

# Methodological Problems in Free-Response ESP Experiments

J. E. KENNEDY<sup>1</sup>

(Original publication and copyright: *Journal of the American Society for Psychical Research*, 1979, Volume 73, pp 1-15.)

**ABSTRACT:** Various methodological problems have occurred in some recent free-response experiments in parapsychology. Statistical errors have involved improper assumptions of independence, multiple analysis of data, and, to a lesser extent, data selection and misapplication of the normal approximation to the binomial. The inability to accurately measure information in free-response judging procedures has led to a general lack of statistical sensitivity and to incorrect inferences about ESP "information rates." Further, the possibility of sensory cues has not been completely eliminated in several experiments. Most of these problems can be easily avoided with slight alterations in experimental design.

## INTRODUCTION

Free-response methods have become very popular in parapsychological research during the last decade. As with any fairly new experimental method, various types of errors in experimental procedure or analysis of data have occurred. This paper discusses several methodological errors that have appeared in published reports. Most of the faulty methods arise primarily in free-response experiments, but a few are well-known problems that also arise in most other applications of statistics. Numerous examples are drawn from the literature not only to show the existence of the problems, but also to indicate results that must be interpreted with these methodological questions in mind.

## IMPROPER ASSUMPTIONS OF INDEPENDENCE

In several free-response experiments a matching type of judging procedure has been improperly analyzed. Typically, after free-

<sup>1</sup> I wish to thank Charles Akers, Rex Stanford, and Debra Weiner for valuable comments that led to extensive improvements of earlier drafts of this paper.

response protocols (transcripts of verbal reports and perhaps drawings) had been collected for several different targets, a judge (the subject or an independent judge) was given *all* the target pictures used for a set of trials and *all* the response protocols for that same set. The order of the targets was randomized and the judge was asked to match the responses to the targets according to similarity of content. In a "forced-matching" procedure (as discussed in Burdick and Kelly, 1977) the judge would match a response with one and only one target and the pair would be removed from the judging pool. It is easy to see that under these conditions the probability of obtaining a correct match for one target would not be independent of the probability of a correct match for the other targets in the set. This dependency must be considered in the statistical analysis. Scott (1972) has published a table for evaluating significance under these conditions and Burdick and Kelly (1977) also discussed appropriate methods.

In most experiments, however, the judges have been given all targets and all responses and were told to rank all responses against each target and to treat each target independently of the others in the set; thus, a response could be ranked equally high with more than one target. The data have then been analyzed by statistics that assume that the ranking for each target is independent of the rankings given to the other targets in the pool. The significance level has been found by either evaluating the number of direct hits (correct ranks of one) using the binomial distribution or by employing a preferential ranking method which uses equations provided by Stuart (1942) and Morris (1972) to evaluate the sum of the ranks given to the correct responses. (The preferential-ranking method is often considered to be more sensitive since it utilizes near hits.)

However, as both Stuart and Morris clearly stated and has been discussed by others (Burdick and Kelly, 1977; Scott, 1972), the response rankings given for the different targets under these conditions cannot be considered independent. Morris (1972) noted:

In the particular case in which the same judge does all the rankings of a closed pool of  $N$  targets and protocols, his rankings may or may not be independent, regardless of any set of instructions he may receive. As Stuart (1942) pointed out, the main danger occurs in the tendency of a judge to avoid assigning any target a ranking of one for more than one protocol (pp. 402-403).

In discussing the use of his equations and a table he published, Morris (1972) commented:

This table is based on the assumption that the rankings are assigned independently. Table 1 should be used only as a rough estimate in the case of studies involving a constant judge responding to a

constant target-protocol pool, especially if (a)  $N$  is small, six or less; or (b) the judge has not in fact assigned any target a ranking of one on more than one occasion, and if at the same time more than one-third of the rankings of number one are correct and therefore contributing to the small size of  $s$  [sum of ranks]. Procedures using a constant judge and target pool are not recommended unless an appropriate statistical tool is to be used (p. 406).

Neither the binomial nor the preferential ranking method is appropriate under these conditions. Further, it is apparent that the preferential ranking method can be the most misleading under the very conditions needed to get highly significant results (i.e., many correct ranks of one).

One of the best strategies to overcome this possible dependence is to change the judging process. For each response protocol, a number of dummy targets drawn from the overall target pool can be submitted with the real (randomly selected) target; the judge ranks the possible targets according to correspondence with the response. This method is perfectly valid and various forms of it have become the most common procedures for evaluating free-response experiments.

Another method that appears valid at first glance is to use a number of judges (equal to the number of targets), with each judge ranking all the responses in the set against only one target. This procedure would seem to eliminate dependencies since each judge sees only one target. However, some dependency would still be possible under these conditions.<sup>2</sup> For example, if the judges know each other, they may consciously or unconsciously be influenced by their knowledge of the preferences of the other judges. This situation could lead to problems if there were no ESP in the data and, as a result, the judges based their ranks primarily on personal preferences for the responses. Since various other types of subtle artifacts can be imagined, it seems best to avoid this procedure.

If the data were not or cannot be judged with a procedure assuring independence, several statistical methods can be applied. In situations in which a fairly large number of target-response pools are judged, the mean of the sum of the ranks in each pool can be tested against MCE ( $N/[N + 1]/2$ , where  $N$  is the number of target-response pairs in each pool) with a single mean  $t$ -test. The difference between groups can be compared with a  $t$ -test or analysis of variance. This approach is particularly desirable for research that is process-oriented rather than merely looking for evidence of ESP (Stanford and Palmer, 1972).

Greenwood (see Stuart, 1942) developed a preferential ranking

<sup>2</sup> Charles Akers pointed out the problems with this procedure.

method to correct for first order dependence which assumes dependence only for the ranks of one. I concur with Burdick and Kelly (1977, p. 114) that it is best to avoid this method since it requires "potentially controversial" assumptions about the judges' behavior. Burdick and Kelly (1977) also pointed out that Greville's (1944) method corrects for all existing dependencies in the single judge, closed target-response pool situation (as well as essentially any other situation) and described the required calculations in detail. This method has the distinct advantage that the rankings from several judges can be incorporated into the analysis. The only cautionary note is that small sample sizes can, under certain conditions, lead to inflated results with Greville's method (Burdick and Kelly, 1977; Scott, 1972). Unfortunately, little work has been done to explore the extent of this potential problem. As noted by Burdick and Kelly (1977, p. 116), computer modeling to find exact probabilities can be applied in many cases with small  $N$ .

For an analysis comparable to the binomial method applied to direct hits, one can assume the worst case dependency and evaluate the direct hits with the forced-matching analysis mentioned above. This test would be most powerful if the judge had actually followed the forced-matching procedure, and it may be *quite* conservative for cases in which the judge was asked to treat each target independently. For most of the published reports that have the dependence problem, the forced-matching method is the only way to estimate the significance from published data.

Three reports of hypnotic dream studies (Honorton, 1972; Honorton and Stump, 1969; Parker and Beloff, 1970) have used binomial or preferential ranking methods when independence of ranks cannot be assumed. In all three experiments each session consisted of four clairvoyant, hypnotic dream trials following which the subject for the session ranked the four protocols according to correspondence with each of the four targets. In the earliest experiment (Honorton and Stump, 1969), seven sessions were completed by six subjects. The results gave 13 direct hits ( $MCE = 7$ ) out of 28 trials, and this was evaluated with a normal approximation to the binomial of  $CR = 2.40$ ,  $p < .02$ , two-tailed. The sum of ranks gave  $p < .02$ , two-tailed. Applying the conservative forced-matching analysis, 13 direct hits in 28 trials gives  $p < .054$ , two-tailed; thus, given the published data, the actual significance level probably lies between the questionable .02 and the correct but possibly conservative .054.<sup>3</sup>

<sup>3</sup> It should be clearly understood that the results obtained when studies are reanalyzed using the forced-matching analysis are *not* meant to be definitive. Rather, the forced-matching analysis is presented to indicate the maximum range of error that *may* be present in the published reports and, more importantly, to indicate

The above experiment was replicated and extended by Honorton (1972). In this case the condition equivalent to the previous study gave 15 direct hits in 40 trials ( $MCE = 10$ ) and the significance was reported as  $CR = 1.83, p < .04$ , one-tailed. However, this  $CR$  is slightly inflated since it, unlike the  $CR$  of the previous study, was calculated without using a continuity correction. The exact binomial gives  $p = .054$ , one-tailed, and the matching case gives  $p = .085$ , one-tailed.

Another attempt to replicate the Honorton-Stump study was reported by Parker and Beloff (1970). Eight subjects each participated in two sessions. The first session alone gave 13 direct hits out of 32 trials which were reported as  $CR = 1.84, p = .06$ , two-tailed (the authors chose the two-tailed analysis). The matching case gives  $/? = .128$ , two-tailed. The significance of the sum of the ranks for the first session was reported as  $p = .02$ , two-tailed. The individual data were reported so a more appropriate  $t$ -test could be carried out, yielding  $t = 2.707, 7 df; p < .05$ , two-tailed. Parker and Beloff also reported a significant decline ( $p = .011$ ) between sessions using a chi-square analysis, but it is not apparent to me how that analysis was done. A  $t$ -test on the difference between sessions finds only  $p < .1$ , two-tailed.

A precognitive dream study reported by Krippner, Ullman, and Honorton (1971) used binomial methods to evaluate direct hits when a closed pool of eight target-response pairs was judged. The authors reported  $CR = 3.74, p = .00018$ , two-tailed; however, the normal approximation ( $CR$ ) cannot be used reliably with a  $P$  of  $1/8$  and only eight trials. The exact binomial gives  $p = .002$  and the matching case yields  $p = .008$ .<sup>4</sup> At least one other dream telepathy study made questionable use of the binomial method (Ullman and Krippner, 1970, pp. 100-101), but the results are also clearly significant with the conservative analysis.

Several experiments using the remote viewing procedure have incorrectly used the preferential ranking method (Bisaha and Dunne, 1977; Dunne and Bisaha, 1978; Puthoff and Targ, 1976; Targ and Puthoff, 1976). In these cases each experiment consisted of the judging of one target-response pool. Five of the six experi-

---

which studies are clearly significant even with the conservative analysis. A more sensitive and appropriate statistical analysis would likely find a significance level lying between the binomial and the forced-matching values. Unless otherwise noted, all probabilities for the forced-matching analyses were obtained using the table in Scott (1972).

<sup>4</sup> This and other Maimonides dream telepathy experiments were also analyzed using analysis of variance procedures (see Ullman and Krippner, 1970; Ullman and Krippner, with Vaughan, 1973). Exactly how the ANOVAs were applied and what assumptions of independence were made are not clear to me.

ments carried out by Puthoff and Targ (reported in Puthoff and Targ, 1976) were significant at the .05 level with the preferential ranking method, but only two are significant with the conservative forced-matching analysis. Of the two remote viewing studies by Bisaha and Dunne, the first (Bisaha and Dunne, 1977) is quite significant even with the conservative analysis (based on unpublished data in the short report handed out at the 1976 convention of the Parapsychological Association). The second remote viewing study (Dunne and Bisaha, 1978) and a ganzfeld study (Dunne, Warnick, and Bisaha, 1977) are not significant with the conservative analysis.

A few reports of remote viewing studies have also improperly combined results of multiple calling or multiple judging of data (Hastings and Hurt, 1976; Puthoff and Targ, 1975). In such situations the responses of the different subjects or judges are not independent and must be treated accordingly (see Burdick and Kelly, 1977).

#### DIFFICULTIES FROM MULTIPLE ANALYSIS

Multiple analysis of data arises because there are many different procedures for statistically evaluating free-response experiments. For example, several judges (e.g., the subject and one or more independent judges) can judge the data, the binomial method can be used to find the significance of the number of hits ( $P = 1/2$ ) or direct hits ( $P = 1/N$ ), the preferential ranking method can be applied, and various forms of rating methods (e.g., the judge rates the correspondences between target-response pairs on a scale from one to a hundred; see Burdick and Kelly, 1977) can be used. With so many options, it is not surprising that experiments are often analyzed in several different ways.

Evaluating experiments in which many statistical analyses have been performed is a problem that occurs in most statistical experimental work, and it certainly is neither new nor unique to parapsychology. Multiple analyses, however, are not necessarily misleading and, in fact, may be desirable in free-response studies since they can provide confidence in the reliability of the scoring procedures. Thus, a dream telepathy experiment (the second study reported in Ullman, Krippner, and Feldstein, 1966) that apparently included eight different methods for analyzing the overall ESP effects in the data is not difficult to interpret since six of the eight analyses gave results significant at the .05 level.<sup>5</sup>

<sup>5</sup> The first study reported 12 analyses looking for an overall ESP effect. The eight basic analyses involved the ranking and rating by the subjects and the independent

The results of some other experiments, unfortunately, are not so consistent and indicate that differences among various scoring procedures need to be investigated. In the first study reported in the paper by Ullman, Krippner, and Feldstein (1966), only one of 12 analyses was significant at the .05 level. (Confidence in these results is increased by the fact that apparently another judge later produced an effect at the .01 level in one of two or three evaluation methods; Ullman and Krippner, 1970, pp. 69-70).<sup>6</sup> Three hypnotic dream studies have found conspicuously different results with different judges (Honorton and Stump, 1969; Keeling, 1971; Krippner, 1968). An experiment reported by Braud and Wood (1977) using the ganzfeld procedure presents another instance of inconsistent results with different methods of analyzing the data. Targets from the Maimonides binary target pool (see Honorton, 1975a) were scored in the Braud and Wood study according to the binary scoring method and also by the binomial method applied to the subjects' rankings ( $P = 1/2$ ). Although there were significant effects, the results with the two scoring methods were not significantly correlated (p. 421) and for several conditions were not even in the same direction (p. 417).

While various post-hoc hypotheses can be proposed to explain why a certain analysis worked in a particular situation, until the process of judging free-response data and the effectiveness of various statistical procedures are better understood, studies yielding inconsistent results must be interpreted with the multiple analyses problem in mind. Evaluation of the efficacy of different methods of analysis is hampered by a lack of discussion of the number of analyses attempted and of the reasons for applying particular techniques. For example, the first ganzfeld study from Maimonides (Honorton and Harper, 1974) used a direct hits ( $P = 1/4$ ) analysis. The second ganzfeld study (Terry and Honorton, 1976) used a binomial  $P = 1/2$  analysis which gave  $p = .00296$ ; a footnote mentioned that the direct hits analysis gave  $p = .05$ . (In fact, the exact binomial probability is .053.) The third ganzfeld study (Terry

---

judges for dream protocols and dream protocols plus associations (p. 421). These analyses were evaluated by carrying out eight separate analyses of variance (p. 425). Also, the four analyses involving ranks (i.e., subjects and independent judges for dream protocols and dream protocols plus associations) were evaluated by the binomial method (p. 425). For the description of the statistical analyses in the second study the authors state only that "statistical confirmation was expected in the form already described for Experimental Study I" (p. 429). However, the results for the eight analyses of variance were reported without mentioning results from the four binomial analyses.

<sup>6</sup> Of the total of 14 or 15 analyses, two involved the dubious application of the binomial method to a closed target-response pool situation. Both results were nonsignificant, however.

and Honorton, 1976) again used the direct hits analysis. When and why the  $P = 1/2$  analysis was selected for the second study is not discussed. Was this just an arbitrary decision of the experimenters, or are there known situations for which the  $P = 1/2$  method is better?

The primary danger of multiple analysis occurs when one attempts to appraise an entire line of research. For example, Honorton (1977) recently evaluated several lines of research related to psi and internal states (meditation, p. 442; relaxation, p. 457; ganzfeld, p. 464) by combining the probability values of the studies using Fisher's method and by comparing the number of studies that were significant at the .05 level with the number expected by chance. *These statistical procedures, however, assume that only one analysis was carried out for each experiment, a situation not true for most of the experiments.* Further, it is not clear how the significance value was selected for experiments that reported several analyses. The figures for the overall significance and reliability of the various lines of research are certainly somewhat inflated and perhaps extremely so.

One solution to the dilemma of multiple analyses would be to select before an experiment is carried out one and only one specific statistical analysis to be used to represent that experiment when evaluating the line of research. Other analyses would, of course, be carried out and would have much value, but they would not enter into any statistical evaluation of the line of research as a whole. *Carrying out many exploratory analyses in a pilot experiment is quite proper, but selecting the most significant result for use when evaluating the line of research is not.* Also, clear distinctions between planned and post-hoc analyses are needed. This discussion clearly applies to all experiments in parapsychology, not just to free-response studies.

#### INABILITY TO PRECISELY MEASURE INFORMATION IN FREE- RESPONSE TARGETS

While free-response procedures may be more interesting for subjects and the target information more life-like than in forced-choice tests, it is very difficult to quantify the amount of information transferred. Honorton (1975a) has commented that the typical free-response judging procedures are

. . . grossly insensitive to genuine psi effects since a perfect correspondence between target and subject report may receive the same weight as a partial correspondence. Such procedures are time-consuming and cumbersome and, in addition, provide no quantita-

live data concerning (a) the information content of the target, (b) the information content of the subject's report, or (c) the information content which is shared by the target and subject report (pp. 353-354).

In the usual judging methods, a large amount of information is lost as the data are reduced into forced-choice trials. For example, if a judge's ranks are analyzed with the binomial ( $P = 1/2$ ) method, the entire information content of the target and response is reduced to one bit, losing essentially all the information that may have been involved in the psi process. As a step in overcoming such problems, Honorton (1975a) developed a pool of target pictures, each of which could be evaluated according to the presence or absence of 10 specific attributes. This is an improvement, but the measurable information is still limited to 10 bits and, as Burdick and Kelly (1977) noted, this procedure "only grossly characterizes the content of any given target" (p. 114).

This inability to precisely measure information with free-response procedures has led to some misleading comparisons of free-response and forced-choice methods. Honorton (1975b, 1977, pp. 450-451) has made such a comparison by using a psi quotient percent (PQ%) measure. The PQ (Schmidt, 1970) for each experiment was divided by the maximum PQ that could have been obtained, i.e., the PQ if all trials had been hits.<sup>7</sup> The PQ measure can be related to information transmitted per trial (Schmidt, 1970) and Honorton took the PQ% as a measure of "information rate." He compared the mean PQ% for free-response studies with that for forced-choice studies and found the PQ% of the free-response studies to be significantly larger, indicating that the "information rate" was much higher in free-response studies.

This comparison, however, is both conceptually and statistically erroneous. The PQ% measures only the information that is utilized by the particular statistical analysis employed in the experiment. As discussed above, the common methods of analyzing free-response data deal with only a small fraction of the total information content of the targets and responses, so the PQ% is only remotely related to the actual information rate that may be involved in the psi process. The comparison with forced-choice studies, which allow relatively precise quantification of information, is misleading.

From a mathematical point of view, the mean chance expected value for PQ% (and PQ) is inversely proportional to the number of

<sup>7</sup> The PQ is a measure of efficiency, defined by Schmidt as  $PQ = 1000 CRW$ , where  $N$  is the number of trials and CR is the CR obtained on the  $N$  trials.

trials.<sup>8</sup> Since forced-choice experiments have many more trials than free-response experiments, the corresponding PQ%'s have very different *a priori* expected values, and a direct comparison is not meaningful. Also, large values of PQ% can come from experiments that have no evidence for paranormal communication if the number of trials is small.<sup>9</sup> The fact that the mean PQ% for free-response studies is one to two orders of magnitude larger than that of forced-choice experiments does not indicate higher information rates and may merely reflect the fact that they have one to two orders of magnitude *fewer* trials than forced-choice experiments. Reducing the number of trials in the forced-choice experiments by carrying out majority-vote procedures would lead to dramatic increases in their PQ%'s.<sup>10</sup>

### IMPROPER USE OF CR

The small numbers of trials found in many free-response experiments require careful application of the normal approximation to the binomial distribution. The continuity correction should be used and the approximation becomes more tenuous as  $P$  differs from  $1/2$ . Instances of inflated CR values when the sample size was too small for  $P = 1/8$  and/or when the continuity correction was not used were mentioned above. In a more recent example (Braud, 1977), two hits in 10 binary trials were reported as suggestive with  $CR = 1.90$ ,  $p = .057$ , two-tailed. However, the continuity correction was not used. The exact binomial gives  $p > .1$ , two-tailed.

<sup>8</sup> Since  $CR^2$  is approximately chi square distributed with one degree of freedom, the expected value is one. Therefore, the expected value for PQ is  $E[PQ] = 1000/W$ . The maximum PQ is equal to 1000 ( $QIP$ ) independent of the number of trials, since  $PQ_{max} = \frac{CR^2_{max} \cdot N}{N \cdot NPQ} = \frac{1000 (N-NP)^2}{N \cdot NPQ} = \frac{1000 (1-P)^2}{P}$

Thus  $E[PQ\%] = P/QN$  and for a given  $P$ , PQ% differs from PQ only by a constant factor.

<sup>9</sup> Douglas Stokes pointed out to me that the observed PQ% depends only on the scoring rate and is independent of the number of trials. If  $X$  is the scoring rate, then, From footnote 8,  $PQ\% = \frac{P}{Q} \frac{(XN-NP)^2}{N^2 PQ} = \frac{(X-P)^2}{Q^2}$ . Thus, for example,

an experiment with  $P = 1/2$  and .2 hits in 20 trials ( $CR = n.s.$ ) has the same PQ% as an experiment with 600 hits in 1000 trials ( $CR = 6.29$ ). Also, an experiment with  $P = 1/2$  and only one trial will always have  $PQ\% = 100$ , i.e., 100% "information rate."

<sup>10</sup> Carrying out a majority vote procedure on data leads to a higher scoring rate, a smaller number of trials, and a slightly smaller CR (Scott, 1960). That this procedure will result in higher PQ and PQ% is easily seen from footnotes 8 and 9. The comparison of majority vote data reduction with free-response judging procedures is more than a superficial analogy.

### DATA SELECTION

Instances of possible data selection seldom appear in current parapsychological work, but they do occasionally occur under certain circumstances. When several subjects are tested in multiple sessions, the experimenter may, for various reasons, choose to discard the data for subjects who did not complete the intended number of sessions. This selection is not intrinsically improper, but the results for the discarded data should be reported. In ESP tests, those subjects who do poorly on the first sessions may become discouraged and drop out, while those initially doing well will finish the required number of sessions. This situation would create a biased sample. A confirmatory ganzfeld experiment (Terry and Honorton, 1976, p. 211) is a recent example in which results for the discarded data were not reported.

### SENSORY CUES

In many free-response telepathy experiments, the subject judged his own responses using a target pool which included the same targets that were viewed and handled by the agent. Leaving aside questions of possible collusion, this procedure does not *completely* eliminate the possibility of sensory cues. One wonders, for example, whether in several "internal-states" experiments there might have been slight unintentional cues as to which of the possible targets had been previously handled by the agent, and, if so, whether the supposed "psi-conducive" states might have increased the subjects' sensitivity to such subliminal sensory cues. The use of a duplicate set of targets for the subject would eliminate the need to consider such possibilities. Unfortunately, the two laboratories that have carried out most of the successful free-response, internal-states experiments have reported several experiments in which the same target material was used by both subject and agent.

Stokes (1978) has recently noted other mechanisms by which the judging procedure may have allowed sensory cues in some free-response experiments. In a remote viewing study (Bisaha and Dunne, 1977), judges ranked a target-response pool consisting of photographs of the target locations and the corresponding response transcripts. Since the photographs were taken by the agent on the day of the ESP trial, and the trials occurred on different days, both the photographs and the response transcripts might have contained cues about the weather on the day of the trial. This speculation is made more plausible by the experimenters' comment that "since most of the trials were performed in early spring, a week or two could make a significant difference in the appearance of the loca-

tion due to climate conditions and developing foliage" (p. 85). This criticism also applies to another remote viewing experiment carried out by the same experimenters (Dunne and Bisaha, 1978).

Apparently referring to a dream study reported by Child, Kanthamani, and Sweeney (1977), Stokes (1978) commented:

. . . one of the judges acted as agent and was later asked to rank each target against the subject's dreams on successive nights. As a person's dream may incorporate certain "day residues," this judge might have been able to infer unconsciously which dream took place on which night (he, being the agent, of course knew the target for each night) (p. 73).

As a general guideline, the judges should not know or be able to get possible cues from the target materials as to the order or timing of the use of particular targets.

#### CONCLUSION

The only conclusion to be drawn from this paper is the obvious one that time needs to be devoted to critical discussions of methodology. Not only do we need to avoid repeating errors in future research, but scientific advancement requires that we accurately assess the strengths and limitations of all our published work. Since free-response procedures may offer an important method for investigating certain aspects of the ESP process, it is fortunate that designing experiments which avoid the pitfalls discussed above is a fairly simple matter.

#### REFERENCES

- BISAHA, J. P., AND DUNNE, B. J. Precognitive remote viewing in the Chicago area: A replication of the Stanford experiment. In J. D. Morris, W. G. Roll, and R. L. Morris (Eds.), *Research in Parapsychology 1976*. Metuchen, N. J.: Scarecrow Press, 1977. Pp. 84-86.
- BRAUD, W. G. Long-distance dream and presleep telepathy. In J. D. Morris, W. G. Roll, and R. L. Morris (Eds.), *Research in Parapsychology 1976*. Metuchen, N. J.: Scarecrow Press, 1977. Pp. 154-155.
- BRAUD, W. G., AND WOOD, R. The influence of immediate feedback on free-response -GESP performance during ganzfeld stimulation. *Journal of the American Society for Psychical Research*, 1977, 71, 409-427.
- BURDICK, D. S., AND KELLY, E. F. Statistical methods in parapsychological research. In B. B. Wolman (Ed.), *Handbook of Para-*

- psychology. New York: Van Nostrand Reinhold, 1977. Pp. SI-130.
- CHILD, I. L., KANTHAMANI, H., AND SWEENEY, V. M. A simplified experiment in dream telepathy. In J. D. Morris, W. G. Roll, and R. L. Morris (Eds.), *Research in Parapsychology 1976*. Metuchen, N. J.: Scarecrow Press, 1977. Pp. 91-93.
- DUNNE, B. J., AND BISAHA, J. P. Multiple channels in precognitive remote viewing. In W. G. Roll (Ed.), *Research in Parapsychology 1977*. Metuchen, N. J.: Scarecrow Press, 1978. Pp. 146-151.
- DUNNE, B. J., WARNOCK, E., AND BISAHA, J. P. Ganzfeld techniques with independent rating for measuring GESP and precognition. In J. D. Morris, W. G. Roll, and R. L. Morris (Eds.), *Research in Parapsychology 1976*. Metuchen, N. J.: Scarecrow Press, 1977. Pp. 41-43.
- GREVILLE, T. N. E. On multiple matching with one variable deck. *Annals of Mathematical Statistics*, 1944, 15, 432-434.
- HASTINGS, A. C., AND HURT, D. B. A confirmatory remote viewing experiment in a group setting. *Proceedings of the IEEE*, 1976, 64, 1544-1545.
- HONORTON, C. Significant factors in hypnotically-induced clairvoyant dreams. *Journal of the American Society for Psychological Research*, 1972, 66, 86-102.
- HONORTON, C. Objective determination of information rate in psi tasks with pictorial stimuli. *Journal of the American Society for Psychological Research*, 1975, 69, 353-359. (a)
- HONORTON, C. Receiver-optimization and information rate in ESP. Paper presented at the 141st annual meeting of the American Association for the Advancement of Science, New York City, January 27, 1975. (b)
- HONORTON, C. Psi and internal attention states. In B. B. Wolman (Ed.), *Handbook of Parapsychology*. New York: Van Nostrand Reinhold, 1977. Pp. 435-472.
- HONORTON, C., AND HARPER, S. Psi-mediated imagery and ideation in an experimental procedure for regulating perceptual input. *Journal of the American Society for Psychological Research*, 1974, 68, 156-168.
- HONORTON, C., AND STUMP, J. P. A preliminary study of hypnotically-induced clairvoyant dreams. *Journal of the American Society for Psychological Research*, 1969, 63, 175-184.
- KEELING, K. R. Telepathic transmission in hypnotic dreams: An exploratory study. *Proceedings of the Parapsychological Association*, 1971, 8, 55-58.
- KRIPPNER, S. Experimentally-induced telepathic effects in hypnosis and non-hypnosis groups. *Journal of the American Society for Psychological Research*, 1968, 62, 387-398.

- KRIPPNER, S., ULLMAN, M., AND HONORTON, C. A precognitive dream study with a single subject. *Journal of the American Society for Psychological Research*, 1971, 65, 192-203.
- MORRIS, R. L. An exact method for evaluating preferentially matched free-response material. *Journal of the American Society for Psychological Research*, 1972, 66, 401-407.
- PARKER, A., AND BELOFF, J. Hypnotically-induced clairvoyant dreams: A partial replication and attempted confirmation. *Journal of the American Society for Psychological Research*, 1970, 64, 432-442.
- PUTHOFF, H. E., AND TARG, R. Remote viewing of natural targets. In J. D. Morris, W. G. Roll, and R. L. Morris (Eds.), *Research in Parapsychology 1974*. Metuchen, N. J.: Scarecrow Press, 1975. Pp. 30-32.
- PUTHOFF, H. E., AND TARG, R. A perceptual channel for information transfer over kilometer distances: Historical perspectives and recent research. *Proceedings of the IEEE*, 1976, 64, 329-354.
- SCHMIDT, H. The psi quotient (PQ): An efficiency measure for psi tests. *Journal of Parapsychology*, 1970, 34, 210-214.
- SCOTT, C. An appendix to the repeated guessing technique. *International Journal of Parapsychology*, 1960, 2 (No. 3), 37-45.
- SCOTT, C. On the evaluation of verbal material in parapsychology: A discussion of Dr. Pratt's monograph. *Journal of the Society for Psychological Research*, 1972, 46, 79-90.
- STANFORD, R. G., AND PALMER, J. Some statistical considerations concerning process-oriented research in parapsychology. *Journal of the American Society for Psychological Research*, 1972, 66, 166-179.
- STOKES, D. M. Review of *Research in Parapsychology 1976*. *Journal of Parapsychology*, 1978, 42, 70-76.
- STUART, C. E. An ESP test with drawings. *Journal of Parapsychology*, 1942, 6, 20-43.
- TARG, R., AND PUTHOFF, H. E. Replication study on the remote viewing of natural targets. In J. D. Morris, W. G. Roll, and R. L. Morris (Eds.), *Research in Parapsychology 1975*. Metuchen, N.J.: Scarecrow Press, 1976. Pp. 33-37.
- TERRY, J. C., AND HONORTON, C. Psi information retrieval in the ganzfeld: Two confirmatory studies. *Journal of the American Society for Psychological Research*, 1976, 70, 207-217.
- ULLMAN, M., AND KRIPPNER, S. *Dream Studies and Telepathy. An Experimental Approach*. New York: Parapsychology Foundation, 1970.

- ULLMAN, M., KRIPPNER, S., AND FELDSTEIN, S. Experimentally-induced telepathic dreams: Two studies using EEG-REM monitoring techniques. *International Journal of Neuropsychiatry*, 1966, 2, 420-437.
- ULLMAN, M., AND KRIPPNER, S., WITH VAUGHAN, A. *Dream Telepathy*. New York: Macmillan, 1973.

*Institute for Parapsychology*  
*College Station*  
*Durham, North Carolina 27708*

Return to: [Paranormal Phenomena Articles](#)