consistent with the descriptions of many distinguished research workers, such as Rhine and Carington, with the positive correlation obtained by Nicol and Humphrey (1955) between happy-go-lucky disposition and ESP and with the positive correlation obtained by the Nashes (1958) between ESP scores and general activity. It overlaps with Rasch's (1955) description of eight

"publicly known" sensitives, that "the sensitive subjects proved to be predominantly lively, outgoing, reproductive, and emotionally unstable personalities." Many other findings consistent with these could readily be cited. Its theory does not need elaboration, since in effect it is the converse of the preceding suggestion. We may summarize it by saying that freedom from inhibitions in a subject who is directing his efforts toward effective performance at a certain task permits utilization of even faint (and apparently incongruous) stimuli which are relevant to that task. It is therefore conducive to breadth, flexibility and accuracy in learning and cognition, in perception and in ESP.

Finally comes a broad generalization, for which each of the preceding ones is a special case.

9. Any factor which makes a subject inhibited or withdrawn in his approach to an ESP task is associated with poor ESP scores; any factor which makes a subject freely responsive in his approach to an ESP task is associated with good ESP scores. Such factors (a) may be long lasting personality traits (as of extreme self-control or of relaxed receptivity) which are most clearly observed when attitude to the task and mood are controlled; or (b) may be attitudes (such as acceptance or rejection of the specific task) which are most clearly observed in subjects who make effective use of their abilities in other forms of behavior; or (c) may be momentary variations of mood.

SUMMARY

The work reported here is part of a larger project, in which the initial finding was that subjects who accepted the possibility of paranormal success under the conditions of the experiment (sheep) tended to have higher ESP scores than subjects who rejected this possibility (goats). Observation of the subjects indicated that personality factors were related to ESP success for both sheep and goats, and further work was aimed at investigating such factors. The chief technique was the administration of projective tests of personality. This monograph reports the findings from the 1062 subjects who took the Rorschach Test.

Procedure

ESP targets were concealed randomized lists of 25 items. For earlier subjects, the items on the lists represented the five conventional ESP symbols. Subjects were instructed to respond to three lists as a unit. The standard session consisted of nine lists. Short psychological tests or questionnaires were usually interpolated between successive units. For later subjects, the items on the lists were paired colors and symbols. Each pair consisted of one of five colors and one of the five ESP symbols. The order of colors and of symbols was independently randomized. For these later subjects a pair of lists (composed of 25 paired items and thus of fifty separate targets) constituted a unit. Psychological tests or questionnaires were interpolated between successive units. The standard session consisted of four pairs of lists.

All responses were written. All had at least two independent scorings against the target.

All subjects were tested in college classroom groups.

All subjects were required, before they were informed of their ESP scores, to respond to questions showing whether they were sheep or goats.

The standard procedure for administering the Rorschach was to show slides representing the Rorschach cards and require the subjects to write their responses to those slides. In most cases, the written record was supplemented by an interview during which inquiry was made about ambiguous responses. Except in the preliminary group of 58 subjects, all Rorschachs were scored by the examiner without knowledge of the ESP scores.

The Munroe Inspection Technique was used to give a measure of social adjustment. This technique consists of filling out a check list of 24 Rorschach items and entering checks if the subject's responses fall outside the normal limits for that item. The total number of checks is an inverse measure of social adjustment.

Hypotheses

Examination of the Rorschachs of the preliminary group indicated that sheep with fewer check list entries (better social adjustment) had higher ESP scores than other sheep, and that goats with fewer check list entries had lower ESP scores than other goats. A hypothesis was formally stated to this effect; and the subsequent 1004 subjects were the test group for this hypothesis. The dividing line between good and poor adjustment was originally set between 10 and 11 check list entries.

After the first 250 records of the test group for the adjustment hypothesis had been gathered, they were examined to find if any other Rorschach categories seemed to relate to ESP scores. Seven scoring categories were taken as "signs" that the postulated sheep-goat difference was unlikely to appear. The hypothesis was stated that sheep whose records were free of those signs would tend to make higher ESP scores than other sheep; and that goats whose records were free of those signs would tend to make lower ESP scores than other goats. The subsequent 754 subjects were the test group for this hypothesis.

Interim Analyses of the Data

An experienced Rorschach analyst who had previously been unfamiliar with the Inspection Technique and the scoring method on which it was based, independently scored the Rorschachs of 83 subjects. The correlation between her check list totals and the experimenter's was +.88. Her division of subjects into those with good or poor adjustment, and into those free of signs or with signs, gave ESP results in conformity with both hypotheses. This indicates that the scoring method is communicable.

Division of records into those with check list entries of 10 or fewer and of 11 or more supported the hypothesis at a high level of significance. However it was also found that subjects with 10 and with 11 check list entries, that is, those just below and just above the arbitrary pre-set boundary between good and poor adjustment, also showed a significant difference in ESP scores in the predicted direction. Consequent re-examination of the records indicated that (a) most of the records could legitimately, within the broad directives of the Inspection Technique, have received one more or one less check than they did; and (b) there was a tendency for the experimenter to assign 10 entries to those records which could legitimately have been scored at 10 or at 11 and which showed liveliness (even if there was marked hostility or guilt) and to assign 11 entries to those records which could legitimately have been scored at 10 or at 11 and which showed constraint and self-control. There was no reason to believe that similar judgments of liveliness and constraint markedly affected the assignment of check list entries at any level other than that of 10 and 11. To prevent contamination of social adjustment ratings with liveliness-constraint ratings, the formal examination of the hypothesis of social adjustment in relation to ESP scores put all subjects with 8-12 check list entries into the same broad adjustment category (average adjustment).

Examination of the check list entries of the test group for signs showed that females had a significantly lower number of entries than males. (It is suggested that this is due to the high proportion of prospective school teachers among the females of this group.) No such difference was found in the test group for the adjustment hypothesis. (It is suggsted that this is due to the high proportion of returned veterans among the male students who were tested earlier.) Because of the obtained difference in the test group for signs, it was considered doubtful whether scores of males and females could legitimately be pooled. Formal tests of both hypotheses therefore separated male and female records.

Suggestively higher ESP scores were found for subjects who responded to colors and symbols than for those who responded to symbols alone (though no difference was observed between scores on colors and scores on symbols). Separate analyses of variance for these groups showed similar patterns in respect to check list entries and signs. Because of this it was decided that their scores could be pooled in the formal analyses.

Formal Tests of the Hypotheses

Analysis of variance of the test groups showed:

(a) Significantly higher ESP scores for sheep than for goats ($P < .001$).

(b) Significant interaction between social adjustment scores and the sheep-goat classification ($P < .001$). Sheep with 2-7 and 8-12 check list entries had higher ESP scores than goats with 2-7 or 8-12 check list entries; there was a slight difference in the opposite direction for subjects with check list entries totalling 13 or more.

(c) Significant interaction between the presence or absence of the signs described above and sheep-goat classification, in the predicted direction ($P < .001$).

The formal hypotheses are therefore considered to be confirmed by the data.

Other Findings

Analysis of variance of the test group for the adjustment hypothesis showed significant interaction ($P < .005$) between sex and sheep-goat classification. Female sheep had higher ESP scores and female goats had lower ESP scores than male sheep and goats. This is interpreted as due to the special testing conditions. Most tests were administered by the experimenter, one of the few female instructors at the college. It is inferred that the girls in her classes would therefore feel a more personal interest in supporting or repudiating her unconventional ESP experiments than would the men, and that this tendency toward identification with the experimenter or rejection of her was reflected in the ESP scores.

On the basis of Waldron's and Anderson's work, it was hypothesized that subjects with a strong need for intellectual achivement, given ESP tests by their own instructor, would have higher ESP scores if their grades in her course had been high than if their grades had been low. This hypothesis was restricted to the sheep. Two of the seven Rorschach signs were taken as indicators of a strong need for intellectual achievement. For the 101 sheep whose grades had been retained and whose Rorschach protocols showed these signs, the data were suggestively supportive ($P = .05$) in that the 26 with high grades tended to have higher ESP scores than the 16 with low grades. The effect seemed more pronounced among the girls than among the men.

Two of the signs were interpreted as indicators of marked intellectual (inner) activity, three as indicators of marked restraint or inhibition, and two as indicators of marked impulsiveness or responsiveness to outer stimuli. Inspection of the ESP scores of the subjects with these signs led to the following tentative generalizations:

Subjects with both a sign of overcontrol and a sign of over-responsiveness, that is, subjects with a balanced pattern of signs, contributed to the sheep-goat difference.

Subjects with a sign of overcontrol but no counterbalancing sign of over-responsiveness tended to have ESP scores near chance expectation.

Subjects with a sign of over-responsiveness but no sign of overcontrol tended to have ESP scores slightly above chance, whether they were sheep or goats.

Subjects with signs both of marked inner (intellectual) activity and of (emotional) responsiveness to outer stimuli tended to have high ESP scores whether they were sheep or goats.

Conclusions

Three hypotheses stated before the data were gathered are supported at a level of significance better than $P = .001$. It is therefore concluded that, with ESP tests conducted under conditions similar to these, sheep will tend to have higher ESP scores than goats; this tendency will be more pronounced for subjects whose social adjustment is good than for those whose social adjustment is poor; and the tendency will be more pronounced for subjects who do not show signs of marked inhibition or of marked over-responsiveness.

The fact that the obtained differences between groups were small and that there was a great deal of overlap in scores is attributed to several causes. One, of course, is "chance" coincidence. Chief among the others is believed to be our lack of information as to whether the attitudes and response tendencies inferred from the psychological tests influenced the subjects' moods as they made each individual ESP response.

General conclusions stated hesitantly about the relation between psychological dynamics and ESP success are: Feelings of constraint, withdrawal or negativism are associated with near-chance ESP scores or with psi-missing; and feelings of free responsiveness are associated with successful ESP scoring. Among the various possible indicators of negativism, constraint or withdrawal in a particular situation are: (a) calling oneself a goat, (b) obtaining low college grades in a similar situation if grades are valued highly, (c) showing general personality tendencies toward overcontrol if there are no counterbalancing tendencies toward over-responsiveness. Among the various possible indicators of free responsiveness in a particular situation are: (a) calling oneself a sheep, (b) showing general tendencies toward good social adjustment in a similar situation, (c) obtaining good college grades in a similar situation if grades are valued highly, (d) showing general personality tendencies toward free responsiveness.

RÉSUMÉ

Arriere-Plan

Le travail dont il est rendu compte ici n'est qu'une partie d'une plus vaste entreprise dont la conclusion première fut que les sujets qui admettaient la possibilité d'un succès paranormal dans les conditions de l'expérience (Brebis) tendaient à donner des résultats en ESP supérieurs à ceux des sujets qui rejetaient cette possibilité (Boucs). L'observation des sujets indiquait qu'un rapport existait entre les facteurs de personnalité et le succès en ESP aussi bien pour les boucs que pour les Brebis, et les travaux furent poursuivis en vue de l'examen de ces facteurs. La technique principale consista à appliquer les tests de personnalité projectifs. La présente monographie relate les conclusions fournies par les 1062 sujets qui se soumirent au Test de Rorschach.

Procedure

Les buts d'ESP furent des listes de 25 stimuli cachés et pris au hasard (randomized). Pour les sujets du début, les stimuli des listes représentaient les cinq symboles ESP conventionnels. Les instructions données aux sujets fixaient comme unité la réponse à 3 listes. La "session" type comportait 9 listes. De courts tests psychologiques ou questionnaires étaient généralment introduits entre unités successives. Pour les sujets suivants, les stimuli des listes furent des couleurs et des symboles associés. Chaque paire consistait en l'une de cinq couleurs et un des cinq symboles ESP. L'ordre des couleurs et des symboles avait été indépendamment déterminé par le hasard (randomized). Les instructions données aux sujets fixaient comme unité la réponse à une liste (de 25 paires de stimuli, soit 50 buts séparés). Des tests psychologiques et des questionnaires étaient introduits entre unités successives. La "session" type comportait 4 listes.

Toutes les réponses étaient faites par écrit. Pour toutes, l'écart du but faisait l'objet d'au moins deux pointages indépendants. Pour tous les sujets, les tests portaient sur des groupes de classes de collège.

Tous les sujets avaient à répondre, avant d'être informés de leurs résultats d'ESP, aux questions montrant s'ils étaient Brebis ou Boucs.

La procédure type pour l'application du Rorschach consistait à faire passer des clichés représentant les cartes Rorschach et à demander aux sujets d'écrire leur réponse à ces clichés. Dans la plupart des cas, le compte rendu écrit était complété d'une interview pendant laquelle on interrogeait sur les réponses ambigues. Sauf pour le groupe préliminaire de 58 sujets, les notes des Rorschach ont été données par l'examinatrice sans qu'elle connaisse les résultats d'ESP.

La technique d'inspection Munroe a été employée pour mesurer l'adaptation sociale. Cette technique consiste à remplir une liste ques-

tionnaire de 24 articles Rorschach en attribuant à chacun 1, 2 ou 3 points si les réponses du sujet sortent des limites normales pour cet article. Le total des points est en rapport inverse de l'adaptation sociale.

Hypotheses

L'examen des Rorschachs du groupe préliminaire indiquait que les Brebis totalisant le moins de points dans la liste questionnaire (meilleure adaptation sociale) donnaient des résultats ESP supérieurs à ceux des autres Brebis et que les Boucs totalisant le moins de points dans la liste questionnaire donnaient des résultats ESP plus bas que ceux des autres Boucs. Une hypothèse a été formulée à cet effet, et les 1004 sujets suivants ont formé le groupe de vérification de cette hypothèse. La ligne de démarcation entre bonne et mauvaise adaptation fut fixée originairement entre 10 et 11 points de la liste questionnaire.

Lorsque le 250 premiers enregistrements du groupe de vérification de l'hypothèse de l'adaptation eurent été recueillis, on les a examinés pour voir si quelque autre catégorie de Rorschach pouvait être en relation avec les résultats d'ESP. 7 catégories furent prises comme "signes" que la différence postulée "Brebis-Boucs" ne semblait pas devoir apparaître. L'hypothèse fut émise que les Brebis dont les réponses ne comportaient pas ces signes tendraient à donner des résultats d'ESP plus élevés que les autres Brebis, et que les Boucs dont les réponses ne comportaient pas ces signes tendraient à donner des résultats d'ESP plus bas que les autres boucs. Les 754 sujets suivants formèrent le groupe de vérification de cette hypothèse.

Analyse Provisoire des Faits

Une analyste ayant une grande expérience du Rorschach, qui ne s'était pas familiarisée jusqu'ici avec la Technique d'Inspection et les méthodes de chiffrage sur lesquelles elle est basée, a noté indépendamment les Rorschachs de 83 sujets. Le rapport entre les totaux de ses listes et ceux de l'expérimentatrice fut 4,88. Sa division des sujets en bien ou mal adaptés et en exempts ou atteints de signes donna des résultats d'ESP conformes à l'une et l'autre hypothèses. Cela indique que la méthode de chiffrage est communicable.

La division des résultats entre ceux comportant 10 réponses à la liste questionnaire ou moins et ceux comportant 11 ou plus confirme l'hypothèse à un haut degré de signification. Cependant, on a trouvé que des sujets ayant 10 et ayant 11 réponses à la liste, c'est à dire ceux qui sont juste en dessous ou juste au dessus de la ligne arbitrairement tracée entre bonne et mauvaise adaptation, montraient également une différence significative de leurs résultats d'ESP dans la direction prévue. Un nouvel examen des enregistrements indiqua que:

a) un point de plus ou de moins aurait pu être affecté légitimement à la plupart des totaux sans sortir des larges limites de la technique d'Inspection,

b) que l'expérimentatrice avait tendance à allouer 10 points à ceux des totaux qui auraient pu légitimement être évalués à 10 ou 11 et qui montraient une vive réactivité (même s'il y avait hostilité ou culpabilité marquée) et à allouer 11 points à ceux des totaux qui auraient pu légitimement être évalués à 10 ou à 11 et qui montraient une contrainte ou un contrôle de soi. Nous n'avons pas de raison de penser que des jugements semblables de vivacité et de contrainte ont influé

appréciablement sur l'attribution des réponses à la liste questionnaire à aucun échelon autre que 10 et 11.

Afin d'éviter toute contagion des estimations de l'adaptation sociale à celles de vivacité-contrainte, l'examen formel de l'hypothèse de l'adaptation sociale en liaison avec les résultats d'ESP a placé tous les sujets se situant dans la zône moyenne des réponses à la liste questionnaire (y compris ceux ayant des totaux de 10 et de 11) dans une même large catégorie d'adaptation.

L'examen des réponses à la liste questionnaire du groupe de vérification des signes a montré que les femmes donnaient un chiffre de réponse significativement plus bas que les hommes. (On suggère que ceci est dû à la forte proportion de futures enseignantes dans les femmes de ce groupe). On n'a pas trouvé de différence de ce genre dans le groupe pour la vérification de l'hypothèse de l'adaptibilité. (On suggère que ceci est dû à la forte proportion d'anciens combattants dans les étudiants examinés plus tôt). En raison de la différence obtenue dans le groupe de vérification des signes, on a considéré qu'il était douteux que des résultats hommes et femmes puissent légitimement s'additionner. Les tests formels des deux hypothèses ont donc séparé les résultats hommes des résultats femmes.

Des chiffres d'ESP plus élevés semblent avoir été trouvés avec des sujets réagissant aux couleurs et aux symboles qu'avec ceux qui ne réagissent qu'aux seuls symboles. Des analyses séparées des variations pour ces groupes ont montré des modèles semblables quant aux réponses à la liste questionnaire et aux signes. En raison de ceci, il fut décidé que leurs totaux pouvaient être additionnés dans les analyses formelles.

Tests Formels des Hypotheses

L'analyse des variations des groupes de tests a montré:

a) des résultats d'ESP significativement plus élevés pour les brebis que pour les Boucs ($P < 0{,}0001$)

b) une intéraction significative entre les chiffres d'adaptation sociale et la classification Brebis-Boucs ($P < 0{,}0001$). Les Brebis ayant de 2 à 7 et de 8 à 12 réponses à la liste questionnaire ont eu des chiffres d'ESP supérieurs à ceux des Boucs ayant de 2 à 7 et de 8 à 12 réponses à la liste questionnaire; il y eut une légère différence dans la direction opposée pour les sujets dont les réponses à la liste questionnaire totalisaient 13 ou plus.

c) une interaction significative entre la présence et l'absence de signes et la classification Brebis-Boucs, dans la direction prévue ($P < 0{,}001$).

Les hypothèses formelles sont par conséquent considérées comme confirmées par les faits.

Autres Constatations

L'analyse des variations pour le groupe de vérification de l'hypothèse de l'adaptation montre une interaction ($P < 0{,}005$) entre le sexe et la classification Brebis-Boucs. Les Brebis femmes ont atteint des chiffres d'ESP supérieurs et les Boucs Femmes des chiffres d'ESP inférieurs à ceux des Brebis hommes et Boucs Hommes. Ceci s'interprète comme étant dû aux conditions spéciales du test. La plupart des tests étaient administrés par l'expérimentatrice, une des rares instructeurs-femmes du collège. Nous supposons que les filles dans leurs

classes devaient donc éprouver un intérêt plus personnel à soutenir ou à rejeter ses expériences inconventionnelles d'ESP que ne le faisaient les hommes, et que cette tendance à la reconnaissance ou à la négation s'est reflétée dans les résultats d'ESP.

En partant des tragaux de Waldron et d'Anderson, on a émis l'hypothèse que les sujets ayant un vif désir de réussite intellectuelle, recevant des tests d'ESP des mains de leur propre instructeur, atteindraient des chiffres d'ESP plus hauts si leurs notes dans sa classe avaient été élevées plutôt que si leurs notes avaient été faibles.

Cette hypothèse était limitée aux Brebis. Deux des 7 signes de Rorschach ont été pris comme indicateurs d'un vif désir de réussite intellectuelle. Pour les 101 Brebis dont les notes ont été retenues et dont les protocoles de Rorschach ont montré ces signes, les faits suggèrent l'exactitude ($P = 0{,}004$) en ceci que les 26 ayant des notes élevées ont tendu à rendre des résultats d'ESP meilleurs que les 16 ayant des notes faibles. L'effet semble avoir été plus prononcé chez les filles que chez les hommes.

Deux des signes ont été interprêtés comme étant des indicateurs d'activité intellectuelle (intérieure) marquée, 3 comme étant des indicateurs de retenue ou d'inhibition marquée, et 2 comme étant des indicateurs d'impulsivité ou de réaction aux stimuli extérieurs marquée. L'étude des chiffres d'ESP des sujets présentant ces signes a conduit aux généralisations provisoires suivantes:

—les sujets présentant un signe d'excès de contrôle de soi avec un signe d'excès de responsivité, c'est à dire les sujets avec un agencement de signes équilibré, participaient à la différence Brebis-Boucs.

—les sujets présentant un signe d'excès de contrôle de soi mais aucun signe le contrebalançant d'excès de responsivité ont tendu à donner des chiffres d'ESP voisins de la probabilité.

—les sujets présentant un signe d' 'excès de responsivité mais aucun signe le contrebalançant d'excès de contrôle de soi ont tendu à donner des chiffres d'ESP un peu au dessous de la probabilité, qu'ils aient été Brebis ou Boucs.

—les sujets présentant à la fois des signes d'activité intérieure marquée et de responsivité aux stimuli extérieurs ont tendu à donner des chiffres d'ESP élevés, qu'ils aient été Brebis ou Boucs.

Conclusions

Trois hypothèses énoncées avant que les faits aient été recueillis sont confirmées à un dégré de signification supérieur à $P = 0{,}001$. Nous en concluons donc que, dans des tests d'ESP conduits dans des conditions semblables à celles-ci, les Brebis tendront à atteindre des chiffres d'ESP plus élevés que ceux des Boucs; cette tendance sera plus prononcée pour les sujets dont l'adaptation sociale est bonne que pour les sujets dont l'adaptation sociale est mauvaise, et la tendance sera plus prononcée pour les sujets qui ne présentent pas de signes d'inhibition marquée ou d'excès de responsivité marquée.

Le fait que les différences constatées entre les groupes ont été faibles, et qu'il s'est produit un bon nombre de chevauchements dans les résultats peut être attribué à plusieurs causes. La principale d'entre elles, croyons-nous, est notre manque de renseignements sur le point de savoir si les attitudes et les penchants à répondre qui ont été déduits des tests psychologiques ont influé sur l'humeur des sujets au moment où chacun faisait sa réponse personnelle d'ESP.

Les conclusions générales formulées, non sans hésitation, quant au rapport entre la dynamique psychologique et la réussite en ESP sont:

—les sentiments de contrainte et de repli s'associent aux résultats d'ESP voisins de la probabilité ou au "PSI missing"

et les sentiments de libre responsivité s'associent à la réussite en ESP.

Au nombre des divers indicateurs possibles de contrainte ou de repli dans une situation particulière, on trouve:

a) l'impression d'être un Bouc

b) l'obtention, dans une situation semblable, de notes de classe basses, si l'on attribue une grande valeur aux notes.

c) une personnalité dont la tendance générale est l'excès de contrôle de soi alors qu'il n'y a pas de tendance équilibrante à l'excès de responsivité.

Au nombre des divers indicateurs possibles de libre responsivité dans une situation particulière, on trouve:

a) l'impression d'être une Brebis

b) une tendance générale à la bonne adaptation sociale dans une situation semblable.

c) l'obtention, dans une situation semblable, de bonnes notes de classe, si l'on attribue une grande valeur aux notes.

d) une personnalité dont la tendance générale est la libre responsivité.

ZUSAMMENFASSUNG

Hintergründe

Die Arbeit, über die hier berichtet wird, bildet ein Teilstück einer umfassenderen Untersuchung, an deren Anfang sich ergeben hatte, dass Vp, welche die Möglichkeit erfolgreicher parapsychologischer Experimente unter den angewandten Bedingungen bejahten ("Schafe"), eine höhere Trefferzahl bei ASW aufzuweisen pflegen, als Vp, welche diese verneinen ("Böcke"). Eine Beobachtung solcher Vp ergab, dass sowohl bei Schafen als auch bei Böcken Persönlichkeitsfaktoren am Erfolg der ASW beteiligt waren. Diese sollten nun näher untersucht werden. Es wurden hierbei hauptsächlich Projektionstests angewendet. Die vorliegende Monographie berichtet über die Ergebnisse bei 1062 Vp, die dem Rorschachtest unterworfen wurden.

Vorgangsweise

Das Ziel bestand aus versteckten, nach dem Zufall zusammengestellten Listen mit 25 Einzelheiten. Bei den ersten Vp stellten die Einzelheiten auf den Listen die 5 üblichen ASW-Symbole dar. Die Vp mussten sich auf drei Listen als Einheit einstellen. Eine Normalsitzung umfasste neun Listen. Kurze psychologische Tests oder Befragungen wurden gewöhnlich zwischen die aufeinanderfolgenden Einheiten eingeschoben. Bei den späteren Vp bestanden die Einzelheiten auf den Listen aus in Paaren zusammengefassten Farben und Symbolen. Jedes Paar bestand aus einer von fünf Farben und einem der fünf ASW-Symbole. Die Reihenfolgen der Farben und Symbole wurde unabhängig von einander aufs Geratewohl zusammengestellt. Die Vp wurden angewiesen, sich auf je eine Liste als Einheit (25 paarweise Einzelheiten, somit fünfzig einzelne Punkte) einzustellen. Zwischen die aufeinanderfolgenden Einheiten wurden psychologische Tests oder Befragungen eingeschoben. Die Normalsitzung bestand aus vier Listen.

Alle Antworten erfolgten schriftlich. Alle zeigten mindestens zwei von einander unabhängige Angaben, die das Ziel verfehlten.

Alle Vp wurden in Gruppen von College-Klassen getestet.

Vor der Mitteilung ihrer ASW-Trefferzahlen wurde allen Vp ein Fragebogen vorgelegt, um zu ermitteln, ob es sich um Schafe oder Böcke handelte.

Die übliche Vorgangsweise bei der Anwendung des Rorschachtests bestand aus dem Ablaufenlassen von projizierten Rorschach-Klecksbildern, und der Aufforderung an die Vp, ihre Reaktion auf dieselben schriftlich niederzulegen. In den meisten Fällen wurde die Niederschrift ergänzt durch eine Befragung, in deren Verlauf mehrdeutige Reaktionen aufgeklärt wurden. Ausser bei der ersten Gruppe von 58

Vp wurden alle Rorschachtests ohne Kenntnis der ASW-Treffer durch den Prüfer ausgewertet.

Das "Munroe-Inspektions-Verfahren" wurde angewendet, um die soziale Anpassungsfähigkeit zu bewerten. Bei diesem Verfahren muss eine Kontrolliste von 24 Einzelheiten aus dem Rorschachtest ausgefüllt werden, wobei für jeden Punkt entweder ein, zwei oder drei Kontrollziffern vermerkt werden, je nachdem, inwieweit die Antworten der Vp aus dem normalen Rahmen für den betreffenden Punkt herausfallen. Die Gesamtsumme der Kontrollziffern steht in umgekehrtem Verhältnis zu der sozialen Anpassungsfähigkeit.

Hypothesen

Die Auswertung der Rorschachtests der ersten, vorläufigen Gruppe wies darauf hin, dass Schafe mit geringerer Zahl von Kontrollpunkten (besserer sozialer Anpassungsfähigkeit) höhere ASW-Trefferzahlen hatten, als die anderen Schafe, und das Böcke mit weniger Kontrollpunkten niedrigere ASW-Trefferzahlen zeigten, als andere Böcke. Dies wurde in einer Hypothese in Form einer Regel niedergelegt; die Untersuchung der folgenden 1004 Vp sollte diese Hypothese überprüfen. Die Trennungslinie zwischen guter und schlechter Anpassung wurde ursprünglich bei 10 und 11 Kontrollpunkten angesetzt.

Nachdem die ersten 250 Untersuchungsprotokolle der Gruppe zur Überprüfung der Hypothese hinsichtlich der Anpassungsfähigkeit eingesammelt worden waren, wurden sie daraufhin durchgesehen, ob etwa irgendwelche andere Kategorien des Rorschachtests eine Beziehung zu den ASW-Trefferzahlen besitzen könnten. Sieben Kategorien der Auswertung wurden als "Anzeichen" dafür aufgefasst, dass ein Auftreten der vorweggenommenen Schaf-Bock-Unterscheidung unwahrscheinlich sei. Es wurde die Hypothese aufgestellt, dass Schafe, deren Auswertung diese Anzeichen nicht enthielten, eine höhere ASW-Trefferzahl aufweisen würden, als andere Schafe und Böcke, die dieser Anzeichen in ihren Auswertungen entbehrten, niedrigere ASW-Trefferzahlen aufweisen würden, als andere Böcke. Bei den folgenden 754 Vp wurde diese Hypothese überprüft.

Vorläufige Analyse der Ergebnisse

Eine erfahrene Auswerterin der Rorschachtests, die bisher mit dem Inspektions-Verfahren und der Methode der Kontrollpunkte, auf dem es fusst, nicht vertraut war, hat unabhängig davon die Rorschach-Tests von 83 Vp ausgewertet. Die Vergleichsziffer zwischen ihrer Gesamtauswertung und der der Versuchsleitung betrug + .88. Ihre Einteilung der Vp in solche mit guter und schlechter Anpassung und in solche, bei denen die Anzeichen sich fanden oder bei denen sie fehlten, ergaben ASW-Treffer, die mit beiden Hypothesen übereinstimmten. Dies zeigt, dass die Auswertungsmethode anwendbar ist.

Die Einteilung der Auswertungen in solche, bei denen die Kontrollpunkte 10 oder weniger, beziehungsweise 11 oder mehr betragen, bestätigte die Hypothese in hohem Grade. Es ergab sich jedoch ebenfalls, dass Vp mit 10 und 11 Eintragungen in der Kontrolliste, d.h. die gerade unter und über der willkürlich gezogenen Grenzlinie zwischen guter und schlechter Anpassungsfähigkeit lagen, ebenfalls einen bedeutungsvollen Unterschied in Richtung der vorweggenommenen ASW-Treffsicherheit aufwiesen. Eine folgende Nachuntersuchung der Auswertungen zeigte, dass a) die meisten Auswertungen innerhalb der

weiten Grenzen des Inspektions-Verfahrens einen Punkt mehr oder weniger hätten erhalten können, als tatsächlich der Fall war, und (b) dass die Versuchsleiterin dazu neigte, 10 Punkte bei den Auswertungen einzusetzen, bei denen 10 oder 11 Punkte angemessen waren und bei denen eine lebhafte Reaktionsfähigkeit (auch bei ausgesprochener Feindseligkeit oder Schuldbewusstsein) vorlag,—und den Auswertungen 11 Punkte zu geben, bei denen lo oder 11 Punkte angemessen waren, die jedoch Zurückhaltung oder Selbstbeherrschung aufwiesen. Es lag jedoch kein Grund zu der Annahme vor, dass ausser bei 10 und 11 Kontrollpunkten, eine derartige Einschätzung der Lebhaftigkeit und Zurückhaltung die Bewertungslisten irgendwo sonst wesentlich beeinflusste. Um die Störung in der Bewertung der sozialen Anpassung durch die Einschätzung der Lebhaftigkeit und Zurückhaltung zu unterbinden, wurden bei der grundsätzlichen Untersuchung der Hypothese über das Verhältnis der sozialen Anpassungfähigkeit zu der Zahl der ASW-Treffer alle Vp an der mittleren Grenze der Kontrolliste, einschliesslich derer mit insgesamt 10 und 11 Punkten, in einer einzigen Kategorie für allgemeine Anpassungsfähigkeit untergebracht.

Eine Untersuchung der Eintragungen in den Kontrollisten hinsichtlich der besonderen Anzeichen, zeigte dass weibliche Vp in bedeutsamem (significant) Grade deren weniger aufwiesen, als männliche. (Man vermutet, dass dies auf die grosse Anzahl von künftigen Lehrerinnen innerhalb dieser Gruppe von Vp zurückzuführen ist.) Es wurden keine derartigen Unterschiede bei der Gruppe zur Überprüfung der Hypothese hinsichtlich der sozialen Anpassungsfähigkeit gefunden. (Man vermutet, dass dies durch die grosse Zahl von Kriegsheimkehrern bedingt ist, die sich unter den früher getesteten männlichen Studenten befanden.) Angesichts der Unterschiede in der Testgruppe für Anzeichen, erschien es zweifelhaft, ob ein Zusammenwerfen der Auswertungen für weibliche und männliche Vp zulässig sei. In den grundsätzlichen Überprüfungen beider Hypothesen wurden deshalb die männlichen und weiblichen Auswertungen getrennt.

Auffallend höhere ASW-Trefferzahlen wurden bei Vp angetroffen, die sowohl auf Farben als auf Symbole ansprachen, verglichen mit solchen, die nur auf Symbole reagierten. Besondere Analysen der Abarten bei diesen Gruppen zeigten gleiche Unterabteilungen im Vergleich mit den Kontrollisten und Anzeichen. Es wurde deshalb verfügt, dass ihre Auswertungen bei der grundsätzlichen Analyse vereinigt werden könnten.

Grundsätzliche Überprüfung der Hypothesen

Die Analysen der verschiedenartigen Testgruppen ergaben:

a) Charakteristisch höhere Trefferzahlen für Schafe als für Böcke ($P < .001$).

b) Charakteristische Beziehungen zwischen der sozialen Anpassungsfähigkeit und der Einteilung in Schafe und Böcke ($P < .001$). Schafe mit 2-7 und 8-12 Eintragungen in der Kontrolliste zeigten höhere ASW-Trefferzahlen, als Böcke mit 2-7 oder 8-12 Eintragungen. Es bestand ein geringfügiger Unterschied bei Vp mit insgesamt 13 oder mehr Kontrollpunkten.

c) Charakteristische Wechselwirkungen zwischen dem Vorhandensein oder Fehlen der Anzeichen und der Einteilung in Schafe und Böcke in der erwarteten Richtung ($P < .001$).

Es ergibt sich also eine Bestätigung der grundlegenden Hypothesen durch die einzelnen Tatsachen.

Weitere Ergebnisse

Eine Analyse der Verschiedenheiten innerhalb der Testgruppe für soziale Anpassungsfähigkeit zeigte bedeutsame (significant) Wechselwirkungen ($P < .005$) zwischen dem Geschlecht und der Einteilung in Schafe und Böcke. Es wird dies durch die Besonderheiten der Versuchsbedingungen erklärt. Fast alle Experimente wurden von der Versuchsleiterin durchgeführt, einer der wenigen weiblichen Dozenten des College. Es wird deshalb angenommen, dass die Mädchen in ihren Klassen ein grösseres persönliches Interesse daran hatten, ihre aussergewöhnlichen ASW-Versuche zu unterstützen oder abzulehnen, als dies bei männlichen Studenten der Fall sein würde, und dass diese Tendenz zur Identifizierung oder Ablehnung sich in den ASW-Treffern spiegelte.

Unter Zugrundelegung der Arbeit von Waldron und Anderson wurde vermutet, dass Vp mit einem grossen Bedürfnis nach intellektueller Bewährung bei ASW-Versuchen unter Leitung ihres eigenen Lehrers höhere ASW-Trefferzahlen aufweisen würden, wenn ihre Erfolge in der Klasse gross waren, als wenn sie sich als gering erwiesen. Diese Hypothese wurde auf die Schafe beschränkt. Zwei der sieben Rorschach-Anzeichen wurden als starke Hinweise auf ein Verlangen nach intellektueller Bewährung aufgefasst. Bei den 101 Schafen, deren Noten aufgehoben worden waren und deren Rorschachbewertung diese Anzeichen enthielt, waren die Tatsachen in hohem Masse überzeugend ($P = .04$), weil die 26 mit guten Noten dazu neigten, höhere ASW-Trefferzahlen aufzuweisen, als die 16 mit schlechten Noten. Diese Zusammenhänge schienen bei den weiblichen Vp stärker zutage zu treten, als bei den männlichen.

Zwei von den Anzeichen wurden als Hinweise auf eine ausgesprochene (innere) intellektuelle Aktivität aufgefasst, drei als Hinweise auf ausgesprochene Zurückhaltung oder Hemmung, und zwei als Hinweise auf ausgesprochene Impulsivität, oder Ansprechbarkeit auf äussere Anregungen. Eine Untersuchung der ASW-Treffer der Vp mit diesen Anzeichen ergaben folgende versuchsweisen Verallgemeinerungen:

Vp mit einem Anzeichen von zu starker Selbstbeherrschung und einem Anzeichen zu starker Ansprechbarkeit, also Vp mit ausgleichenden Anzeichen, trugen zu der Schaf-Bock-Unterscheidung bei.

Vp mit einem Anzeichen für zu starke Selbstbeherrschung ohne ausgleichende Anzeichen von zu starker Ansprechbarkeit, neigten zu ASW-Treffern in der Nähe der Zufallserwartung.

Vp mit Anzeichen von zu starker Ansprechbarkeit, ohne Zeichen zu starker Selbstbeherrschung, neigten zu ASW-Treffern etwas über der Zufallserwartung, gleichgültig, ob es Schafe oder Böcke waren.

Vp mit Zeichen sowohl von ausgesprochener innerer Aktivität und von Ansprechbarkeit auf äussere Reize, neigten zu hohen ASW-Trefferzahlen, gleichgültig, ob es Schafe oder Böcke waren.

Schlussfolgerungen

Drei der vor Zusammenstellung der Tatsachen aufgestellten Hypothesen werden mit einer Höhe der Bedeutsamkeit (significance), die über $P = .001$ liegt, unterstützt. Daraus wird gefolgert, dass bei ASW-Versuchen, die unter gleichartigen Bedingungen in Angriff

genommen werden, Schafe zu höheren Trefferzahlen neigen werden, als Böcke. Diese Tendenz wird stärker ausgeprägt sein bei Vp mit gutem sozialem Anpassungsvermögen, als bei solchen, bei denen dieses gering ist, ferner wird diese Tendenz stärker sein bei Vp, die keine Zeichen von starker Gehemmtheit oder zu starker Ansprechbarkeit zeigen.

Die Tatsache, dass die erhaltenen Unterschiede zwischen verschiedenen Gruppen nur gering waren, und die Auswertung der Trefferzahlen vielfache Überschneidungen aufwies, wird auf verschiedene Ursachen zurückgeführt. Eine der wichtigsten scheint in unserer Unkenntnis des Einflusses der Verhaltensweisen und Tendenzen der Ansprechbarkeit—wie sie aus den psychologischen Tests erschlossen werden—auf die Stimmung der Vp während der einzelnen Versuche zu liegen.

Allgemeine, wenn auch mit Vorsicht aufgestellte Schlussfolgerungen über die Beziehungen zwischen psychologischen Dynamismen und dem Erfolg bei ASW lauten: Gefühle der Einengung und des Zurückweichens sind mit ASW-Trefferzahlen verknüpft, die in der Nähe der Zufallserwartung oder des psi-Verfehlens liegen, während Gefühle freier Ansprechbarkeit mit erfolgreichen ASW-Versuchen verbunden sind. Zu den vielen möglichen Anzeichen von Einengung oder Zurückweichen in einer bestimmten Situation, gehört es: a) sich als Bock zu bezeichnen, b) in der Schule schlechte Noten zu erhalten, sofern diesen grosse Bedeutung beigelegt wird, c) allgemeine Persönlichkeitsmerkmale für zu starke Selbstkontrolle, wenn sie nicht durch zu starke Ansprechbarkeit ausgeglichen werden. Zu den vielen möglichen Anzeichen von freier Ansprechbarkeit in einer bestimmten Situation gehören: a) sich als Schaf zu bezeichnen, b) allgemeine Tendenzen zu guter sozialer Anpassung in einer derartigen Situation, c) gute Noten in einer derartigen Situation, wenn Noten hoch bewertet werden, d) allgemeine Persönlichkeitsmerkmale für freie Ansprechbarkeit.

RESUMEN

Antecedentes

El trabajo que aquí se describe forma parte de un más ambicioso proyecto, en el cual descubrimiento inicial fué que aquellos sujetos que aceptaban la posibilidad de éxitos paranormales, bajo las condiciones del experimento (ovejas), tendían a tener más altos resultados de ESP que quienes rechazaban esa posibilidad (cabras).

La observación de los sujetos indicó que tanto para los ovejas como para los cabras ciertos factores de la personalidad estuvieron relacionados con los éxitos de la ESP, por cuyo motivo, se efectuaron trabajos posteriores para investigar dichos factores. La técnica empleada principalmente fué la administración de tests proyectivos de la personalidad.

Esta monografía da cuenta de los hallazgos de la investigación de 1062 sujetos a quienes se les administró el Test de Rorschach.

Procedimiento

Como objetivos de ESP se utilizaron listas tapadas de 25 items distribuídos en forma aleatoria. Para los primeros sujetos, los items de la lista representaban los cinco símbolos convencionales de ESP. Los sujetos fueron instruídos para responder a tres listas como una unidad. La sesión standard fué de nueve listas. Breves tests o cuestionarios psicológicos se interpolaron, generalmente, entre las unidades sucesivas. Para los sujetos posteriores, los items de las listas fueron colores y símbolos emparejados. Cada par se formaba de un color (de entre cinco) y un símbolo (de entre los cinco símbolos de ESP). El orden de los colores y los símbolos fué aleatorizado independientemente. Los sujetos fueron instruídos para responder a una lista (de 25 ítems emparejados, o sea, de cincuenta objetivos considerados separadamente) como una unidad. Tests o cuestionarios psicológicos fueron interpolados entre unidades sucesivas. La sesión standard fué de cuatro listas.

Todas las respuestas fueron escritas. Todas tuvieron por lo menos dos confrontaciones independientes de las respuestas con los objetivos.

Todos los sujetos fueron probados en forma colectiva, en grupos formados por los alumnos de una misma aula.

A cada sujeto se le requirió, antes de informarle de sus resultados de ESP, que respondiera un cuestionario de donde resultaba si era oveja o cabra.

El procedimiento standard seguido en la administración del Rorschach fué el de mostrarles proyecciones de las cartas de Rorschach, requiriéndoles que dieran sus respuestas a estas proyecciones, por escrito. En la mayoría de los casos ese procedimiento fué suplementado por entrevistas tendientes a aclarar las respuestas ambiguas.

Excepto en el grupo preliminar de 58 sujetos, todos los Rorschach fueron interpretados por el examinador sin conocer los resultados de ESP.

Para obtener la medida del ajustamiento social se utilizó la Técnica de Inspección de Munroe. Esta técnica consiste en llenar una lista de puntaje de 24 items de Rorschach, adjudicando a cada item uno, dos, o tres tildes, cuando las respuestas caen fuera de los límites normales para ese item. El número total de tildes es una medida inversa del grado de ajustamiento social.

Hipótesis

El examen de los Rorschach del grupo preliminar indicó que los ovejas con menor puntaje en la lista de puntaje (mejor ajustamiento social) tuvieron más altos resultados de ESP que los otros ovejas, y que los cabras con menor puntaje en la lista de puntaje tuvieron resultados de ESP más bajos que los otros cabras. Una hipótesis fué formalmente formulada con relación a este efecto; y los 1004 sujetos subsiguientes fueron el grupo de prueba para esa hipótesis. La línea divisoria entre el buen y mal ajustamiento fué originariamente fijada entre 10 y 11 puntos de la lista de puntaje.

Después de obtenerse los primeros 250 protocolos del grupo de prueba de la hipótesis del ajustamiento, se los examinó para encontrar si otras categorías del Rorschach mostraban signos de estar relacionadas con los resultados de ESP. Siete categorías de marcas se tomaron como "signos" de que era improbable que apareciera la diferencia oveja-cabra postulada. La hipótesis formulada fué que los ovejas cuyos protocolos no presentaban esos signos tenderían a obtener resultados de ESP superiores a los de los demás ovejas; y que los cabras cuyos protocolos no presentaban esos signos tenderían a obtener resultados de ESP inferiores a los otros cabras. Los 754 sujetos subsiguientes fueron el grupo de prueba para esta hipótesis.

Análisis Provisorio de los Datos

Un experimentado analista del Rorschach que no estaba familiarizado con la Técnica de Inspección y el método de computación en el que ésta se basaba, computó independientemente los Rorschach de 83 sujetos. La correlación entre sus listas de puntaje total y las del experimentador fué de + 0,88. Su división de los sujetos en con buen y mal ajustamiento y, dentro de éstos, en con signos o sin signos, dió resultados de ESP de conformidad con ambas hipótesis. Esto indica que ambos métodos de computación eran equivalentes.

La división de los protocolos en dos clases. con puntaje en la lista de puntaje de 10 o menos, y con 11 o más, apoya la hipótesis con un alto grado de significación. Sin embargo, se encontró que los sujetos con 10 y con 11 en la lista de puntaje, es decir, aquellos que estaban inmediatamente por debajo o por encima de la línea divisoria arbitrariamente preestablecida entre el buen y el mal ajustamiento, también mostraban una significativa diferencia en los resultados de ESP, en la dirección prevista. Un reexamen de los protocolos indicó que (a) la mayoría de los protocolos podían legítimamente, dentro de las amplias directivas de la Técnica de Inspección, haber recibido un punto más o un punto menos que el que recibieron; y (b) hubo una tendencia por parte del experimentador a asignar 10 puntos a aquellos protocolos que podían haber sido legítimamente valorados con 10 ó

con 11 y que mostraban tendencias vivazmente reactivas (aún cuando hubiera marcada hostilidad o culpa) y a asignar 11 puntos a aquellos protocolos que podrían haber sido valorados legítimamente como 10 ó 11 y que mostraban coartación y autocrontrol. No hubo razón para pensar que juicios similares de expansividad y retracción afectaran marcadamente el otorgamiento de puntos a las listas de puntaje en ningún otro nivel que el de 10 y 11. Para prevenir la contaminación de la estimación del ajustamiento social con la estimación del grado expansión-retracción, el examen formal de la hipótesis del ajustamiento social en relación con los resultados de ESP puso a todos los sujetos que se encontraban en la zona media del puntaje de las listas de puntaje, incluyendo a aquellos con totales de 10 y de 11, dentro de la misma amplia categoría de ajustamiento.

El examen del puntaje de las listas de puntaje del grupo de prueba para los signos mostró que las mujeres tuvieron un número de puntos significativamente menor que los hombres. (Se sugiere que esto se debe a la mayor proporción de futuros maestros entre las mujeres de este grupo.) Ninguna diferencia de esta clase se encontró en el grupo de prueba de la hipótesis para el ajustamiento. (Se sugiere que esto es debido a la más alta proporción de veteranos entre los estudiantes varones que se probaron primero.) A raíz de la diferencia obtenida en el grupo de prueba para signos, se consideró dudoso si los resultados de los varones y de las mujeres podían legítimamente ser tomados en conjunto. Por tal motivo, se realizaron tests formales para ambas hipótesis, separando los protocolos de los varones de los de las mujeres.

Se obtuvieron resultados de ESP sugestivamente más altos con los sujetos que respondieron a colores y símbolos que con los sujetos que respondieron sólamente a los símbolos. Los análisis de la varianza efectuados separadamente para esos grupos mostraron patrones similares con respecto al puntaje y los signos de las listas de puntaje. A raíz de esto se consideró que sus resultados podían ser juntados en el análisis formal.

Tests Formales de la Hipótesis

El análisis de la varianza del test de grupos mostró:

a) Resultados de ESP significativamente mayores para los ovejas que para los cabras ($P < 0{,}001$)

b) Significativa interacción entre el puntaje de ajustamiento social y la clasificación oveja-cabra ($P < 0{,}001$). Los ovejas con 2-7 y 8-12 en las listas de puntaje tuvieron resultados de ESP más altos que los cabras con 2-7 ó 8-12 en las listas de puntaje; hubo una pequeña diferencia en la dirección opuesta para los sujetos con 13 o más en las listas de puntaje.

c) Significativa interacción entre la presencia o ausencia de signos y la clasificación oveja-cabra, en la dirección predicha ($P < 0{,}001$)

La hipótesis formal es por esto considerada como confirmada por los datos.

Otros Hallazgos

El análisis de la varianza del grupo de prueba para la hipótesis del ajustamiento mostró significativa interacción ($P < 0{,}005$) entre el sexo y la clasificación oveja-cabra. Las mujeres ovejas tuvieron resultados de ESP más altos y las mujeres cabras más bajos que los varones ovejas

y cabras. Esto se interpreta como debido a las condiciones especiales de los tests. La mayoría de los tests fueron administrados por el experimentador, uno de los pocos instructores de sexo femenino de la Facultad. Se infiere que las muchachas, en sus clases, pudieron de esta manera sentir un interés más personal que los muchachos en apoyar o rechazar sus experimentos de ESP no convencionales, y que esa tendencia hacia la identificación o rechazo se haya reflejado en los resultados de ESP.

Sobre la base de los trabajos de Waldron y de Anderson, se formuló la hipótesis de que los sujetos con mayores necesidades intelectuales, en tests de ESP realizados por sus propios instructores, podrían tener más altos resultados si sus notas en el curso habían sido altas que si habían sido bajas. Esta hipótesis se restringió a los ovejas. Dos de los siete signos del Rorschach fueron tomados como indicadores de intensa necesidad intelectual. Para los 101 ovejas cuyas notas fueron conservadas y cuyos protocolos del Rorschach mostraban esos signos, los resultados fueron sugestivamente confirmatorios ($P = 0{,}04$), puesto que los 26 con altas notas tendieron a tener más altos resultados de ESP que los 16 con bajas notas. El efecto pareció ser más pronunciado entre las mujeres que entre los varones.

Dos de los signos fueron interpretados como indicativos de una marcada actividad intelectual (interior), tres como indicativos de una marcada represión o inhibición, y dos como indicativos de una marcada impulsividad o reactividad a estímulos exteriores. El examen de los resultados de ESP de los sujetos con estos signos permite la siguiente tentativa de generalización:

Sujetos con signos de sobrecontrol y de sobrerespuesta, es decir, sujetos con una marca de signos balanceada, contribuyeron a la diferencia oveja-cabra.

Sujetos con signos de sobrecontrol no contrablanceados por signos de sobrerespuesta tendieron a tener resultados de ESP cercanos al esperado por azar.

Sujetos con signos de sobrerespuesta pero sin signos de sobrecontrol tendieron a dar resultados ligeramente superiores al azar, ya fueran ovejas o cabras.

Sujetos con signos tanto de marcada actividad interior como de marcada respuesta a estímulos exteriores tendieron a tener altos resultados de ESP, ya fueran ovejas o cabras.

Conclusiones

Tres hipótesis formuladas con anterioridad a la observación de los datos fueron confirmadas con un nivel de significación superior a $P = 0{,}001$. De este modo se llegó a la conclusión de que, con tests de ESP realizados en condiciones similares a las del presente experimento—: los ovejas tenderán a tener resultados de ESP superiores a los cabras; esta tendencia será más pronunciada para los sujetos cuyo ajustamiento social es pobre; y la tendencia será más pronunciada para los sujetos que no muestran signos de marcada inhibición o marcada sobrerespuesta.

El hecho de que las diferencias obtenidas entre los grupos fuera pequeña y de que hubiese una gran superposición en los puntajes se atribuyó a distintas causas. Es posible que entre ellas se encuentre principalmente el hecho de nuestra falta de información acerca de si las actitudes y las tendencias de las respuestas, inferidas de los tests

psicológicos, influencieron en el estado de ánimo con que los sujetos efectuaron sus respuestas de ESP individuales.

Una conclusión general, no muy segura, acerca de la relación entre la dinámica psicológica y los resultados de ESP es la siguiente: Los sentimientos de coartación y repliegue están asociados con los resultados de ESP cercanos a lo esperado por azar o con psi-missing; y los sentimientos de libertad de respuesta están asociados con los resultados exitosos de ESP. Entre los varios indicadores posibles de coartación y repliegue en una situación particular se encuentran:

a) llamarse a si mismo un cabra;

b) obtener bajas notas escolares en una situación similar si las notas son altamente estimadas;

c) mostrar tendencias generales de la personalidad hacia el sobrecontrol si no hay, como contrapartida, tendencias hacia una sobrerespuesta.

Entre los varios posibles indicadores de libertad de respuesta en una situación particular se encuentran:

a) llamarse a si mismo un oveja;

b) mostrar tendencias generales hacia un buen ajustamiento social en una situación similar;

c) obtener buenas notas escolares en una situación similar si las notas son altamente estimadas;

d) mostrar tendencias generales de la personalidad hacia una libertad de respuesta.

SOMARIO

Premessa

Questo lavoro non è che una parte di uno studio più vasto—la cui prima conclusione fu che i soggetti che ammettevano la possibilità di risultati paranormali nelle condizioni dell'esperienza ("pecore") tendevano a dare cifre di ESP superiori a quelle dei soggetti che respingevano tale possibilità ("capre"). Lo studio dei soggetti indicò l'esistenza di un rapporto tra i fattori di personalità e i successi in ESP—tanto per le capre che per le pecore—; e le indagini sono state continuate in vista dell'esame di tali fattori. La tecnica principale consisté nell'applicazione di *tests* proiettivi di personalità. La presente monografia riferisce le conclusioni relative ai 1062 soggetti che vennero esaminati con il *test* di Rorschach.

Procedimento

I "bersagli" ESP furono liste di 25 immagini-stimolo nascoste e mescolate. Per i soggetti di una prima serie, le immagini rappresentavano i 5 simboli ESP convenzionali. Le istruzioni impartite ai soggetti stabilivano che essi dovessero rispondere a tre liste considerate come un'unità. La "seduta-tipo" constava di 9 liste. Brevi prove psicologiche o questionarî furono introdotti, per solito, fra due unità successive. Per i soggetti esaminati in un secondo tempo, gli stimoli consistettero in colori e simboli associati. Ogni paio era costituito da uno dei cinque colori e uno dei cinque simboli ESP. L'ordine dei colori e dei simboli era stato determinato indipendentemente e a caso. Le istruzioni impartite ai soggetti stabilirono come unità la risposta a una sola lista (di venticinque paia, ossia di 50 singole immagini-stimolo). Fra unità successive furono introdotti *tests* psicologici e questionarî. La "seduta-tipo" includeva 4 liste.

Tutte le risposte furono scritte. Per tutte, furono date almeno due possibilità indipendenti di punteggio nei riguardi dell'immagine-stimolo. Le prove, per tutti i soggetti, si svolsero nell'ambito di gruppi e classi di College. Prima d'essere informati sui loro risultati di ESP, tutti i soggetti dovevano rispondere ai quesiti indicanti se erano pecore o capre.

Il procedimento-tipo per l'applicazione del Rorschach fu quello di far passare delle lastre rappresentanti le immagini Rorschach, e chiedere ai soggetti di scrivere le loro risposte. Nella maggior parte dei casi, il resoconto scritto fu completato da un'intervista, durante la quale il soggetto veniva interrogato sulle risposte ambigue. Tranne che per il gruppo preliminare di 58 soggetti, le note relative ai Rorschach furono prese da chi esaminava senza previa conoscenza dei risultati di ESP.

La Tecnica Ispettiva di Munroe fu adoperata per misurare l'adattamento sociale. Questa tecnica consiste nel riempire una lista-questionario di 24 elementi Rorschach, attribuendo a ciascuno 1, 2 o 3 punti se le risposte del soggetto escono dai limiti normali relativi a tale elemento. Il totale dei punti è in rapporto inverso con l'adattamento sociale.

Ipotesi

L'esame dei Rorschach del gruppo preliminare indicò che le pecore che totalizzavano meno punti nelle liste-questionari (miglior adattamento sociale) davano risultati ESP superiori a quelli delle altre pecore, e che le capre che totalizzavano meno punti nelle liste davano risultati ESP più bassi di quelli delle altre capre. A tale riguardo è stata formulata un'ipotesi, e i successivi 1004 soggetti hanno costituito il gruppo di verifica di tale ipotesi. La linea di demarcazione tra adattamento buono e adattamento cattivo fu fissata dall'inizio tra i 10 e gli 11 punti della lista-questionario.

Dopo che furono raccolte le 250 prime registrazioni del gruppo di verifica dell'ipotesi relativa all'adattamento, esse furono esaminate per vedere se qualche altra categoria di Rorschach poteva essere in rapporto con i risultati di ESP. Furono prese 7 categorie come "segni" che la presunta differenza pecore-capre non sarebbe dovuta apparire. Si formulò l'ipotesi secondo la quale le pecore le cui risposte non comprendevano tali segni tenderebbero a dare risultati di ESP più alti delle altre pecore, e che le capre le cui risposte non comprendevano quei segni tenderebbero a dare risultati di ESP più bassi delle altre capre. I successivi 754 soggetti formarono il gruppo di verifica di tale ipotesi.

Analisi Provvisoria dei Dati

Una specialista con grande esperienza di Rorschach, che non si era ancora familiarizzata con la Tecnica ispettiva e con i criteri di valutazione numerica su cui essa si basa, notò indipendentemente i Rorschach di 83 soggetti. Il rapporto fra i totali delle sue liste e quelli della sperimentatrice fu di 4,88. La sua divisione dei soggetti in bene adattati o male adattati, e in soggetti con segni e soggetti senza segni, diede risultati di ESP conformi all'una e all'altra ipotesi. Ciò indica che il criterio di valutazione numerica è comunicabile.

La divisione dei risultati tra quelli che comportavano 10 risposte o meno alla lista-questionario e quelli che ne comportavano 11 o più conferma l'ipotesi con un alto grado di significatività. Tuttavia si è trovato che dei soggetti che avevano 10, oppure 11, risposte alla lista (e cioè che si trovavano appena sotto o appena sopra la linea arbitrariamente tracciata fra buon adattamento e cattivo adattamento) mostravano anch'essi una differenza significativa dei loro risultati ESP nella direzione prevista. Un nuovo esame delle registrazioni indica che:

a) si sarebbe potuto legittimamente assegnare un punto di più o di meno alla maggior parte dei totali senza uscire dai larghi limiti della tecnica ispettiva;

b) che la sperimentatrice aveva la tendenza ad assegnare 10 punti a quei totali che avrebbero potuto legittimamente essere valutati 10 o 11 e che mostravano una viva reattività (anche se c'era spiccata ostilità o colpevolezza); e ad assegnare 11 punti a quei totali che avrebbero potuto legittimamente essere valutati 10 o 11 e in cui si manifestava

costrizione e auto-controllo. Non abbiamo motivo di pensare che simili giudizi di vivacità e di costrizione abbiano influito in modo apprezzabile sull'attribuzione delle risposte della lista-questionario ad alcun altro livello tranne che 10 o 11. Per evitare qualsiasi contaminazione delle valutazioni dell'adattamento sociale con quelle di vivacità-costrizione, l'esame formale dell'ipotesi dell'adattamento sociale in relazione con i risultati ESP ha messo tutti i soggetti appartenenti alla zona media delle risposte alla lista-questionario (compresi coloro che avevano totali di 10 e di 11) in una sola grande categoria di adattamento.

L'esame delle risposte alla lista-questionario del gruppo di verifica dei segni ha mostrato che le donne davano un numero di risposte significativamente più basso degli uomini. (Si suppone che ciò sia dovuto alla forte proporzione di future insegnanti nelle donne di questo gruppo). Non si è trovata una differenza di tal genere nel gruppo per la verifica dell'ipotesi dell'adattabilità. (Si suppone che ciò sia dovuto alla forte proporzione di ex combattenti fra gli studenti esaminati in un primo tempo). Data la differenza ottenuta nel gruppo di verifica dei segni, si è considerato dubbio che i risultati relativi a uomini e donne potessero legittimamente essere sommati. Le prove formali delle due ipotesi hanno perciò separato i risultati dei primi e quelli delle seconde.

Cifre ESP più elevate sembrano essere state trovate con soggetti che reagivano ai colori e ai simboli rispetto a quelli che reagivano ai soli simboli. Analisi separate delle variazioni per questi gruppi hanno mostrato modelli simili per ciò che riguardava le risposte alla lista-questionario e ai segni. Per questa ragione, fu deciso che i loro totali potevano essere addizionati nelle analisi formali.

Prove Formali Delle Ipotesi

L'analisi delle varianti dei gruppi di prova ha mostrato:

a) risultati ESP significativamente più alti per le pecore che per le capre ($P < 0{,}0001$).

b) interazione significativa tra le cifre relative all'adattamento sociale e la classifica pecore-capre ($P < 0{,}0001$). Le pecore che avevano da 2 a 7 e da 8 a 12 risposte alla lista-questionario hanno avuto cifre di ESP superiori a quelle delle capre aventi da 2 a 7 e da 8 a 12 risposte alla lista-questionario; c'è stata una lieve differenza nel senso opposto per i soggetti le cui risposte alla lista-questionario totalizzavano 13 o più.

c) interazione significativa tra la presenza e l'assenza di segni e la classifica pecore-capre, nella direzione prevista ($P < 0{,}001$).

Le ipotesi formali sono pertanto considerate come confermate dai dati.

Altre Constatazioni

L'analisi delle variazioni per il gruppo di verifica dell'ipotesi dell'adattamento mostra un'interazione ($P < 0{,}005$) tra il sesso e la classificazione pecore-capre. Le pecore donne hanno raggiunto cifre di ESP superiori, e le capre donne cifre di ESP inferiori, rispettivamente a quelle delle pecore uomini e delle capre uomini. Ciò viene interpretato come dovuto alle condizioni speciali del test. Per la maggior parte, i tests venivano curati dalla sperimentatrice, una delle poche donne-istruttori del College. Noi supponiamo pertanto che le ragazze

delle rispettive classi dovessero provare un interesse personale a sostenere o a respingere le sue esperienze "non ortodosse" di ESP, più di quanto non facessero gli uomini, e che questa tendenza all'identificazione o al rifiuto—si sia riflettuta nei risultati di ESP.

Partendo dai lavori di Waldron e di Anderson, è stata emessa l'ipotesi che i soggetti aventi un vivo desiderio di successo intellettuale, e che venivano sottoposti alle prove di ESP da parte del loro stesso istruttore, avrebbero raggiunto risultati di ESP più elevati se i loro voti nella sua classe fossero stati alti, e viceversa. Questa ipotesi fu limitata alle pecore. Due dei 7 segni di Rorschach sono stati presi come indicatori di un vivo desiderio di successo intellettuale. Per le 101 pecore i cui voti sono stati registrati e i cui protocolli Rorschach hanno mostrato quei segni, i fatti indicano l'esattezza ($P = 0{,}004$) dell'ipotesi, nel senso che i 26 che avevano voti alti hanno mostrato tendenza a dare risultati di ESP migliori dei 16 con voti bassi. L'effetto sembra essere stato più pronunciato con le ragazze che non con i giovani.

Due segni sono stati interpretati come indicazioni di notevole attività intellettuale (interiore), 3 come indicatori di ritegno o di notevole inibizione, e due come indicatori di impulsività, o di notevole reattività agli stimoli esteriori. Lo studio dei risultati di ESP dei soggetti che presentavano tali segni ha portato alle seguenti generalizzazioni provvisorie:

—i soggetti che presentavano un segno di eccesso di auto-controllo insieme con un segno di eccesso di reattività, e cioè i soggetti con situazione-segni equilibrata, partecipavano alla differenza pecore-capre.

—i soggetti che presentavano un segno di eccesso d'auto-controllo senza alcun segno di eccesso di reattività che lo controbilanciasse hanno avuto tendenza a dare cifre di ESP prossime alla probabilità.

—i soggetti che presentavano un segno di eccesso di reattività senza alcun segno di eccesso d'auto-controllo che lo controbilanciasse hanno avuto tendenza a dare cifre di ESP un po' al disotto della probabilità, sia che fossero pecore, sia capre.

i soggetti che presentavano insieme segni notevoli di attività interiore e di reattività agli stimoli esterni hanno avuto tendenza a dare cifre elevate di ESP, sia che fossero pecore, sia capre.

Conclusioni

Tre ipotesi, enunciate prima della raccolta dei dati, sono state confermate con un grado di significatività superiore a $P = 0{,}001$. Se ne conclude che, in prove di ESP condotte in condizioni simili a queste, le pecore avranno tendenza a raggiungere risultati di ESP più alti di quelli delle capre; questa tendenza sarà più pronunciata per i soggetti aventi un buon adattamento sociale che non per quelli il cui adattamento sociale è cattivo, e la tendenza sarà più pronunciata per i soggetti che non presentano segni di notevole inibizione o eccesso notevole di reattività. Il fatto che le differenze constatate fra i gruppi sono state deboli, e che si è prodotto un buon numero di accavallamenti tra i risultati, può essere attribuito a varie cause. La principale—crediamo—è la nostra mancanza d'informazioni sul seguente punto: se, cioè, gli atteggiamenti e le tendenze a "rispondere," dedotti dai *tests* psicologici, hanno influito sull'umore dei soggetti nel momento in cui ognuno dava le sue risposte personali nelle prove di ESP.

Le conclusioni generali formulate, non senza esitazione, circa il rapporto fra la dinamica psicologica e il successo in fatto di ESP, sono:

—i sentimenti di costrizione e di ripiegamento si associano ai risultati di ESP vicini alla probabilità, o al cosiddetto "*psi* missing"; e i sentimenti di reattività libera si associano al successo in ESP.

Tra i diversi indicatori possibili di costrizione o di ripiegamento in una particolare situazione, si trovano:

a) l'impressione di essere una capra;

b) l'ottenere, in una tal situazione, voti di studio bassi, se si attribuisce un grande valore ai voti;

c) una personalità la cui tendenza generale è l'eccesso di auto-controllo, mentre non vi è una tendenza all'eccesso di reattività che la controbilanci.

Nel novero dei diversi indicatori possibili di libera reattività in una situazione particolare si trova:

a) l'impressione di essere una pecora;

b) una tendenza generale al buon adattamento sociale in una situazione simile;

c) l'ottenere, in una tal situazione, buoni voti di studio, se si dà grande importanza ai voti;

d) una personalità la cui tendenza generale è la libera reattività.

REFERENCES

Anderson, M. The relationship between level of ESP scoring and student class grade. J. Parapsychol., 1959, 23, 1-18.

Boring, E. G. Statistical frequencies as dynamic equilibria. Psychol. Rev., 1941, 48, 279-301.

Hartley, R. E. The acceptance of new reference groups; Technical Report 2. Distributed under Contract nonr-1597 (01). Washington, D. C.: Group Psychology Branch of the Office of Naval Research, 1956.

Humphrey, B. M. The relation of some personality ratings to ESP success: a review of recent research, J. Soc. psychic. Res., 1951, 36, 453-66.

———, Introversion-extraversion ratings in relation to scores in ESP tests. J. Parapsychol., 1951, 15, 252-62.

Klein, G. S. Need and regulation. In Jones, M. R. Nebraska Symposium on Motivation. Lincoln: Univ. of Nebraska Press, 1954, 224-74.

Munroe, R. L. Prediction of the adjustment and academic performance of college students by a modification of the Rorschach method. Appl. Psychol. Monog., No. 7, 1945.

———, Private Communication.

Murphy, G. Personality. New York: Harper and Brothers, 1947.

Nash, C. B. and Nash, C. S. Checking success and the relationship of personality traits to ESP. J. Amer. Soc. psychic. Res., 1958, 52, 98-107.

Nicol, J. F. and Humphrey, B. M. The repeatability problem in ESP-personality research. J. Amer. Soc. psychic. Res., 1955, 49, 125-157.

Rasch, W. D. Psychodiagnostic examination of sensitives. Proceedings of the First International Conference of Parapsychological Studies. New York: Parapsychology Foundation, Inc., 1955.

Schmeidler, G. R. Rorschach variables in relation to ESP scores. J. Amer. Soc. psychic. Res., 1947, 41, 35-64.

———, Additional data on sheep-goat classification. J. Soc. psychic. Res., 1959, 40, 63-72.

Schmeidler, G. R. and McConnell, R. A. ESP and Personality Patterns. New Haven: Yale Univ. Press, 1958.

Schmeidler, G R., Nelson, M. J. and Bristol, M. Freshman Rorschachs and college performance. Genet. Psychol. Monog., 1959, 59, 3-43.

Waldron, S. Clairvoyance of sheep versus goats when subject's attitudes toward the experiments and the purpose of the experiment are manipulated. Unpublished thesis, Harvard University, 1958.