TELEPATHY IN THE WAKING STATE

AN EXPERIMENTAL DESIGN

THELMA MOSS, PH.D.

The Neuropsychiatric Institute
University of California
Los Angeles, California

LUDWIG: Dr. Thelma Moss will speak on telepathy experiments in the waking state. Following her presentation, which will include a demonstration of the stimulus material she uses, Dr. Strauch and Dr. Kamiya will comment on her work. Dr. Moss, please.

MOSS: I am going to describe six different experiments that have extended over a period of four years. During that time adjustments were made in technique as the work progressed. So what I am going to present to you is not a specific experiment that has been replicated six times, but rather variations on a theme.

The paradigm that seemed to evolve from reading the literature on spontaneous telepathic experiences was that the "receiver" is usually in a relaxed state, either generating a lot of alpha rhythm, or sleep, or in a hypnagogic fantasy state—in some kind of passive, open relaxation. The transmitter, on the other hand, is almost invariably in a highly charged emotional state. For example, when Mrs. Garrett got an impression of Aldous Huxley's home burning, I am sure that his emotional arousal was very high at that particular moment, as that was an extremely tragic episode for him. Using such a model, I tried to create in the laboratory a situation in which the transmitter would be actively aroused and the receiver would be passive but open, and willing to associate freely. These were the basic conditions. Superimposed on this basic paradigm was the problem of what kind of person would respond best in a telepathic situation, if it were demanded of him.

My pilot study was done primarily to see whether a measurable physiological reaction aroused in a person exposed to a highly charged emotional stimulus would be replicated in a "receiver." I also introduced the "relatedness" versus "non-relatedness" variable, hypothesizing that pairs of individuals who had a close relationship prior to the time of the experiment would be more likely to show telepathic communication than pairs of individuals who did not know each other. I used 13 transmitter-receiver pairs (6 teams of strangers, 7 teams of closely-related individuals), both team members being "wired" for measures of respiration, heart beat, GSR, and plethysmography (finger blood volume).

The subject serving as transmitter (T) was seated in an isolation booth facing a blank screen. He was equipped with earphones through which he could hear sounds accompanying slides projected onto the screen. At the end of each slide-sound episode, his comments on what he had experienced were tape recorded.

The receiver (R) was asked to relax by lying down in a tilt-back chair in another room. At a given signal (synchronized with the start of the "show" to the transmitter), the receiver was asked by another experimenter (E_2) to "free associate," i.e. to say anything which came into his mind, images, thoughts, sensations or feelings. These free associations were recorded on another channel of the same tape recorder. Each trial lasted about 5 minutes, including a 2-minute interval between episodes. Some form of reinforcement was given by informing the Rs of their performance after the whole experiment was over.

The four emotional episodes to which each T was exposed were chosen to elicit different kinds of emotion and were shown in a random sequence: no one except the experimenter (E_1) knew which episode was shown at which moment. Here is a brief description of these first four stimuli:

1) *Disneyland*: A series of slides showing Disneyland: cable cars, Matterhorn, Mad Tea Party, children, balloons, etc. Honky tonk music. Expected reaction: pleasure, amusement.

2) *Frustration*: E_1 asked T to multiply mentally two 2-digit numbers and give the answer within half a minute. Expected reaction: frustration, pressure. (This episode was particularly weak; the expected reaction did not always occur. One T gave the correct answer in 10 seconds, and was very pleased to solve more problems within the time allotted.)

3) *Fear*: An eerie musical composition interlaced with heart beats, groans and gasps, climaxed with a man's piercing scream and a very startling picture of a coiled snake with fangs bared. Expected reaction: startle, fear.

4) *Colors*: T was asked to watch a stroboscope's flickering white light until it changed into brilliant colors, as the frequency was being manipulated. Expected feelings: pleasure, excitement. (This was the least successful episode; several Ts complained that the light hurt their eyes and never saw colors.)

Quantitative analysis of the data was obtained by asking six judges (professional psychologists and psychiatrists) to blindly match the randomized Rs' with the Ts' protocols. This judging procedure is similar to the one used by the Maimonides group. The matching was essentially an intuitive one, the judges making their guesses on the basis of their knowledge of psychoanalytic theory, symbolic representation and distortion. The judges' correct matchings were then counted up. The Rs' responses fell roughly into three categories: 1) exactly correct (identical to Ts' protocols) or veridical responses—these were very rare occurrences; 2) responses incorporating "primary process" symbolism, or other interpretative distortions of Ts' reactions; 3) fantasies having no relation to anything said by the T—this was the most frequent type of response.

As to the results of this pilot study, the psychological reactions of the T-R pairs as expressed in their verbalizations were similar enough, as matched by the judges, to reach a .001 significance. Some telepathic communication did seem to be operating. However, there were several unexpected findings in this experiment. First, there was a negative correlation in all physiological measures: as T, receiving highly charged emotional stimuli, became increasingly aroused, R, lying down, became increasingly relaxed. Second, there was no difference between the closely-related and the non-related groups. Last, there were time displacements in Rs' responses. For example, one R said: "I saw balloons . . . and I was riding around in a teacup—honestly! It felt so strange, riding around in a teacup!" *Disneyland* was not shown to the T of this team until two episodes later. A few times, an R described an episode after the T had seen it.

On the basis of the results of the pilot study it was decided to evaluate the emotional impact through T-R verbalizations only, not using physiological measures, in the next experiment.[1] Several other changes were incorporated in the experimental design:

1) To give Rs, who usually envied Ts' role, a chance to participate more actively, it was decided to use all Rs as Ts after they had served as Rs. The reverse was avoided for obvious reasons.

2) As *Disneyland* and *Fear* had been the most successful episodes, it was decided to use slides-and-sound in all subsequent work to attack two sensory modalities at the same time.

3) Controls were introduced. One consisted of teams whose Ts saw no picture; responses from Rs were recorded as usual. Independent T responses to the episodes were recorded separately at a later time, for matching purposes. In another control group, Rs' responses were recorded while a mannikin was being shown the episodes. The judges matched them against a description of the slides, rather than against both slide descriptions and T responses as in the experimental groups.

4) Despite the lack of difference between performances of related versus non-related teams, it was decided to keep studying this variable, since the vast majority of spontaneous cases occur between strongly related persons.

5) For reasons of statistical evaluation, it was decided to increase the number of episodes from four to six. Some new episodes were created, and the ones used are listed here, in order of number of correct matchings elicited:

A) *JFK Assassination*: This tragedy had occurred nine months earlier. Expected reaction: grief, horror.

B) *Ladies*: A song, "The Stripper," accompanying several slides of partially undraped or nude girls. Expected reaction: erotic, sensual feelings.

C) *Hawaii*: The ocean from dawn to sunset, chiefly around Hawaiian Islands; surf, sailboats, bathers. Throughout, Respighi's "Fountains of Rome." Expected reaction: serenity, exhilaration.

D) *Disneyland*: As in the pilot study.

E) *Nazis*: A passage from Mossolov's "Iron Foundry" repeated over and over, discordant and piercing, acompanying slides of concentration camps. Expected reaction: disgust, hatred.

F) *Fear* (as in the pilot study); or *Cold*: T was asked to keep the left foot in a bucket of ice and water for 40 seconds. We still thought that a strong physical reaction might be picked up by Rs. (The poor results obtained dispelled this idea, and no physical stimuli were used after that time.)

In the summer of 1964, 93 persons participated in this second experiment. The results again showed no difference between related and non-related teams. Six judges matched the first control group (Rs without Ts) far less successfully than the genuine T-R teams. The mannikin control group was matched by six other judges at a level slightly above chance. This may be due to the fact that E_1, who was administering the stimulus episodes (to the Ts or the mannikin), unwittingly might have acted as an extra T.

On certain occasions R impressions reflected very accurately the episodes being transmitted. In a few cases the descriptions were clearly referring to an episode already shown or still to come. Much more frequently judges made their choices on the basis of symbolic representation. For example, sexual feelings elicited by the *Ladies* episode were described by one R as "a volcano erupting" and by another as ". . . Whiskey à Go Go, one of the nightclubs on the Strip."

In one instance, when Dr. Gardner Murphy (as T) was shown this episode, the R (who was a firm non-believer in ESP) said: "I have the impression of a small, curved stone bridge with a single-masted sailboat sailing underneath. Then a sort of rubbing feeling along the side or end of the bridge. That's all." This does not seem like very much, but if you are a psychiatrist or a psychologist looking for imagery, this would have more sexual implications than many other kinds of comments. Anyway, all six judges matched these responses correctly.

To bypass the shortcomings inherent in the matching procedure, we wanted to find some way by which the Rs would score themselves. While investigating ways to eliminate judges from the design, I was asked to hold a lecture at Stanford University. Rather than a formal lecture, an informal experimental demonstration was requested. As time would limit the number of individual T-R teams which could be tested, we decided to use a group of people as Ts, and a different group as Rs, in two separate rooms. The members of each group, instead of speaking their impressions, would write them down independently from each other. After each trial, the Rs were shown five slides (four dummy slides plus the one which had been shown to the T group); without consultation, each R wrote down which slide corresponded best to his own impressions.

Even in this inadequately controlled informal design, the forced-choice rating procedure proved so simple and effective that it was decided to introduce it in our subsequent studies, with some modifications. Three pairs of contrasting episodes were used. In each session one member of each pair, chosen randomly, was shown to the T. After each trial that R matched his own impressions against two contrasting slides shown simultaneously, one corresponding to the episode seen by the T, the other to the one not used. The chances of picking the correct slide for any particular episode would be 50%, but the limitation inherent in such a high probability was by-passed by using a greater number of subjects and trials. The episode-pairs in our third experiment[2] were as follows:

1) *Madonna* (various religious pictures of Madonna and Child; music: "Silent Night." Expected to convey maternal, reverent feelings) and *Van Gogh* (self-portraits, accompanied by music and a voice saying, "Van Gogh" with increasing intensity).

2) *Space* (rockets, satellites, astronauts, and electronically distorted music. Expected reaction: excitement, fascination) and *Drunk* (man and glass, with distorted music. Expected reaction: amusement).

3) *Wild Animals* (various voracious beasts devouring their prey, music, screams and roars. Expected reaction: fear) and *Tree* (series of Mondrian paintings and a poem by Joyce Kilmer, to convey an impression of expectation, of building towards accomplishment).

Also introduced into this experiment, which used 72 individual teams, was the sheep-goat dichotomy. The subjects were asked beforehand whether they did believe in ESP, did not believe in ESP, or were undecided, and were divided on the basis of their responses into "ESP," "No ESP," and "ESP?" The related/non-related variable was also kept in the design. Results showed that relatedness again made no difference. ESP believers did show more successful individual performances than did non-believers, even if their results as a group were not significant, but the "ESP?" group did not do significantly better than the "No ESP" group.

While collating the data, after the experiment had been completed, another division of the teams was attempted, "creative" and "non-creative," according to the profession of the receiver. We re-examined the protocols on the basis of the post hoc hypothesis that creative people's performance would be better.[3] The results were spectacular; of the 26 teams including "creative" Rs, 25 scored two or three out of three correct.

It was then decided to do an experiment[4] in which the hypothesis would be made beforehand that creative T-R teams would do better than non-creative ones. We used the forced-choice design, with six new episodes, divided into three pairs of contrasting emotional tone. These were, as paired: *Crucifixion-Sports*; *Israeli War-Rodin Love Sculptures*; *LSD Trip-Race Riots*. We again divided the teams into related and non-related, and also according to their own personal experience with ESP. Thus, "ESP" and "No ESP" referred to actual experience or no experience, rather than belief per se. This experiment proved overall so successful that the issue of what kind of Rs perform best is still unsolved. All performed significantly above chance, whether creative or non-creative, with or without ESP experience, related or strangers.

We then ran a long-distance study between Los Angeles, New York City, and Sussex (England). In this fifth experiment[5] as in the following

one, we used *group* T-R teams. The procedure utilized was essentially the same as in the short-distance studies, with the modifications introduced during the Stanford demonstration, and the 2-slide forced-choice. There were 28 Rs in Los Angeles, 15 in New York, 14 in Sussex (57 in all) and 22 Ts in Los Angeles. The six episodes were:

Water Sports (pleasurable aquatic activities accompanied by Debussy's "La Mer")—*Van Gogh* (described previously).

Space-Wild Animals (both described previously).

War (Israeli-Arab conflict scenes with appropriate sound track)—*Love* (Rodin's sculptures of human bodies, with music from Holst's "Venus").

There was a total of six trials in the experiment. In three of them one episode out of each pair was shown to the Ts; in the other three control trials nonsense material (such as random numbers, letters, or slanted lines) was projected onto the screen of the Ts. The Rs, however, were still given a choice of two "emotionally charged," "phony" slides, not corresponding to anything seen by the Ts. If ESP were really operative, it was thought that there would be considerably greater accuracy in choosing the correct slide for the legitimate episodes than for the phonies.

Our hypotheses were the following:

1. That distance would not make any significant difference in accuracy of reception.

2. That Rs would choose the right slides more accurately for the real episodes than for the phonies.

3. That subjects in a creative profession would do better than non-creative ones.

4. That receivers with ESP experience would do better than those without such experience.

The factor of relatedness was not investigated in long-distance experiments.

The best results were shown by the Los Angeles group (nearest), the Sussex group (farthest away) being the next best, and the New York one poorest. Receivers did make the correct choices for the real episodes significantly more accurately than for the phonies, as predicted. However, the other two hypotheses were not borne out by the data.

Our sixth and last project was also a long-distance experiment, involving Los Angeles, Stanford, Fresno, Berkeley, Durham (N.C.), and Edinburgh, with a procedure similar to that used in the experiment I just described. There was a total of 45 Ts, and 99 Rs divided in six

groups. The episode sequence was as follows: first *Marilyn Monroe* (actually shown)—*Soviet Union* (army parades, etc., with corresponding music); the second was a phony; the third was a choice between *Africa* (actually shown) and *Space*; the fourth *Racing Cars* (actually shown)—*Skin Diseases*; the fifth and sixth were phonies. Statistics from the forced-choice failed to show any significant results. In general, content analysis of the Rs' impressions showed no correlation with the targets.

The qualitative analysis of the data was still the crucial problem. In order to make it more rigorous, we tried to introduce a systematization of the verbal material collected. We decided to define each episode by means of 15 key words. Then we simply counted how many of them had been used by the R in his associations to that particular target. To determine whether the words chosen by the Rs were specific to what was shown to the Ts, we examined how many times the same words occurred in the verbalizations of other Rs in another experiment. We think that the problem of assessing the value of the Rs' responses in a qualitative way somewhat accurately is a very difficult one and our 15 words technique certainly did not solve it.

I will now show you two episodes which were actually used in the course of the experiments described, namely *Marilyn Monroe* and *Africa*, in the hope that you will be able to appreciate the type of stimulus to which the Ts were exposed.

[*DEMONSTRATES*]

STRAUCH: Before commenting on your experiments, I would like to ask you to clarify a few points in your procedure. The stimulus selection you made for your experiments is very interesting and undoubtedly represents a progress over the Zener cards. But, if I understand correctly, when you say that the experimental results were significant, you disregard the qualitative assessment and refer only to the Rs' forced choice or to the judges' ranking in your first experiments. What indication do you get from this treatment of your data that the subjects not only made a good guess in choosing between two targets, but that the emotional impact of the targets actually contributed to the Rs' decisions?

MOSS: The nature of their responses gives us a clue in this direction.

STRAUCH: I remember from one of your papers the example of the Kennedy assassination, where the receiver's response corresponded almost word-for-word to one of the episodes shown. At that time you could

not exclude that your subjects knew the target pool, therefore the qualitative aspects of their responses could not be taken into consideration.

Moss: Right. Therefore I have changed the design.

Strauch: In order to substantiate your quantitative evaluation of qualitative results, it is important to prove that it was the emotional impact which produced the occurrence of significant results. An emotionally charged target (as the ones you showed us) might arouse a wide range of emotions. I think that it would be extremely difficult to describe a uniform emotion being aroused in different persons by these stimulus episodes.

Moss: There were individual differences in the Ts' responses, but the important thing is what the Rs wrote down. If they wrote down something like, "I see a picture of a blond lady in a nightclub," and then they saw a picture of *Marilyn Monroe* and a picture of *Space*, they chose Marilyn Monroe, because this corresponded to their free associations.

Strauch: But how do you assess whether the emotion was the motor for the right choices? Maybe they were pure guesses, where emotion did not play a part.

Moss: As I said, the choice in the forced-choice situation is dictated by the feelings previously described in the free associations and by the similarities of one of the slides to these free associations.

Cavanna: I would be very grateful to you if you could clarify how you deal with the qualitative assessment of the Rs' responses. You used judges only in the pilot study and in the first major study, is that right?

Moss: Yes. However, in the later experiments, we had somehow to evaluate qualitatively the content of the protocols, with respect to the actual content of the slides.

Strauch: But in these later experiments you also used 15 words to characterize each contrasting episode. How did you make use of them? How did you select the 15 words?

Moss: The 15 words were not used until the group experiments. Let me describe briefly how they were selected.

I showed all the episodes to a group of five students and asked them to write down independently of each other all the words that came to their minds as representative of each sequence of slides. We

then pooled all the words and found that certain words had been written by most people. We would use them. Once we had the 15-word lists, we took the Rs' protocols and checked to see if any of the words appeared in them.

STRAUCH: Did you use the same lists for the controls?

MOSS: Yes.

STRAUCH: So this would really be an indication that it was the qualitative aspect of the target which led to the association?

MOSS: Yes.

VAN DE CASTLE: You say that you did not use judges in your later experiments, but what you describe seems like a judging procedure. Someone had to rate whether the free associations included any of the 15 tag words.

MOSS: It was not a judging. It was a simple counting procedure.

VAN DE CASTLE: Yes, but sometimes words can be used synonymously. For example, the tag word could be "suicide" but somebody would have to decide whether "self mutilation" could be considered a synonym for "suicide." I would call that a judging situation.

MOSS: In that sense it is. There were two independent counters who were actually my research assistants, and they always agreed. Nevertheless, using the 15-word lists presents many problems. We are trying to simplify the procedure. While in most of the experiments the quantitative statistical analysis showed significant results, I do not consider that to be the most important aspect of my experiments. Dr. Rhine has worked for more than 30 years using such an approach. What seemed to me relevant about my design was the opportunity to get verbalizations or free associations. I had to find a way to evaluate them.

I do believe that responses which can be considered strong evidence for ESP are limited. They could never become statistically significant even using the great numbers of people we do. When ESP is present, it comes through in the choice of words that people use. So it seemed a good idea to track down the number of relevant words that people would write pertaining to one particular episode. This method presents a lot of difficulties.

Let me give you an example of the word lists used for two contrasting episodes, which were paired.

SOVIET UNION

1. power (strong, struggle)
2. red
3. weapon (missile)
4. patriotism
5. peasants (workers, masses)
6. Russia (USSR, Soviet)
7. parade (festival, May Day)
8. pride (spirit, triumph)
9. anger (aggression)
10. military (army)
11. Communism
12. foreign
13. fear
14. march (marching song)
15. politics (leadership)

MARILYN MONROE

1. actress (star)
2. blond
3. roses
4. Marilyn Monroe
5. sad (tragic)
6. blues
7. woman
8. fame (famous)
9. movies (films)
10. roles (parts)
11. drugs
12. drinking (wine)
13. neurotic (sick)
14. glamour
15. suicide (death)

As you see, we had to include some synonyms which conveyed the same feeling or mood. If the *Marilyn Monroe* episode had been shown, for example, we considered it evidence for ESP if more words characteristic of *Marilyn Monroe* were used than words characteristic of the *Soviet Union* episode. Results were encouraging, but inconclusive.

OWEN: May I call your attention to the fact that in making up a list of attributes like the ones you showed us, you would have to make them independent of each other. You have "actress," "blond," "woman" and "glamour" in the same list. They tend to be correlated and this correlation might boost your significance.

MARGENAU: Another danger would be with words straddling the two episodes. They would also give rise to falsely high correlations.

STRAUCH: I agree with Dr. Moss that the important aspect of her experimental design is certainly not the quantitative statistical analysis of the receivers' ability to determine which one of two slides corresponded to the target episode presented to the transmitter. Although she collected some significant data with regard to this aspect, their impact is rather limited. Thousands of experiments have been done with such an approach, not only involving much larger numbers of trials and subjects, but also usually relying on an initial probability considerably smaller than 0.5.

In my opinion, the crucial question is to observe whether there is a co-variance between the verbalizations of the receivers and those of the transmitters (which hopefully should co-variate with the targets).

Dr. Moss's methodological variations on this theme reveal the general difficulties which we meet when we try to submit qualitative material to an objective assessment. Dr. Owen just pointed out that the proposed lists of attributes should only include items which are independent of each other. Similarly, difficulties may arise if complex slides are used in the target episodes. First, because complex stimuli may arouse complex feelings, giving rise to equivocal responses; and second, because many small details appearing in slides pertaining to a given episode may prove not to be contrasting but actually similar to details of slides belonging to the other episode in the pair. To quote an example from Dr. Moss's material, for me the *Space* and *Drunk* episodes are not in clear-cut contrast: for instance, astronauts often move just like drunken people.

I would suggest using stimuli to which the majority of people reacts with one clear-cut emotion. In order to select only target episodes which do not arouse complex and ambiguous feelings, some empirical evaluation of their emotional impact ought to be carried out on a population similar to that of the subjects taking part in the psi experiments. To stay with the above example, slides depicting the topic *Drunk* might not bring forth the expected reaction of amusement. For some potential subjects alcohol may present a major emotional problem, either directly or because of its effects on people to whom they are attached.

Therefore, I think that in order to make the evaluation of qualitative data more meaningful one would have to further improve on both target selection and word list procedure. I understand that this is not easy. Perhaps one should not limit oneself only to counting the items

pertaining to the target episode, but should also take into account those statements which are not pertinent to the target ("misses"). Evaluating the "hits" out of such a total score would help to correct for the possibility that some statements refer to the target by chance.

Furthermore, it would be desirable to assign a relative "weight" to the different items composing each list. Obviously, if a subject said "Marilyn Monroe," he would have made a more valid statement than if he had only said "drugs." Some characteristics seem to me essential for the definition of each episode, while other items can be viewed as attributes. These may be used in conjunction with the essentials to arrive at more precise scorings, but their intrinsic defining value is open to questioning.

Finally, my major objection is concerned with the judges' psychoanalytic interpretations of the receivers' verbalizations. I think that the problem is not how well judges agree in their matching procedures. Their unilateral symbolic interpretation will be worthless as long as it will disregard the highly differentiated frame of reference in which the receivers' "association" ought to be evaluated. As you know, interpretations are possible at several levels, and they have to be formulated taking into account the subject's life situation which is made of present elements interwoven with and modulated by past personal history. The example given by Dr. Moss of a sailboat under a bridge being scored as a hit for an episode consisting of nude or partially clothed women is a good illustration of this point. Practically any imagery can have sexual implications; but if we want to give evidence of a telepathic transfer we must have a more articulated, realistic and differentiated correspondence.

KAMIYA: I have a few comments about the design of Dr. Moss's first major experiment, as published in the *Journal of Abnormal Psychology.*

The real issue is not how well the judges, either individually or collectively, could match T and R protocols, but rather whether the subjects as T-R pairs had protocols which co-varied in content over a sequence of six successive episodes. Since group E subjects were well known to each other, they may have been similar in temperament, and perhaps shown similar mood shifts over six successive reports. If this tendency were strong enough, it might have an effect even if the specific stimulus sequence (episodes) tended to counteract it.

As a check on this, Dr. Moss could have required her judges to match protocols from each T in group E with the protocols from a randomly selected R of the same group. She could also have done the

same with R protocols of group A, judged against randomly selected sets of T protocols from group E. The rather small number of hits obtained in group A should be matched by a similarly small number of hits of random T-R pairs in group E.

A p-value ought to be computed for each subject pair, with hits for random pairs being the chance figure, and not the calculated figures used in Table 1. Furthermore, random pairings of all R protocols from groups A and M with episode cards from group E, with and without T protocols, would provide important clues.

Very much needed is also a complete table listing every T-R pair in rows and each judge in columns. This would permit direct comparisons of the different T-R pairs. The judges' hits in group E should be systematically more frequent for some T-R pairs than for others in the same group. If they were not, I would be suspicious, since individual differences in subjects should be expected.

Finally, the higher scores in group M than in group A suggest communication through the walls, from sound equipment or from E_1, or systematically different behavior of E_2 in group A as opposed to group M.

Concerning the scoring procedure used in later experiments, if I understood correctly, the issue is how to discriminate adequately target A from target B using lists of words defining each of the two members which were presented to the receiver as two slides after each trial in the forced choice situation. It seems to me that there is an ideal way to assess the discrimination, that is to compute the chi-square of relative frequencies of A-type responses for A pictures and for B pictures, plus B-type responses for A pictures and for B pictures. That would give a measure of the co-variation between stimulus and response. I wonder whether Dr. Moss has thought of this type of statistical approach.

I am concerned with the physical process by which information is transmitted from Los Angeles to London. There are some fairly straightforward experiments that could be done to check the hypothesis that some unknown physical energy is involved here.

You could have two conflicting sources of energy or information, one group of Ts being shown one episode, while another group is being shown a contrasting episode. If something was being communicated by some form of energy, a sort of jamming should result; you should get essentially random responses from the Rs.

Dr. Moss's experiments are conceived with the implication that the response bias of the recipients is crucial. This could be checked in long distance experiments just by asking a colleague, who does not know that an ESP experiment of any sort is being conducted, to act as a

"control receiver." He would be asked to write down a random list of 15 words that he might pick out of the air. This would provide some kind of baseline and indicate whether the Rs' intentional bias contributes to focussing toward the Ts and their targets. It would have some far-reaching implications if you found out that any R might come up with the same distribution of guesses as your experimental subjects.

MARGENAU: I think we can use physical analogies to rule out the possibility of jamming. You get jamming in radio transmission if the signals are sufficiently strong. Dealing with light impulses, single photons going from source to destination in a random fashion, there can be no jamming. Jamming is confined to a stream of photons or other electromagnetic signals which can be interfered with. Individual photons cannot be jammed. A similar situation might be operative in psi phenomena. Besides, if your analogy is based upon the conviction that psi is manifested through electromagnetic impulses, there are various simple tests by means of which this possibility can be ruled out.

RAO: I wonder whether your scheme would work for clairvoyance and all other situations in which there is a psi transfer of information from an object to a person without any transmitter? Furthermore, we already have an inbuilt control in the T-R design. The experimenter selects a specific message for the T to send and the R's responses must be evaluated according to how well they reflect this message.

KAMIYA: I am concerned with experiments in which there is a transmitter. If a card is the transmitter, we have a difficult problem. But in those cases where there is a person looking at the target, you ought to be able to jam that information channel simply by having two different transmitters sending conflicting messages. I still think this is a fundamental point. If there is no agreement, I fail to see why experiments are done with a sender and a receiver.

BELOFF: Dr. Soal tried this kind of rival transmission working with the subject Gloria Stewart, using two agents simultaneously. He got significant results with one agent and not with the other.[6]

LUDWIG: In yesterday's discussion there were comments on the fact that psi seemed to be a weak force, if any; that it was difficult to predict when it would appear; and that before you could really study it, you had to have a situation where it just stood out like black against white.

In Dr. Moss's experiments highly significant results were obtained

using a particular type of technique. Would this type of methodology satisfy the criteria set forth in this conference to pinpoint certain aspects of psi phenomena? Do we need to go further towards conceptualizing other experiments, or should our attention be directed along these particular lines, in terms of isolating the phenomenon, defining its characteristics and manipulating variables? I would like the group to direct its comments not only towards Dr. Moss's findings but also more generally towards the methodology involved in assembling evidence in this field.

VAN DE CASTLE: I do not think we can agree on the basis of one approach. In psi research the experimenter is a critical variable. Rosenthal[7] has described the profound effect of the experimenter on his subjects, when he possesses certain information, although he cannot directly communicate his expectations to his subjects. I do not think we can mold experimenters to "The Design" because to do so may decrease the individual experimenter's motivation.

LUDWIG: I agree to some extent, yet there is a definite need for replication. Unless results can cross over from one experimenter to another, following certain conditions that are outlined, so that one experimenter can validate what another is doing, there is going to be only a number of individual studies with individual conceptualizations, and very little knowledge will be accumulated.

BELOFF: I entirely agree with Dr. Ludwig. However, before we can get an experimental design which will work for other people, we must isolate the critical variables which gave Dr. Moss her results. We do not yet know them. We have no control comparison to determine whether it was the dramatic nature of the target material that produced these results. The phenomena are still so elusive that we are not in a position to answer Dr. Ludwig when he challenges us to say whether this rather than another would be the ultimate design.

KRIPPNER: We plan to use Dr. Moss's procedure in our laboratory, but will try to select more precise stimuli. For example, in the *Marilyn Monroe* episode, we would use a record of Miss Monroe singing songs from her movies instead of using music from other sources. We would eliminate most slides in which Miss Monroe is disguised as someone else or in which she appears with someone else.

In our dream studies we have evolved toward the utilization of target pictures which are fairly simple in nature but still contain emotion, color, and archetypal symbols. When we use more complex art prints, the evaluation becomes difficult; an imaginative judge can find

correspondences between many dream images and small details of such Chagall paintings as "Purim," or "Soldier Drinking."

LUDWIG: Let me ask a very naive question, Dr. Krippner. Why should Dr. Moss change some variables in her experimental setup when her results came out so highly significant?

KRIPPNER: I am not suggesting that Dr. Moss introduce any changes into her design. I am merely describing the adaptations we would make for our own purposes, which involve somewhat different methods of evaluation. It may be that for daytime experiments, such as those done by Dr. Moss, the stimulus material must be far richer and more complex than needed for night-time experiments, in which the subject provides his own variations and associations.

LUDWIG: Any other comments? Dr. Tart.

TART: There are two aspects to psi communication. One is the mechanism of information transfer per se, which is completely unknown. The other is selectivity; not just any information, but the correct information must be transmitted and received. In Dr. Moss's experiments after the Rs tried to get an impression at a particular time, they were shown two slides and asked to choose one of them. Often many Rs found that their imagery matched one of the slides rather accurately, only to realize that it was the "wrong" one. I suspect that the expectation of seeing these two slides may turn them into ESP targets, in preference to the mental machinations of an unknown group of people (the transmitters) in some far away place. Furthermore, having to make a forced choice so soon after the attempt to "receive" a telepathic message might be too diverting. I do not have an easy solution for it.

MOSS: Neither do I.

TART: There is another shortcoming in Dr. Moss's design. The Maimonides experiments have often shown carry-over effects from one target to another. So have subliminal stimulation experiments with complex material. In Dr. Moss's experiments several targets are used in relatively rapid succession. This may make targets less discriminable. Maybe one would increase the significance level and understandability of results if one used one target per day or even per week. I have a bias myself to collect a large amount of data in the least possible time; that is what our training conditions us to do.

MOSS: That is my bias, too.

LUDWIG: I am a little confused. Why do you suggest to Dr. Moss ways to get higher significance levels? Some of the statistical results she mentioned are extremely impressive.

TART: I am not so much interested in raising significance levels as in clarifying possible underlying mechanisms by certain changes in design. If material from target 1 would still appear in target 5, the responses might be very difficult to evaluate correctly.

MARGENAU: I have no particular criticism of the results that have been achieved. However, I wonder whether similar experiments could not be performed with less complex and less dynamic targets, as Dr. Strauch pointed out. Some semantic ambiguities could thus be eliminated.

MOSS: We could use Zener cards, or similar targets. However, the whole point with my experiments is to evoke emotions and thus elicit qualitative material.

SERVADIO: I am very intrigued by the fact that there is no difference between the performance of related and non-related couples. All the literature on so-called spontaneous phenomena is very much in support of psi occuring between emotionally linked people. If this is not so, we would have to revise our basic thoughts in this area.

RAO: This could possibly be an artifact. When telepathic experiences take place between two relatives, they are likely to be verified, whereas if you had psi occurrences between strangers at a distance, they would be more difficult to observe and to verify.

ULLMAN: As far as relatedness is concerned, you may have lived with your wife for a quarter of a century and never experienced any ESP. Whatever facilitates psi may not encompass sympathy and contact as such, but impinge on more complex personality variables. As a matter of fact, in a number of instances the individuals involved in psi communication did not know each other.

MOSS: Professional mediums can work with people they have never seen before and produce very accurate material, so relationship sometimes has nothing to do with it.

MAUPIN: Did you assess the closeness between the two people on any basis other than the fact that they were married or lovers? Did you make some evaluation of their degree of closeness?

MOSS: No. I simply divided subjects into groups on the basis of

their relationship. I did not use psychological testing to make sure that mother and son, for example, were in the middle of an Oedipal complex. I assumed that if the relationship had existed for a number of years, something was going on between the two people. The results showing no significant difference between related and non-related groups do puzzle me.

VAN DE CASTLE: Maybe the "strangers" liked one another, were attracted to one another, although they did not know each other. When I served as a subject in the Maimonides Laboratory, I did not know anyone there, but I would pick as an agent somebody with whom I thought I could work. One might work with someone one knows well but does not really care for and obtain only disappointing results, whereas it is possible that very good results could be obtained with someone who is a stranger whom one would like to know much better.

MOSS: This might be a good point, but we did not pursue it further.

LUDWIG: Perhaps Dr. Rechtschaffen would like to comment on Dr. Moss's experiments.

RECHTSCHAFFEN: I have only a few technical questions. In long distance experiments how many people knew what the slides were? Were the Rs told which slide was the correct choice?

MOSS: No. And no one knew what the slides were except the projectionist, but he did not know which was the correct choice.

RECHTSCHAFFEN: Apart from the really specific hits, is there significance in the results? Is the psi force distributed along a continuum of "bits" of information?

MOSS: I think there is a continuum. Often bits of information are given in symbolic form. Psychoanalysts call that primary process mechanism. This kind of scattered, dreamlike symbolic representation is much more frequent than the actual precise hit.

FLEW: I am still confused about the boats and the girls. Now, the *Ladies* episode was the target, wasn't it? And the response was the boat rubbing against the bridge. This was rated by the judges as a hit. If a boat rubbing against a bridge is a hit for a target that shows girls, it seems to me that almost anything could be a hit, granted sufficiently able judges.

RECHTSCHAFFEN: What about the unsuccessful experiments? Did they give you any clues?

MOSS: No. Except that they were both group experiments. Perhaps so many people gathered together and trying to receive individually might interfere with each other.

BELOFF: There was another shortcoming in the design of the group experiments. Not only were there the usual response biases due to communal background, but the people were sitting rather squashed in the same room, and one could not preclude the occurrence of subliminal communication among the group members. Yet, as I understand it, in the assessment of the results, you counted each R's responses as if they were independent. I am not a statistician; I am just a psychologist, but to me this seems to be an illegitimate assessment.

MOSS: I understand your point. But still, when you score your students' examinations, do you not assume that their work is independent, or do you assume that there is some type of communication going on among them?

BUCHSBAUM: I have several questions. When the people were free associating after the slides were projected, were they questioned by an investigator, and if so, did the investigator know which series was the correct series?

MOSS: No.

BUCHSBAUM: When the slides were projected in the forced choice situation, was the correct slide always projected on one side or the other?

MOSS: No. The slides were always randomly presented.

BUCHSBAUM: When did you stop running a series of subjects? In other words, did you just do a certain random number of trials and stop when the results were at some particular point, or was there a predetermined series?

MOSS: It depended on the experiment. For some experiments (3, 5 and 6) the sequence was decided beforehand. For the others it was not.

MARGENAU: I speak now as a physicist whose experience is in fields far from parapsychology. One thing is certain in your research; you do not know as yet what the mechanism of transmission is. We do know, however, that a most intricate physiological mechanism, the human brain, is involved. I gather that in the perceptual processes a very small

group of molecules is responsible for enhancement of a stimulus. An indeterminacy principle dominates physical microcosmic processes. If psi phenomena are carried by microscopic, microcosmic physiological processes, you would encounter situations in which the uncertainty principle does have dominance. In other words, you would be in the same position as quantum mechanics for the last 30 years.

The state of any physical system cannot be set up so that the probability distribution of all results will collapse to certainty (which we call delta functions). There seems to be no way to collapse dispersive responses (like those of the brain) to a single answer with certainty. Therefore, I think that at the present stage of your knowledge, a search for complete replicability may be unwarranted. However, the dispersion within given distributions, obtained under certain experimental circumstances, ought to be the same for different observers. What would be significant would be the standard deviation, or some similar aspect of the probability distribution of the results. Once you achieve replicable standard deviations (or other replicable characteristics of total distributions) you might, under certain conditions, try to achieve complete replicability. In sum, I would not want to insist on replicability, in the sense of classical science, as a necessary attribute of convincing parapsychological measurements.

LUDWIG: I agree with Dr. Margenau. But I would like to see even partial replicability rather than just several isolated experiments. Once you have the significance established, you could begin narrowing down, isolating variables, altering conditions and trying to specify even more. You first must have the phenomenon before you can begin defining it. Until then, you can do nothing to establish the criteria under which it emerges.

If one experimenter does an experiment and gets significant results, and another experimenter does that same experiment and does not get significant results, I think something is grossly wrong. The whole idea of the scientific method is that once you specify conditions, a phenomenon can be produced by any number of people who follow the same procedure. As long as only one experimenter can do it, while it cannot be done by anyone else, why bother doing the experiment?

TART: I want to disagree with Dr. Ludwig. I do not think the problem is that impossible. All you have to do is to treat the experimenter as an experimental variable, for example, by using a number of experimenters with different characteristics.

LUDWIG: I agree with that, Dr. Tart. Recently we conducted an experiment in which we studied the effects of the therapist on therapy.[8] Still, unless a number of experimenters in different labs use the same design, or your design includes a number of different experimenters, and the results hold true, you have nothing.

There is always a need for a conceptualization of some sort prior to any experimentation. Whether the conceptualization proves to be right or wrong, it will have suggested an array of experiments to be done.

REFERENCES

1. MOSS, T. and GENGERELLI, J. A.: "Telepathy and Emotional Stimuli: a Controlled Experiment," *J. Abn. Psychol., 72* (1967), 341.
2. MOSS, T. and GENGERELLI, J. A.: "ESP Effects Generated by Affective States," *J. Parapsychol., 32* (1968), 90.
3. KRIPPNER, S.: "ESP and Creativity," Unpublished manuscript.
4. MOSS, T.: "ESP Effects in 'Artists' Contrasted with 'Non-Artists,'" *J. Parapsychol., 33* (1969), 57.
5. MOSS, T.: "Long Distance ESP: A Controlled Experiment," Paper presented at the 1968 Convention of the Parapsychological Association.
6. SOAL, S. G. and BATEMAN, F.: *Modern Experiments in Telepathy* (Faber, London, 1954), Ch. 13.
7. ROSENTHAL, R.: *Experimenter Effects in Behavioral Research* (Appleton-Century-Crofts, New York, 1966).
8. LUDWIG, A. M.: "Studies on Alcoholism and LSD: Influence of Therapist Attitudes on Treatment Outcome." *Amer. J. Orthopsych., 38* (1968), 733.