The Journal *of*
Parapsychology

VOLUME 31 | SEPTEMBER, 1967 | NUMBER 3

DECLINE OF VARIABILITY OF ESP SCORING ACROSS A PERIOD OF EFFORT

By JAMES C. CARPENTER and JOSEPHINE C. CARPENTER

ABSTRACT: A body of data originally collected for other purposes was examined for evidence of ESP by analyzing fluctuations in scoring (variance). Two variance measures were used: (1) the variance within the experimental session; and (2) the variance of the half-run scores within the run. Each of the two experimenters acted as his own subject. There were 20 clairvoyance series, and nine precognition series, and the targets were ESP symbols. Some of the runs were carried out by the subject in an "expansive" response attitude; some in a "compressive" attitude; and some in a natural or unmodified manner.

(1) It was anticipated on the basis of previous work that if each series were cut in half the run scores in the first halves would show more variance than those in the second halves. This was found to occur with significant consistency ($p < .01$). It was also considered likely that, for psychological reasons, the decline would probably be more pronounced in the shorter series (16-36 runs) than in the longer ones (40-90 runs). Such a decline did occur to a significant extent ($p < .001$) and consistently so ($p < .0001$) in the shorter series. In the longer series, it was not significant.

(2) The difference between the variance of the first-half run scores and that of the second-half run scores was not significant. However, when the data were divided into the different types of response attitude, it was found that the unmodified response style yielded a half-run decline in variance which replicated that found in previous research.

The authors suggest that changes in psychological states, such as spontaneity and enthusiasm, may be reflected in the obtained variance declines.—Ed.

INTRODUCTION

THIS STUDY describes research on the problem of predicting and interpreting ESP scoring in terms of variance—a measure of the scatter of any set of numbers around their central tendency, or mean. Two groups of numbers could have identical means, yet still differ greatly in their amount of scatter. Similarly, a body of ESP data could exhibit as a whole no significant deviation in either a positive or negative direction, yet still contain an internal variability of individual scores which is either too small or too large to conform to binomial ("chance") expectation.

It is by now a conventional assumption in parapsychological research that "psi-hitting" and "psi-missing" are alternate modes of expressing ESP. The effort to find experimental conditions which will differentiate between these two directions of scoring has, of late, been the object of most ESP research. To view a body of data in terms of variance is, however, somewhat different from either examining its overall deviation from chance or searching for criteria predictive of scoring direction. Rather, it involves a consideration of the magnitudes of deviations, irrespective of direction, exhibited by the component parts (e.g., run scores) within the body of data. A genuine extrasensory interaction between a person and his target may be discernible by an analysis in terms of variance, even while the same data, treated by previous means, shows no evidence of ESP. Extrachance variance effects have been reported occasionally in the past (3, 6), but they seem to have been regarded as incidental to the main problems at hand and to have aroused little attention.

The present authors first became interested in variance effects when an unpredicted statistically significant decline in run-score variance was found from the first to second part of a series of runs (5). Overall, the variance was significantly larger than the binomial expectation. In this case, the subject had been under the stimulating effects of a small dosage of dextroamphetamine sulfate, and he underwent a decline in experienced stimulation during the testing which corresponded closely to the decline in variance.

Several months later, it occurred to the authors that the decline in variance might be a normal function in ESP testing irrespective of drug stimulation. They had on hand a sizeable body of data which had been collected by them over a six-month period from November, 1965, to April, 1966, and this was suitable for testing the hypothesis.

These previously collected data are the basis of the present report. Two separate sets of analyses have been made on them. One examines the variability of run scores within a series of runs which were generated in a single "sitting" or uninterrupted period of effort. The other set pertains to patterns of scoring variability within the run by a calculation of half-run scores.

Procedure

The data were generated by the experimenters acting as their own subjects. Altogether, there were 29 series of ESP runs (27 by J.C. and two by J.C.C.), each of which was performed at a single sitting. These data comprised all testing done by the subjects during that school year, up until the time it occurred to them to analyze the data for run-score variance, as described below.

The number of runs to be performed at each sitting was decided before beginning the sitting and the number was noted on the initial page. The series ranged from 16 to 90 runs. Twenty of the series were done with the DT clairvoyance procedure, in which a well-shuffled deck of ESP cards was placed out of sight either in a drawer or behind a stack of books while the calls were made. Each run was checked immediately. The other nine series were of a precognitive type, in which all calls of the series were made before the target orders were determined (on the basis of an entry point in a random number table). The attitude held by the subjects throughout the experiment was that of trying to make as many correct guesses as possible. Because of the informal nature of the testing conditions, this study is presented as in no way conclusive, but as suggestive and exploratory. It is the authors' opinion that self-testing, when carried out in simple good faith, is an important method for the initial phases of parapsychological research. It should also be pointed out that in this case, motivated errors (as in shuffling, recording, etc.), or even sensory "leakage" would not be expected to bias the results presented here since the hypothesis of variance decline had not occurred to the subjects at the time of testing. This hypothesis is, in fact, mathematically independent of the subjects' task set, which was to score positively.

One additional matter of procedure must be mentioned because of its pertinence to the analyses in terms of half-run scores. During the time the testing was going on, the authors were studying with interest the work of Humphrey (2) and others on differential scoring rates associated with the "expansiveness" or "compressiveness" of subjects' response drawings in ESP tests. Briefly, the relative size of drawings was taken as an indication of the subject's mood at the time of testing and was found to be predictive of the direction of scoring. In our own testing, we decided to try to adopt two different psychomotor sets in responding to the ESP cards, deliberate sets analogous to those displayed spontaneously by Humphrey's subjects. This was a purely exploratory effort which was made in order to see if any discernible effects on scoring rate would be produced. In those runs which were written down with an "expansive" psychomotor set, the subject drew the symbols quickly, with considerable pressure on the page, making the five symbols larger than usual and generally overlapping the response space. The state of mind consequent to such a response set was generally increased exhilaration and greater awareness of whatever emotions were present in the subject at that time. Other runs were performed with a compressive or "constrictive" response set. In this set, the subject drew his responses tightly, slowly, and regularly (sometimes timing them by subvocal counting), producing small and precisely-drawn response figures. The state of mind following such runs was generally a vaguely unpleasant one, often attended by "writer's cramp."

In 13 of the 29 series, these expansive and constrictive modes of response were alternated either from run to run or from page to page for the purpose of making a comparison of scoring rates. Six other series were carried out with the expansive set alone in order to compare the scoring rate obtained with the rate expected by chance. The remaining 10 series were carried out with neither the expansive nor constrictive sets. Instead, an unmanipulated manner of response more natural to the subjects and typical of their previous experimentation was employed. (The original hope behind this effort was that the expansive psychomotor set might prove reliably to produce above-chance scoring. However, this did not prove to be true, and while the expansive runs did produce a higher scoring rate than the constrictive runs, the difference was not significant.)

Part I

The Decline of Run-Score Variance Within a Series of Runs

It was anticipated on the basis of the drug study mentioned previously (5) that, if each series were divided in half, the run scores in the first halves would consistently have a larger variance than those in the second halves. It was also suspected that this decline effect would occur most strongly in the shorter series. This was for two reasons: first, the effect had been found in a relatively short series of fairly intense effort (the drug series); and second, if the drop in run-score variance (R-variance) is linked to a drop in enthusiasm or focused attention (as suggested by the drug experiment), it would be expected most regularly in shorter series and less so in longer ones, in which dogged, uninspired effort tended to be the rule throughout and in which rests or diversions were bound to occur and possibly confound the process.

Each series was divided into halves, and an R-variance for each half was calculated. When the series had an odd number of runs, the middle run was dropped from the analysis. In this study the statistic which was used was the "sampling variance" (represented as S^2) which utilizes the obtained empirical mean (or $\overline{X}$) as the measure of central tendency. This is in contrast to a variance which is calculated around the "mean chance expectation" as the measure of central tendency. It has been pointed out that the latter is appropriate in ESP tests only when the null hypothesis under test is that only coincidence is operating in producing the results. In the present analysis, the null hypothesis allows for the existence of overall ESP in the scoring because the alternate hypothesis to be tested is whether ESP in one condition is operating differently from ESP in the other condition. In this case, the mean expectation cannot be assumed to be known, and the empirical estimate is required.[1]

Inspection of the data showed that the anticipated decline in variance from the first halves of the series to the second halves did occur in 22 of the 29 series. This overall result was tested by a sign test, in which a plus was assigned to each series which showed

[1] The formula used was: $S^2 = \text{Sum } (X-\overline{X})^2/N-1$, where N = the number of scores.

the expected effect, a minus to each which did not, and a CR was computed to test the significance of the difference between the sums of pluses and minuses. The obtained CR was 2.59, which is significant at $p < .01$ (two-tailed).

It was also clear from visual inspection that the effect occurred most consistently in the shorter series, as had been anticipated. The first step in evaluating this effect was to separate the series into two groups as to length, using the median number of runs per series as an arbitrary dividing point. There were 12 "short" series from 16 to 36 runs in length, and 15 "long" series from 40 to 90 runs in length. These groups were subjected to analysis separately. In interpreting the results arising from this division, it must be remembered that they are to a certain extent post hoc.

First, a test was carried out to see if the total variance of all the scores in the first-half runs of the series was significantly larger than the variance of all the second-half scores. This was done by dividing the first-half variance by the second-half variance to get an F-ratio which could be tested for significance.

Both sets of series were also tested for the consistency of decline in R-variance from first to second half, using the sign-test procedure described above.

Also, it was determined whether or not the R-variance of the first-half scores was significantly larger than the binomial expectation and whether the R-variance of the second-half scores was significantly smaller. Since here the hypothesis is one of "ESP vs. chance," it is appropriate to use the theoretical mean as the measure of central tendency in calculating the variance. The obtained variance may be divided by the binomial variance if the former is larger, or divided into the binomial variance if it is smaller. The obtained F-ratio may be tested for significance, using N degrees of freedom for the obtained variance and infinite degrees of freedom for the binomial variance.

Results

A. Short Series

The results for the short series are given in Table 1.

1. *Magnitude of Difference Between R-Variance, First and*

Table 1

RUN-SCORE VARIANCE OF FIRST VS. SECOND HALF OF SHORT SERIES

SERIES	NO. OF RUNS	VARIANCE		F	SIGN
		FIRST HALF	SECOND HALF		
1 (J.C.)	16	6.21	2.84	2.19	+
2 (J.C.C.)	20	9.34	6.84	1.37	+
3 (J.C.C.)	20	6.01	2.90	2.07	+
4 (J.C.)	20	11.16	3.17	3.52	+
5 (J.C.)	20	5.34	2.46	2.17	+
6 (J.C.)	20	2.89	2.04	1.42	+
7 (J.C.)	20	7.83	5.07	1.54	+
8 (J.C.)	20	4.90	1.82	2.69	+
9 (J.C.)	20	4.89	2.00	2.45	+
10 (J.C.)	20	5.96	1.73	3.45	+
11 (J.C.)	24	6.15	3.48	1.77	+
12 (J.C.)	28	4.92	3.34	1.47	+
13 (J.C.)	28	4.13	1.60	2.58	+
14 (J.C.)	36	6.50	1.95	3.33	+
Total	312	5.71	2.93	1.94 $p < .001^a$	$p < .0001^a$

[a]One-tailed

Second Halves of Series. The sampling variance of all the scores from the first halves of the series is equal to 5.71; and that of the second-half scores is 2.93. The F-ratio of the two is 1.94, with 155 and 155 degrees of freedom, which is associated with a $p < .001$. First-half variance is significantly larger than second-half variance.

2. *Consistency of Difference.* The sign test gives significant evidence for the generality of distribution of the decline effect. There are 14 pluses (series showing decline) and 0 minuses (series not showing decline). This yields a p-value of $< .0001$ (one-tailed).

3. *Appropriateness of Binomial Model.* The R-variance of all first-half scores (when calculated around the theoretical mean) is 5.78, which is significantly larger ($p < .01$; 156, ∞ degrees of freedom) than the binomial expected value; and the R-variance of second-half scores is 2.82, which is significantly smaller than the binominal value at $p < .01$, with ∞ and 156 degrees of freedom.

B. Long Series

The results for the long series are given in Table 2.

1. *Magnitude of Difference Between R-Variance, First and Second Halves of Series.* The empirical variance of all first-half runs is

Table 2

RUN-SCORE VARIANCE OF FIRST VS. SECOND HALF OF LONG SERIES

Series	No. of Runs	Variance		F	Sign
		First Half	Second Half		
1 (J.C.)	40	5.17	2.87	1.80	+
2 (J.C.)	40	4.75	3.06	1.55	+
3 (J.C.)	40	1.84	3.14	.59	−
4 (J.C.)	40	4.05	3.33	1.22	+
5 (J.C.)	44	2.87	5.01	.57	−
6 (J.C.)	48	4.87	4.35	1.12	+
7 (J.C.)	48	3.19	4.43	.72	−
8 (J.C.)	48	3.17	4.09	.78	−
9 (J.C.)	48	3.21	2.93	1.10	+
10 (J.C.)	48	5.00	2.91	1.72	+
11 (J.C.)	50	2.86	6.38	.45	−
12 (J.C.)	60	4.93	6.53	.75	−
13 (J.C.)	60	5.47	4.19	1.31	+
14 (J.C.)	60	3.96	3.13	1.27	+
15 (J.C.)	88	3.51	5.59	.63	−
Total	762	4.00	4.21	.95 (n.s.)	(n.s.)

equal to 4.00; and that of second-half runs is 4.21. The ratio of the two is .95, which is clearly not significant.

2. *Consistency of Difference.* There are 8 pluses and 7 minuses. Thus there is no consistency of decline or incline of the size of R-variance across halves for these 15 series.

3. *Appropriateness of Binomial Model.* The R-variance of all first-half scores (when calculated around the theoretical mean) is 4.00, which is exactly equal to the expected variance. The R-variance of second-half scores is 4.11, which also is not significantly different from the binomial expectation.

Discussion

It can be seen that the expectations (as measured by three evaluations) regarding a decline in R-variance across an uninterrupted period of effort have been borne out in data where the series are relatively short, but not with longer series. This would seem to suggest that the variables responsible for the decline, whatever they may be, are effective only when the effort is a fairly intense and briefly sustained one. The fact that all three evaluations yielded significant results gives some measure of internal confirmation to the effect, since they are, to some degree, mathematically independent.

PART II

THE DECLINE OF HALF-RUN SCORE VARIANCE WITHIN THE RUN

In a previous article, Carpenter (1) reported a study which dealt with the variability of scoring of half-runs. For such an analysis, each run is divided into a top and bottom half, omitting the middle (13th) cell, and the number of hits for each half-run is obtained. It was found in that research that in seven independent series totaling 660 runs, the top halves of runs tended to yield significantly more variable scoring (i.e., to deviate from chance more in both the psi-hitting and -missing directions) than the bottom halves.

Since the body of data reported here had already been compiled in order to make the evaluations for decline of R-variance reported above, the authors decided to analyze it also to find if the decline in half-run variance (HR) within the run had been repeated. Theoretical variance was the measure employed in this analysis.

Results

Inspection showed that, as a whole, the data did not give a significant repetition of the effect. With an expected variance equal to 1.92, the variance obtained for top HR scores was 1.96, and for bottom HR scores, 1.92.

However, most of the data clearly involved important differences in psychomotor response set from those reported in the earlier study. Both the expansive and constrictive sets of runs involved deliberate modifications of response style and of consequent experience from that "natural" to the subjects. Therefore it was decided to divide the data into three groups in terms of psychomotor set (unmodified, expansive, and constrictive) and to analyze the three groups separately for their HR-score variances.

The obtained HR-variances for the three groups are presented in Table 3 along with the values obtained from the previous study,

Table 3

HALF-RUN SCORE VARIANCE OF RUNS WITH DIFFERENT RESPONSE SETS

Condition	No. of Runs	Variance of Top Half-Run Scores	df	Variance of Bottom Half-Run Scores	df	F	p
Previous Study (1)	660	2.22*	660	1.83	660	1.22	≅.01[a]
Present Study							
1. Unmodified Response	473	2.23*	473	1.86	473	1.20	<.05[b]
2. Expansive Response	434	1.90	434	1.91	434	1.00	n.s.[b]
3. Constrictive Response	193	1.48*	193	2.03	193	1.37	<.05[b]

*Significantly different from chance variance at $p < .02$
[a]One-tailed
[b]Two-tailed

which are included for the sake of comparison. It can be seen that the unmodified-response group of runs, which are comparable in response style to those of the earlier research, closely replicated the earlier values. The top HR-variance is almost exactly equal to the figure obtained earlier (and is significantly larger than the chance value); and the bottom HR-variance is also not appreciably different from that found earlier. The difference between the top and bottom HR-variance in the unmodified responses is in the direction of a decline, and is marginally significant at $p < .05$.

The expansive-response runs yielded top and bottom HR-variances which were both quite close to chance expectation, with no appreciable change in size from top to bottom.

A strong reversal of the effect was found in the constrictive runs, with the top HR-variance significantly smaller than chance, and the bottom HR-variance insignificantly larger than chance. The incline effect is marginally significant at $p < .05$.

DISCUSSION

Because the data analyzed here were not planned originally as an effort to repeat and elaborate upon the earlier findings, and because the division of the data reported here arose from post hoc considerations, these results must be considered to have only suggestive and exploratory value. However, it is the authors' opinion that the results suggest a meaningful picture when considered in the light of other variance findings and are made more interesting by that fact. There appears to be some reason to suspect that relatively large deviations in scoring, irrespective of direction, are associated with factors such as freshness, spontaneity, and alertness (4). If that is the case, then normally a decline in these factors and a corresponding decline in scoring variability would be expected to occur as the effort at response proceeds during a run. Such an effect was found in previous research and was replicated in comparable data here. In cases where an artificially constricted, nonspontaneous mode of response is employed, however, it may be the constriction itself which lessens as guessing progresses, producing an increase of scoring variability. The expansive-response runs gave a result not so readily interpreted. The top and bottom HR-variances neither differed significantly from each other nor from chance expectation. It might be suspected that while the expansive-response style avoided causing any initial inhibition of whatever factors make for high variability, it did not succeed in producing a genuine increase of those factors. These suggestions will have to be clarified and evaluated by future research.

Conclusion

This study is an exploratory effort, investigating the effects of variables which appear to influence ESP scoring by affecting the tendency of scores to deviate in both directions from chance expectation. It was seen that when the series involved relatively few runs, the R-variance tended to decline as testing proceeded, but showed no reliable function when 40 or more runs were performed at a sitting. A hypothesis for future research—which has, in fact, received some confirmation (5)—could be stated as follows: When a subject is asked to persist, uninterrupted, in a card-guessing task in a relatively unvaried manner for a series of 16 to 36 runs in length, a decline in variance of run scores will occur from the first to the second half of the series.

A previous finding, that scoring variability tends to decline within the run, has been repeated with comparable data. The present results indicate that runs which were performed in the deliberately manipulated response styles called "constrictive" and "expansive" produced, respectively, a reversal of the decline effect, and no differentiation between halves at all. Runs performed with an unmodified response style yielded a decline similar to that shown in the previous study.

This report, of course, represents the work of only two subjects and therefore must be approached cautiously as to its potential generalizability. At the same time, it should be remembered that neither of the subjects involved has shown any outstanding ability to produce reliably high- or low-scoring ESP data. They belong to the great majority of ESP subjects whose efforts have been construed in the past as being unexciting or even meaningless. The prospect of generalizing to the unexceptional subjects, who are so much more easily found than the Pearces and Stepaneks, and of "finding ESP" in their data is an exciting one.

Determining just what variables might be involved in producing the effects described here seems to be the task of future, more sophisticated, psychological research. Such work has been initiated, and, as stated above, this report is intended to serve primarily as a preface.

Appendix

Some additional analyses, for which there were no particular expectations, were also made.

The run scores for all runs were pooled, including those middle-of-series runs which were omitted in the analyses of Part I, and the overall scoring rate and R-variance were computed for the sake of a comparison with chance expectation.

The total number of hits obtained was 5,314 in 1,094 runs, for a mean run score of 4.86. A negative deviation of 156 hits was obtained which, when divided by the standard deviation of 66.16, yields a CR of 2.36. Thus, unexpectedly, it appears that psi-missing was characteristic of these data in a small but significant way ($p = .018$, two-tailed). More precisely, it should be said to have been characteristic of the data of the major subject, J.C., since J.C.C.'s two series produced an insignificant positive deviation. When J.C.'s 27 series are considered alone, they show a negative deviation of 163 hits, which is associated with a $p = .012$ (two-tailed).

The R-variance for all runs pooled was 4.12, which is not significantly different from chance expectation.

References

1. Carpenter, J. C. Scoring effects within the run. *J. Parapsychol.*, 1966, 30, 73-83.
2. Humphrey, Betty M. Success in ESP as related to form of response drawings: I. Clairvoyance experiments. *J. Parapsychol.*, 1946, 10, 78-106.
3. Osis, K., and Dean, D. The effect of experimenter differences and subjects' belief level upon ESP scores. *J. Amer. Soc. psych. Res.*, 1964, 58, 158-85.
4. Rogers, D. P. Negative and positive affect and ESP run-score variance. *J. Parapsychol.*, 30, 151-59.
5. Rogers, D. P., and Carpenter, J. C. The decline of variance of ESP scores within a testing session. *J. Parapsychol.*, 1966, 30, 141-50.
6. Van de Castle, R. L. Differential patterns of ESP scoring as a function of differential attitudes toward ESP. *J. Amer. Soc. psych. Res.*, 1957, 51, 43-61.

Dept. Psychology
Columbus State School
1601 W. Broad St.
Columbus, Ohio 43223

Kingswood Farm
Hilliard, Ohio 43026