Journal of Parapsychology, Vol. 47, September 1983

PREDICTION OF FORCED-CHOICE ESP PERFORMANCE

PART II. Application of a Mood Scale to a Repeated-Guessing Technique

By James C. Carpenter

ABSTRACT: Previous efforts to enhance the reliability of forced-choice ESP test performances have relied on the majority-vote technique, the technique of summing multiple calls at a single target to generate a single set of calls. This technique, however, requires a positive scoring trend in the data to be summed. Index sampling to discriminate between psi-hitting and psi-missing trends is of no help unless the relative size of the departure of the total score from chance can be independently predicted. The V scale, developed as a predictor of run-score variance (RSV), will serve in this capacity. Presented is an elaborated majority-vote technique that uses prediction of RSV and index sampling within each run. With this expanded technique, the results of the nine series previously described in Part I of this report are analyzed. The results demonstrate that, given an accurate predictor of RSV, this procedure can boost the reliability of ESP performance by generating single lists of majority-vote calls that are more accurate than the overall body of calls from which they are drawn.

Reliability has been an important issue for experimental parapsychology since the field's beginning. The relative failure to discover reliable, replicable effects has been a continual frustration to researchers and has served as evidence of the field's dismissability to its critics. The enhancement of the reliability of parapsychological phenomena is desirable scientifically, because only systematically reproducible phenomena can be methodically studied. More pragmatic investigators have turned their thoughts to the potential exploitation of more reliable ESP phenomena. The achievement of even a rudimentary system for permitting the reliable acquisition of information by means of ESP would be a step forward. This article reports some elaborations

Part of the work reported here was supported by a grant from the John E. Fetzer Foundation. I also want to thank the following people whose efforts contributed to this part of this research: Richard Broughton, Josephine Carpenter, David Dillman, Ed Kelly, and Donna Rigsbee. A summary of this work was presented at the convention of the Parapsychological Association in 1980.

of earlier efforts in this direction, with data that make the general approach seem promising.

Enormous difficulties have seemed to stand in the way of any such development. Ordinary subjects in guessing tasks produce scoring rates so unreliable as to be apparently useless. Extraordinary subjects are extraordinarily rare, and even they suffer sudden, inscrutable lapses in performance. Despite the difficulties posed by this lack of reliability, even the smallest deviation from chance expectation, if predictable, represents some genuine acquisition of information that might be maximized and serve as the basis for reliable information retrieval.

Two requirements must be met to permit such a process: (1) The subject(s) must produce an extra-chance guessing performance that is stable in mode (i.e., is consistently either above or below chance) or, if it is unstable in mode, there must be some means of ascertaining whether hitting or missing tendencies predominate at any given time. (2) The information must be concentrated in a way that will permit an acceptably accurate retrieval.

The repeated-guessing technique. The method basically used in this study for the task of heightening information accuracy is the "majority-vote technique" (Fisk & West, 1956, 1957). With this procedure, several calls are made for each target, and the majority vote of the calls is taken as the best guess. By this method, a low order of deviation from chance expectation can be "condensed" to higher levels of reliability. For example, if chance expectation for correct guessing is 20%, and guessing success is reliably progressing at 25%, repeated guessing on a single set of targets would tend eventually to "wash out" the incorrect calls and yield a correct majority vote for each target.

However, before this repeated-guessing method will work, our first requirement will have to be met, that is, one will need to know when guessing is proceeding at an extra-chance rate and, given that, in which mode (psi-hitting or psi-missing).

One obvious method of achieving success is to use the extraordinary subject, the person whose guessing performance is fixed at an extra-chance level and is stabilized in mode for long periods of time. A study by Ryzl (1966) illustrates the utility of the majority-vote procedure when a fairly strong and consistent above-chance performance is available. Ryzl worked out a code by which certain sequences of 10 (white or green) colors were allotted to each of the three-digit numbers ranging from 000 to 999. Five of these numbers were picked at random, and their associated color sequences were used as targets in repeated-guessing procedures. All five numbers were "received" correctly by Ryzl's high-scoring subject, Pavel Stepanek.

Not having a consistently high-scoring subject, I turned to another approach that has been suggested for heightening the utility of the repeated-guessing technique, that is, sampling the guessing as it proceeds in the hope that the hitting rate of the sample will be a predictive index of the hitting rate of the remainder (Taetzsch, 1962). By such a procedure, the experimenter at the "receiving" end of a "transmission" has prior knowledge of some targets but not others, the latter being the unknown "message" to be sent. The scoring rate on the sample (or index) is used to predict the scoring rate on the message. Binary targets (white-black, yes-no, etc.) are used. The index calls of the series of guesses are scored, and, if scoring is proceeding in the psi-hitting mode, the calls of the message are assumed to be positive as well and are tabulated as they stand in the repeated-guessing method. However, when the scoring mode of the index calls of a series of guesses is psi-missing, the message guesses are reversed to their alternates (all "white" calls are considered "black" and vice versa), and the reversed calls tabulated. Thus, psi-missing tendencies are separated from psi-hitting ones and are exploited just as fully in the accumulation of information.

The difficulty with any such procedure for predicting scoring mode is that parts of a subject's ESP performance do not tend to be highly correlated with each other. Efforts to demonstrate internal reliability of ESP performance, such as split-half correlations, have generally failed (Schmeidler, 1964). That is, the scoring rate of part of a subject's guessing (analogous to "index" calls) is not generally predictive of scoring rate on the remaining calls. Thus, at least for most subjects, simple sampling will not serve to discriminate scoring mode in message calls.

On the other hand, the scoring mode of index calls is more strongly predictive of mode for other guesses in that run if the size of the deviation from chance expectation of the whole run can be reliably ascertained. For example, if the deviation from chance of a set of calls is large, a randomly drawn index sample will tend to be in the same mode (above or below chance) as the remainder of the run.

To understand this point, consider a single run of 24 binary calls. Chance expectation for the run is 12 hits. Suppose a score with a large deviation is obtained, for example 18 (in the positive mode) or 6 (in the negative); and suppose further that 12 of the calls in the run are randomly selected as an index and the remaining 12 as the message.

With a score of 18, the score on the index calls will tend to be above chance, as will the score on the message calls, because the scoring rate on the whole set is 75% and not the 50% expected by chance. Because, at most, 12 of the 18 hits can appear in either subsample, the possible scores for each sample will range from 6 to 12. One of these possible cases represents exactly MCE (i.e., the score of 6); the other six cases are in the psi-hitting direction. Thus, in six out of seven possible cases, a psi-hitting index score will occur, and five of these will be matched with a psi-hitting message score. In the extreme case, if all 24 calls are correct, the index sample will always score in the positive mode at 100%, which will be an accurate prediction of the scoring mode of the message, also positive at 100%. By the same token, given an overall score with a large negative deviation, index calls will tend to be below chance and match below-chance scoring in the remainder. The larger the deviation for the run as a whole, the more likely that the mode of index and message will be the same.

At the same time, the opposite relationship will be obtained for runs producing, overall, very small or zero deviations. A run of 24 binary calls that contains 12 hits will obviously yield a negative relationship of mode between the scoring on a 12-call index and a 12-call message. A positive score on the index, 7 or more, can only be matched to message scores of 5 or fewer, all in the negative mode. Thus, the use of the sampling technique on runs yielding small deviations will produce a result opposite to that intended.

This point is further elaborated graphically in Table 1, which shows the possible combinations of 12-trial index and 12-trial message scores for a 24-trial run when run-score deviations range from −6 to +6. Note that the distributions are symmetrical for positive and negative deviations of the same size. At the +6 deviation, five of the possible combinations are correctly predictive (in the same mode), none are incorrectly predictive, one case may be considered nonpredictive (i.e., the index score is equal to 6, or MCE), and the final case predicts a message score equal to 6, which may be considered neither correct nor incorrect. The number of *cases* of modal agreement increases directly with the size of the run-score deviation. The percentage of agreement for all possible cases is summarized in Table 2.

Simply scanning the number of cases of modal agreement or disagreement, however, would be misleading because it is based simply on a logical tabulation of the possible cases, which does not reflect their varying probabilities of occurrence. These probabilities follow a hypergeometric distribution. In an experiment with N total trials, n index trials, and H hits, the probability that x of the hits will occur in the index trials is the hypergeometric probability that x "successes" will occur in a set of n items chosen from a set of N items containing H successes. The probability of obtaining x hits in the index cases and $H - x$ hits in the remainder is given by the formula:

$$\frac{\binom{H}{X}\binom{N-H}{n-x}}{\binom{N}{n}}$$

These probabilities for the run-deviation cases of $+6$ to -6 are given in Table 1.

Table 2 summarizes the percentage of modal agreement or disagreement for all possible cases and also gives the summary probabilities of agreement or disagreement for each run-score deviation. These figures show that modal agreement is expected more strongly as the run-score deviations become larger, whereas disagreement is most likely when small or zero deviations are found.

If it should be possible to discriminate predictively between sets of calls that will yield large deviations and those that will yield small ones, then one could apply the index-sampling technique, extracting index calls as discussed above but adjusting the treatment of the message calls depending on the size of the predicted deviation. Specifically, predicted-large deviation runs would be treated as suggested, with the mode of the message calls assumed to be the same as the mode of the index calls. Predicted-small deviation runs, on the other hand, would be treated in the opposite fashion, with the mode of the message assumed to be the opposite of the mode of the index. Thus the three-step sampling process would be expanded by an additional, initial step. That is: (a) binary-target runs of calls would be separated before scoring into those expected to yield large deviations (henceforth: "L-D") and those expected to yield small deviations ("S-D"); (b) a sample of targets would be extracted randomly from the target list (kept constant across all runs within each series) and the calls to those targets would be scored as the index; (c) the calls on the remaining message targets in each run would be assumed to represent a positive mode of scoring if either the run was L-D and the index score was positive, or if the run was S-D and the index score was negative. The mode on the message would be assumed to be negative either if the run was S-D and the index score was positive, or if the run was L-D and the index score was negative; (d) assumed positive message calls would be entered as they were made into a repeated-guessing

TABLE 1. POSSIBLE ARRAYS OF INDEX AND MESSAGE SCORES FOR A 24-TRIAL RUN

6 Hits ($d=-6$)			7 Hits ($d=-5$)			8 Hits ($d=-4$)			9 Hits ($d=-3$)			10 Hits ($d=-2$)		
I	M	p	I	M	p	I	M	p	I	M	p	I	M	p
												10	0	.000034
									9	0	.000168	9	1	.001346
						8	0	.000673	8	1	.004543	8	2	.016658
			7	0	.002288	7	1	.012922	7	2	.039978	7	3	.088841
6	0	.006865	6	1	.032037	6	2	.082918	6	3	.155472	6	4	.233208
5	1	.070611	5	2	.151030	5	3	.236909	5	4	.299838	5	5	.319828
4	2	.242726	4	3	.314645	4	4	.333154	4	5	.299838	4	6	.233208
3	3	.359595	3	4	.314645	3	5	.236909	3	6	.155472	3	7	.088841
2	4	.242726	2	5	.151030	2	6	.082918	2	7	.039978	2	8	.016658
1	5	.070611	1	6	.032037	1	7	.012922	1	8	.004543	1	9	.001346
0	6	.006865	0	7	.002288	0	8	.000673	0	9	.000168	0	10	.000034
p_S	*.993135*			*.963387*			*.806974*			*.599676*			*.319828*	
p_O	*.00000*			*.004576*			*.027190*			*.089378*			*.213758*	
p_{no}	*.006865*			*.032037*			*.082918*			*.115472*			*.233208*	

11 Hits ($d=-1$)			12 Hits ($d=0$)			13 Hits ($d=+1$)			14 Hits ($d=+2$)			15 Hits ($d=+3$)		
I	M	p	I	M	p	I	M	p	I	M	p	I	M	p
			12	0	.000000	12	1	.000005	12	2	.000034	12	3	.000168
11	0	.000005	11	1	.000053	11	2	.000317	11	3	.001346	11	4	.004543
10	1	.000317	10	2	.001611	10	3	.005817	10	4	.016658	10	5	.039978
9	2	.005817	9	3	.017898	9	4	.043627	9	5	.088841	9	6	.155472
8	3	.043627	8	4	.090611	8	5	.157058	8	6	.233208	8	7	.299838
7	4	.157058	7	5	.231963	7	6	.293175	7	7	.319828	7	8	.299838

6	5	.293175	6	6	.315727	6	7	.293175	6	8	.233208	6	9	.155472
5	6	.293175	5	7	.231963	5	8	.157058	5	9	.088841	5	10	.039978
4	7	.157058	4	8	.090611	4	9	.043627	4	10	.016658	4	11	.004543
3	8	.043627	3	9	.017898	3	10	.005817	3	11	.001346	3	12	.000168
2	9	.005817	2	10	.001611	2	11	.000317	2	12	.000034			
1	10	.000317	1	11	.000053	1	12	.000005						
0	11	.000005	0	12	.000000									
p_S	*.00000*			*.00000*			*.00000*			*.319828*			*.599676*	
p_O	*.413650*			*.684273*			*.413650*			*.213758*			*.089378*	
p_{no}	*.293175*			*.315727*			*.293175*			*.233208*			*.155472*	

16 Hits ($d = +4$)			17 Hits ($d = +5$)			18 Hits ($d = +6$)		
I	M	p	I	M	p	I	M	p
12	4	.000673	12	5	.002288	12	6	.006865
11	5	.012922	11	6	.032037	11	7	.070611
10	6	.082918	10	7	.151030	10	8	.242726
9	7	.236909	9	8	.314645	9	9	.359595
8	8	.333154	8	9	.314645	8	10	.242726
7	9	.236909	7	10	.151030	7	11	.070611
6	10	.082918	6	11	.032037	6	12	.006865
5	11	.012922	5	12	.002288			
4	12	.000673						
p_S	*.806974*			*.963387*			*.993135*	
p_O	*.027190*			*.004576*			*.00000*	
p_{no}	*.082918*			*.032037*			*.006865*	

Note. Abbreviations are defined as follows: I, "index" score; M, the remainder ("message"); d, overall score deviation; p_S, the summary p for cases in the same mode within that column; p_O, the p for cases in the opposite mode; p_{no}, the p for no prediction. Note that for both I and M scores, 7 or more hits represents ΨH performance, 5 or fewer is ΨM, and 6 is exactly MCE.

TABLE 2
RELATION BETWEEN MODE OF 12-TRIAL INDEX AND 12-TRIAL MESSAGE IN A 24-TRIAL RUN AS A FUNCTION OF THE OVERALL RUN DEVIATION

	$d \geq 7$	$d=6$	$d=5$	$d=4$	$d=3$	$d=2$	$d=1$	$d=0$
Cases in same mode, %	100	71.43	50	33.33	20	9.09	0	0
Cases in opposite mode, %	0	0	25	44.44	60	72.73	83.33	92.31
Cases when one or both equal MCE, %	0	28.57	25	22.22	20	18.18	16.67	7.69
p of modal agreement	.1000	.9931	.9633	.8069	.5996	.3198	.0000	.0000
p of modal disagreement	.0000	.0000	.0045	.0271	.0893	.2137	.4136	.6842
p of one or both equal to MCE	.0000	.0137	.0640	.1658	.3109	.4664	.5863	.3157

tabulation, and assumed negative calls would be converted to their opposites (e.g., "black" would become "white," etc.) and would then be entered into the tabulation. If an index score equal to chance expectation was obtained, the remaining calls of the run would not be treated at all.

In summary, a crucial prerequisite for the index-sampling procedure is a reliable means of discriminating the size of the deviations in the runs of calls before the runs are scored. Basically, it appears that the majority-vote procedure can be used to enhance information acquisition if either of two independent predictions can be made successfully about the data to be collected: (1) the mode of the data as a whole or the mode of discriminable parts of the data (without index sampling) or (2) the size of the deviations of the subgroupings of data comprising the whole (with index sampling).

Subject mood as a predictor of run-score variance. Several years ago I became interested in studying the variance in ESP run-scores (RSV) as a parameter of ESP performance. As used here, *variance* refers to the tendency of the individual run scores to depart in either direction from the theoretical central tendency as defined by the binomial theorem. Sets of runs whose scores have large variance correspond to the L-D case discussed; sets with small variance scores correspond to the S-D case. The studies described in Part I of this report were aimed in part at developing and validating a set of mood adjectives for predicting RSV (the V scale). They were also intended as a means of testing a majority-vote procedure, focusing upon subsets of data from subjects whose scoring, because of prior results, might be expected to yield strong predictive discriminations.

This paper examines the utility of the RSV-based repeated-guessing procedure by applying it to each series among those studies in which a prior prediction as to the effectiveness of the V scale had been posed. Some pilot series, carried out without benefit of specific predictions, were thus excluded from this analysis.

In each series, subjects had carried out repeated guessing at a single list of targets, making the data amenable to such an analysis. Separate target lists were generated for each series, always after all guessing was completed and turned in to the experimenter, making all of this work precognitive in design.

Although in each case I, the experimenter, had access to both index and message targets, these analyses can be taken as a prototype of an actual information-acquisition situation, in which only the index targets are available at the time of analysis. For this purpose, we may imagine that the analyses of whole run scores and RSVs reported in Part I have not been done and that performance on only *index* targets is available. This would be the case in an actual "transmission" situation, in which an experimenter at the "receiving" end of the procedure would know only the identity of index targets and would be attempting to generate reliable predictions as to the identities of the message targets.[1] In fact, these majority-vote analyses were planned prior to the analyses of whole-run RSVs, so the success of those analyses was not considered a "given," much as it could not be a given in a "transmission" situation. From this perspective, calculation of whole-run RSVs, and the V-scale discriminations thereof, was only a secondary check on how the index-sampling, repeated guessing procedures should have fared. The V scale was used here as a predictor of RSV, which was presumed to be unknown, and thereby as a predictor of the kind of directional relationship expected between index and message-target scoring. Given the V-scale predictions, which of course were taken from mood responses and were independent of ESP performance, the majority-vote procedure described above was carried out.

---

[1] The use of terms such as *transmission* and *message* is only for the sake of clarity of presentation and should not be taken as suggesting that ESP cognizance is some process similar to electromagnetic forces of communication. In fact, I agree with Stanford (1978) that such a model is probably not very useful.

The first study reported in this article was the initial effort to cross-validate the V scale, and its major hypothesis was simply that the scale would function as an effective predictor of RSV for all subjects. The remaining three studies reported here explored the following hypotheses: that the V scale should be predictive when targets are personally salient to the subjects or, second, when subjects are sheep (Schmeidler & McConnell, 1958) or, third, when subjects are low in authoritarianism. What happens when the subject is a student of the experimenter was also addressed in Study 4 of Part I (previous article, this *Journal*). The reader is referred to Part I for procedural details of these studies. Except for the pilots, all of the series described here were analyzed by the majority-vote technique with V-scale predictions. The analysis of Study 1 is presented in more detail than subsequent ones, but the procedures were identical in all series.

Study 1
Initial Cross-Validation of V Scale

One list of 25 targets was randomly determined and used to score all runs. For purposes of testing the modified index-sampling procedure, 12 of the 25 targets were selected at random to serve as index targets, with the remaining 13 acting as message targets. This was done, prior to any examination of the subjects' calls, as follows: The 25 target numbers immediately following the 25 target numbers in the Rand book were inspected. If the set contained 12 odd numbers, those were taken as denoting the index-target locations. If the set of 25 numbers did not contain exactly 12 odd numbers, the next set of 25 was examined, and so on, until a set with 12 was found. This method was used in most subsequent series as well. In a few cases, I used the more informal method of having an uninvolved colleague thoroughly hand-shuffle a deck of 25 (or 24) cards containing 12 +'s and 12 or 13 O's, depending on the number of trials in the run. This placement of the 12 +'s was taken as locating the index trials. Unfortunately, the records denoting the method that was used in each case were lost. This categorization of targets was fixed across all runs.

The mood-adjective checklists (MACL) were scored, counting +1 for each checked item expected to be associated with large RSV and −1 for the items expected to relate to small RSV. The V-scale scores were divided into quartiles, with high-quartile scores (+3 to 0) taken as offering a prediction of large RSV and low-quartile scores (−3 to −9) taken as predictive of small RSV. Sittings that had a midrange V-scale score (+1 or +2) were taken as having no prediction and were omitted from further repeated-guessing analysis (although they were subsequently scored for hits as well).

Index sampling was then carried out as described above. Only index targets were scored. For high V-scale sittings, above-chance index scores for a run were taken as generating a prediction of above-chance scoring in the remaining calls of that run; the calls were entered as given into a repeated-guessing tabulation. Below-chance index scores in high V-scale sittings were taken as predicting below-chance scoring in the remainder of that run, and the calls were therefore reversed (each O became a +, and vice versa) and then entered into the tabulation.

For sittings with a low V-scale score, the opposite process was carried out. Runs with above-chance index scores were treated as offering a below-chance prediction for the remaining calls, and those calls were reversed. Runs with below-chance index scores were treated in the opposite fashion.

In any case, index scores exactly at chance expectation (a score of 6) were treated as offering no prediction, and the remaining calls of that run were not tabulated.

Table 3 reproduces one sitting of guesses from one subject of Study 1, along with illustrations of the target-type distinction and the chain of decisions. The sitting is a high V-scale case.

Results

When the guesses for all 94 usable runs were thus rendered and tabulated, majority votes were reached on all of the 13 message targets. Of the 13 decisions, 11 were correct and 2 incorrect, for an 85% rate of accuracy. Exact binomial p equals .0095. See Table 4.

Index and message scores for each run were then pooled into regular run scores to see if the V-scale predictions of RSV had been successful. The overall RSV associated with high V-scale scores was 5.65 (70 runs). RSV with low V-scale scores was 3.56 (55 runs), which is significantly smaller than σ^2 (CR = 2.04). The difference between them, however, was not significant (CR_d = 1.24; p = .11).

Control tests. Control tabulations of these data were made with the standard repeated-guessing technique described above (Fisk & West, 1956) and with the standard index-sampling modification (Taetzsch, 1961). For standard repeated guessing, calls for all 25 targets were assessed, since the technique does not have an "index–message"

TABLE 3
EXAMPLE OF ONE SITTING OF GUESSES FROM STUDY 1: A HIGH V-SCALE SCORE CASE

Target no.	Target type	Index targets	Runs 1	2	3	4	5	Votes tabulated For +	For O
1	Index	O	+	+	(O)	+	(O)		
2	Message		O	O	+	O	+	3	0
3	Message		+	O	+	O	O	2	1
4	Message		O	O	O	+	+	1	2
5	Message		+	O	+	O	+	2	1
6	Index	O	+	+	(O)	+	(O)		
7	Index	O	(O)	+	(O)	+	(O)		
8	Message		+	O	O	+	O	0	3
9	Index	O	(O)	+	(O)	+	+		
10	Index	+	O	O	O	O	(+)		
11	Message		+	+	+	O	+	2	1
12	Index	O	(O)	+	+	+	+		
13	Index	O	+	(O)	+	+	+		
14	Index	O	(O)	(O)	(O)	+	(O)		
15	Message		+	+	+	O	+	2	1
16	Message		+	O	O	+	+	0	3
17	Message		+	+	+	O	O	1	2
18	Message		O	O	+	O	+	3	0
19	Index	+	O	(+)	O	(+)	O		
20	Index	+	(+)	(+)	O	O	O		
21	Message		O	+	+	+	O	2	1
22	Index	O	+	(O)	(O)	(O)	+		
23	Message		+	O	+	+	+	1	2
24	Message		+	+	O	+	O	0	3
25	Index	+	O	(+)	(+)	O	(+)		
Index score			5 (below chance)	6 (at chance)	7 (above chance)	2 (below chance)	6 (at chance)		
Prediction for message calls given high V-scale score			Below chance	0	Above chance	Below chance	0		
Treatment of message calls for tabulation			Reverse	Omit	Keep	Reverse	Omit		

Note. Symbols in parentheses are the *correct* index guesses.

TABLE 4
STUDY 1: VOTES, DECISIONS, AND TARGETS

Target no.	Votes +	Votes O	Decision	Target
2	47	42	+	+
3	45	44	+	+
4	40	49	O	O
5	47	42	+	+
8	39	50	O	O
11	48	41	+	+
15	40	49	O	O
16	42	47	O	+
17	49	40	+	+
18	31	58	O	O
21	50	39	+	+
23	43	46	O	+
24	43	46	O	O
No. correct	626		11	
No. incorrect	531		2	
Hitting rate	54.1%		85%	

distinction. By this analysis, 11 of the 25 final decisions were correct and 14 incorrect, an insignificant negative hitting rate of 44%. This null result was to be expected by the reasoning given above, since the overall hitting rate for these data was not significantly positive. Whereas mean chance expectation was 4187.5, the observed number of hits was 4167, a hitting rate of 49.8%.

The standard index-sampling method, unaided by variance predictions, yielded 1606 correct votes and 1696 incorrect ones, for a rate of 48.6%. Of the 13 decisions reached, 7 were correct, a 53.8% rate of accuracy. This null result was also to be expected, since the overall RSV of 4.72 was smaller than chance expectation ($\sigma^2 = 6.25$, with 25 calls per run).

Similar control analyses have not been carried out on the studies reported below. Given the unreliable total hitting rates and RSVs obtained, one would expect such analyses to fare no better in the later samples.

Study 2
Personally Salient Targets

Series 1

This study tested the hypothesis that the V scale would be predictive of ESP guessing using personally salient names as targets. The same subject who had performed the pilot study also carried out the first cross-confirmation series. Six 5-run sittings with salient targets, each accompanied by an MACL, were carried out.

When the V scales were scored, only 15 runs (3 sets) were found to be paired with extreme quartile scores. This seemed insufficient for the purposes of the present investigation.

To provide the most thorough test of the repeated-guessing technique being studied, I made the conservative decision to divide the V-scale scores at the median (–1.5), treating all above-median scores as generating large RSV predictions, and all below-median scores as generating small RSV predictions. The arbitrary decision was made (and applied in other series henceforth) to consider 20 extreme quartile runs as a minimum number in this respect. Presumably the use of a median split should serve only to weaken the discriminatory power of the V scale. With such scoring, all six sets were given low-RSV predictions. Index sampling and tabulation were carried out accordingly. For results, see Table 5.

The voting on 11 targets reached majorities. Of those, 10 decisions were correct, 1 was incorrect (binomial $p = .0055$). Among the directionally corrected votes themselves, 173 were correct, 113 incorrect, for a 60% rate of accuracy (including the 2 ties). This is in contrast to the 49.2% hitting rate observed overall (MCE = 375; X = 369) which surely would have made a standard repeated-guessing tabulation a waste of time. On the other hand, scoring of the whole runs showed that the V-scale predictions upon which the present analysis depends were successful ($CR = 1.88$; $p < .05$).

Series 2

Four students in a class taught by the experimenter acted as subjects in Series 2. Beginning with this series and continuing throughout, the length of the ESP runs was changed to 24 calls. A median separation of V scales was resorted to again. This generated large RSV predictions for three sets and small-RSV predictions for

TABLE 5
VOTES, DECISIONS, AND TARGETS FOR STUDY 2

Message target no.	Series 1 Votes Ma	Series 1 Votes Sis	Series 1 Decision	Series 1 Target	Series 2 Votes A	Series 2 Votes B	Series 2 Decision	Series 2 Target
1	8	14	Sis	Sis	20	24	B	A
2	10	.12	Sis	Ma	21	23	B	B
3	5	17	Sis	Sis	24	20	A	A
4	5	17	Sis	Sis	28	16	A	A
5	8	14	Sis	Sis	19	25	B	B
6	9	13	Sis	Sis	26	18	A	B
7	14	8	Ma	Ma	25	19	A	A
8	11	11	Tie	Sis	22	22	Tie	B
9	10	12	Sis	Sis	15	29	B	B
10	9	13	Sis	Sis	15	29	B	B
11	11	11	Tie	Ma	17	27	B	A
12	15	7	Ma	Ma	23	21	A	A
13	12	10	Ma	Ma				
No. correct	173		10		283		8	
No. incorrect	113		1		245		3	
Hitting rate	60.5%		90.9%		53.6%		72.7%	

Note: "Tie" refers to no decision owing to equal number of votes for both targets.

eight sets. Index scoring and message tabulations were carried out accordingly. The results are given in Table 5.

Of the 11 decisions reached, 8 were correct and 3 incorrect, for a rate of 72.7% (exact binomial p = .08). Of the 528 votes, 386 (53.6%) were correct. When total run scores were calculated, the V-scale discrimination was found to have produced a nonsignificant trend in the predicted direction.

STUDY 3
SHEEP VS. GOAT ATTITUDE AS A MODERATING VARIABLE

Series 1

In Series 1, an insufficient amount of data required a median V-scale scoring. Index scoring and tabulation were carried out. All 12 of the message targets received majority votes, 8 correct and 4 incorrect. Exact binomial $p = .1208$. See Table 6.

Of the 732 votes entered into the repeated-guessing tabulation, 386 (52.7%) were correct. When total run scores were calculated, the V-scale discrimination for sheep was in the expected direction but was not significant.

Series 2

V-scale scoring indicated sufficient data to use the presumably more reliable extreme-quartile predictions. Index sampling and tabulation were carried out as before on the sheep data (see Table 6). Twelve decisions were reached, 8 correct and 4 incorrect (exact binomial $p = .1208$). The tabulated votes showed a hitting rate slightly above chance (50.6%). The V-scale discrimination of RSV was again in the predicted direction but not significant.

TABLE 6
VOTES, DECISIONS, AND TARGETS FOR STUDY 3:
SHEEP SUBJECTS ONLY

Message target no.	Series 1 Votes +	Series 1 Votes O	Series 1 Decision	Series 1 Target	Series 2 Votes +	Series 2 Votes O	Series 2 Decision	Series 2 Target
1	32	29	+	+	24	17	+	O
2	33	28	+	+	21	20	+	+
3	31	30	+	O	17	24	O	+
4	35	26	+	+	21	20	+	O
5	32	29	+	O	20	21	O	O
6	34	27	+	O	17	24	O	O
7	27	34	O	O	22	19	+	+
8	29	32	O	O	19	22	O	O
9	35	26	+	+	20	21	O	O
10	34	27	+	O	22	19	+	+
11	37	24	+	+	24	17	+	+
12	26	35	O	O	18	23	O	+
No. correct	386		8		249		8	
No. incorrect	346		4		243		4	
Hitting rate	52.7%		66.7%		50.6%		66.7%	

Study 4
Authoritarianism as a Moderating Variable

Series 1

Ten students in a class taught by me participated in Series 1 of Study 4. Six subjects scored below the median on the California F scale (Adorno et al., 1950), and their data were selected for further analysis.

Extreme quartile V-scale scoring, index sampling, and tabulating were carried out. Majority decisions were reached on all 12 message targets, 8 of which were correct and 4 incorrect (binomial $p = .1208$). Of the 612 modified votes, 328 were correct (53.6%). See Table 7.

Calculation of total run scores showed that the discrimination of RSV by the V scale had produced a trend in the expected direction.

Series 2 and 3

Students in two classes of a colleague served as subjects of Series 2 and 3.

In Series 2, 14 low-F subjects were found. The results of the modified repeated-guessing procedure on their data are given in Table 7. Of the 12 decisions reached, 10 were correct, and 2 incorrect (binomial $p = .0161$). Of the 684 votes cast, 368 (53.8%) were correct. Analysis of RSV in terms of the extreme-quartile V-scale prediction showed a marginally significant discrimination ($CR_d = 1.34$; $p < .10$).

In Series 3, the data of the 17 low-F subjects were rendered as before. The results are given in Table 7. Of the 10 decisions reached, 7 were correct, 3 incorrect (binomial $p = .1172$). Of the 792 votes cast, 419 (52.8%) were correct. When total run scores were calculated, the extreme-quartile V-scale predictions yielded a significant discrimination, ($CR_d = 2.01$; $p < .05$). In both series, V-scale predictions on the data of high-F subjects had produced null results.

Series 4

The data collected thus far in Study 4 were examined in terms of both the F-scale and sheep-goat divisions. It was found that the exclusion of goats from the low-F pool of subjects did strengthen the V-scale discrimination, but only in data collected from my own classes. Perhaps the instructor-student relationship, with the instructor clearly

TABLE 7

VOTES, DECISIONS, AND TARGETS FOR STUDY 4: LOW-AUTHORITARIAN SUBJECTS ONLY

Message target no.	Series 1				Series 2				Series 3				Series 4			
	Votes		Decision	Target	Votes		Decision	Target	Votes		Decision	Target	Votes		Decision	Target
	+	O			+	O			+	O			+	O		
1	28	23	+	O	33	24	+	+	32	34	O	O	11	9	+	+
2	19	32	O	+	33	24	+	+	30	37	O	O	8	12	O	O
3	30	21	+	+	24	33	O	O	35	31	+	O	12	8	+	+
4	26	25	+	O	26	31	O	O	33	33	Tie	O	8	12	O	O
5	25	26	O	O	26	31	O	+	41	25	+	O	9	11	O	O
6	23	28	O	O	25	32	O	O	37	29	+	+	9	11	O	O
7	31	20	+	+	29	28	+	O	42	24	+	+	13	7	+	O
8	20	31	O	O	27	30	O	O	29	37	O	O	14	6	+	+
9	22	29	O	O	30	27	+	+	39	27	+	+	6	14	O	O
10	21	30	O	+	29	28	+	+	33	33	Tie	+	8	12	O	O
11	20	31	O	O	33	24	+	+	32	34	O	+	12	8	+	O
12	34	17	+	+	27	30	O	O	27	39	O	O	14	6	+	+
No. correct	328		8		368		10		419		7		138		10	
No. incorrect	284		4		316		2		373		3		102		2	
Hitting rate	53.6%		66.7%		53.8%		83.3%		52.8%		70.0%		57.5%		83.3%	

Note: "Tie" refers to no decision due to equal number of votes for both targets.

identified with the research, provided a context in which the attitude variable had greater saliency. For my own classes, the V-scale discrimination for low-F sheep was significant, whereas only a slight trend was observed for low-F goats. For subjects of other classes, the low-F sheep and the low-F goats produced discriminations of about the same magnitude.

Because of this, another series with subjects from another of my classes was planned in which it was expected that low-F sheep would produce a strong V-scale discrimination and in which the data from only those subjects were to be analyzed by the revised repeated-guessing procedure.

Twelve students participated. Four of the six low-F subjects were also sheep, and their data were rendered as before.

Decisions were reached on 12 targets, of which 10 were correct and 2 incorrect (binomial $p = .0161$). Of the 240 votes tabulated, 138 were correct (57.5%) (see Table 7). A calculation of total run scores showed that the V-scale discrimination for low-F sheep had been successful ($CR_d = 2.38$; $p < .01$) but that the low-F goats showed only a slight trend in the expected direction.

Summary

Excluding pilot series, nine groups of subjects carried out guessing at nine sets of targets, and their guesses at the message items of each set were subjected to the information retrieval procedure under study. Majority-vote decisions were reached on 105 message items, of which 80 were correct and 25 incorrect. Thus, the decisions represented a 76.2% rate of accuracy (Z corrected for continuity = 5.13; $p < .001$).

One can also examine the hitting rate of the raw votes that entered into the majority-vote tabulation. Across all nine series, 2,770 were correct and 2,553 were incorrect (Z corrected for continuity = 5.62; $p < .001$).

Although still falling short of a truly useful level of efficiency, this 26.2% overall enhancement above the chance level for the majority-vote decisions seemed reliable enough to warrant further study. Part III of this report (a future issue of this *Journal*) will discuss subsequent efforts.

References

Adorno, T. W., Frenkel-Brunswick, E., Levinson, D. J., & Sanford, R. N. (1950). *The authoritarian personality*. New York: Harper & Row.

FISK, G. W., & WEST, D. J. (1956). ESP and mood. *Journal of the Society for Psychical Research*, **38**, 320–329.

FISK, G. W., & WEST, D. J. (1957). Towards accurate predictions from ESP data. *Journal of the Society for Psychical Research*, **39**, 157–164.

RYZL, M. (1966). A model of parapsychological communication. *Journal of Parapsychology*, **30**, 18–30.

SCHMEIDLER, G. R. (1964). An experiment in precognitive clairvoyance: Part II. The reliability of the scores. *Journal of Parapsychology*, **28**, 15–27.

STANFORD, R. (1978). Toward reinterpreting psi events. *Journal of the American Society for Psychical Research*, **72**, 197–214.

TAETZSCH, R. (1962). Design of a psi communication system. *International Journal of Parapsychology*, **4**, 35–66.

Institute for Parapsychology
Box 6847, College Station
Durham, NC 27708