Laboratory Psi Effects May Be Put to Practical Use: Two Pilot Studies

JAMES CARPENTER

Rhine Research Center
carpenterjim@verizon.net

Abstract—I describe two studies that were designed to illustrate the potential applicability of laboratory-derived ESP effects in trying to predict events of practical consequence in the "real world." Both studies attempt to predict the behavior of sets of financial indices over a designated week in the future. The studies followed up on earlier work on the prediction of scoring direction and scoring extremity. Participants are asked to make repeated calls at the same set of targets, and then their responses are combined through a majority-vote analysis to generate a set of "best predictions" to be tested against the actual yoked outcomes. The results of the first study were statistically significant and powerful enough in terms of amplification to have practical consequences. The second study was less effective, but changes in experimental conditions and participant population justified a post-hoc analysis based on the assumption of overall positive scoring. In this case, results were more encouraging. The Discussion addresses possible problems of data analysis and ethical concerns about the application of psi data.

Keywords: psi application—psi enhancement—psi and mood—variance effects

Introduction

Two things are commonly said about parapsychology. One is that laboratory experimental research deals with such weak effects that it cannot be very meaningful. The other is that psychic phenomena are so unpredictable that they cannot possibly have any practical utility. Are these things true? They can both be tested at once by examining efforts that have been made to apply laboratory parapsychological effects to the task of making real-life predictions, and by carrying out further research of the same sort.

Three basic ideas with several variations and elaborations have been proposed to assist in the effort to apply laboratory psi effects. One is to collect numerous guesses at targets and combine the guesses. The second is to sample a participant's guessing and use the success rate of the sample to try to predict

the success rate of the remainder of the guesses. These two ideas both represent what might be called a "bootstrap approach," in which some characteristics of psi responses are used to try to heighten the success rate of the same body of psi responses. The third idea is not a bootstrap approach: to try to establish relationships with independent, measurable variables that have the power to predict how well a participant is performing at ESP. All of these ideas seem intuitively appealing at first blush, and I discuss each of them in turn.

The idea of combining judgments of imperfect reliability in order to cancel out some of the unreliability and improve the correctness of the averaged judgments is a commonplace and is used routinely in psychology and other disciplines whenever measurements with imperfect reliability must be used. (See, for example, Schultheiss, Scott, & Schad, 2008, in the area of measuring implicit motives, or Granhag, 1997, with regard to the reliability of forensic testimony.) A single trial in an ESP test certainly represents a situation of imperfect reliability, even with the most talented performers on their best days. Inevitably, the idea has been applied to ESP data. There it has been referred to as the "repeated guessing technique" (Scott, 1960, Thouless, 1960), the "majority vote technique" (Fisk & West, 1957), and "redundancy" (Kennedy, 1979).

As an illustration of the potential power of this approach, consider the following facsimile situation using pseudo-ESP guesses. I begin with the assumption that I have relatively gifted participants who can perform somewhat reliably above chance. Let us assume that they can perform at a rate of correctness of 60% in a binary ESP test in which chance expectation is 50%. This is high performance by laboratory standards, but still not something to use as a basis for important decisions. In this illustration, I first generated a list of 10 targets that were to be repeatedly "called" made up of 1s and 0s. Then I "asked" the pseudo participants for 10 runs at this target list of 10 binary targets, and did this by making up a little deck of cards in which 6 cards were labeled "C" (for correct) and 4 cards were "I" (for incorrect), and shuffled this deck 10 times and laid out the calls. Each time a "C" appeared I substituted the actual target for that trial of 1 or 0, and each time an "I" appeared I wrote down the non-target. Then I added up the number of "guesses" that were given for 1 and the number for 0 for each of the 10 target positions. The results of this small sample are given in Table 1.

We can see there that from a modest 60% rate of success our majority decisions have risen quickly to a perfectly usable 100% rate of success.

Notice that there are extra columns in the table. This is because this exercise also illustrates another point: that such efficiency, if it could be obtained, could be used for actual communication. In constructing the target list I did not simply assemble an arbitrary list of 1s and 0s. Instead I elected to try to "transmit" a verbal target to my pseudo participants by using Morse code. The word "cats"

TABLE 1
Totals and Majorities of Votes for 10 Targets in a Pseudo-ESP Experiment

PSEUDO TARGET	VOTES FOR 1	VOTES FOR 0	MAJORITY
0	9	11	0
1	12	8	1
0	8	12	0
1	12	8	1
0	8	12	0
1	12	8	1
0	7	13	0
1	13	7	1
1	12	8	1
1	13	7	1
SUM CORRECT			10

in code is represented by dash-dot-dash-dot (C), dash-dot (A), dash (T), and dot-dot-dot (S). I then set "dash" as equivalent to 0 in the language of my ESP targets, and declared that "dot" was the same as 1. See Table 2 as an illustration of these data used as an attempt to retrieve the word.

The results in Table 2 show clearly that if a stable, above-chance rate of scoring can be assumed, then practically useful information can be obtained by combining repeated calls at targets. A modest 60% rate of scoring has been distilled to a satisfying 100% rate and the retrieval of a one-word verbal message.[1] However, a critical requirement is the above-chance rate of overall scoring. This was not appreciated in the first attempts to apply this multiple-calling, call-averaging technique. (Foster, 1943, was apparently the first researcher to report multiple-calling with his study on Plains Indians, but he did not average the calls. He would probably have had good results if he had.) Fisk and West (1956) applied the approach almost as an afterthought in a study

TABLE 2
Data from Pseudo Experiment as an Exercise in Retrieving a Word

TARGET LETTER	MORSE CODE	PSEUDO TARGET	VOTES FOR 1	VOTES FOR 0	MAJORITY	DECISION LETTER
C	DASH	0	9	11	0	C
	DOT	1	12	8	1	
	DASH	0	8	12	0	
	DOT	1	12	8	1	
A	DASH	0	8	12	0	A
	DOT	1	12	8	1	
T	DASH	0	7	13	0	T
S	DOT	1	13	7	1	S
	DOT	1	12	8	1	
	DOT	1	13	7	1	
SUM CORRECT					10	CATS

involving clock targets and mood ratings and got very encouraging results. Two more applications, however, resulted in null results (Michie & West, 1957) and significant psi-missing (Fisk & West, 1957). More recent applications of the basic approach with positive results have been reported by Brier and Tyminski (1970a, 1970b), Puthoff (1985), Puthoff, May, and Thomson (1986), and Radin (1991).

Dramatic results of real-world applications of ESP effects have been reported with the procedure of remote viewing, but almost all of these have appeared in popularly oriented books and periodicals and not in peer-reviewed journals, and details have tended to be scanty. Still, results have often appeared to be impressive, and they have ranged broadly in subject matter, including gathering military intelligence (e.g., McMoneagle, 2002, Targ & Puthoff, 1977), assisting police in solving crimes (Lyons & Truzzi, 1991), predicting silver futures (Harary & Targ, 1985, Targ, 1988), finding good real-estate opportunities (Kasian, 2004), and discovering lost archaeological sites (Schwartz, 2001). Perhaps because of the dramatic nature of the claimed results

and the paucity of details, these reports have sometimes spawned considerable controversy (e.g., Harary, 1992, Marks, 2002, May, 1998, Utts, 1996, Wiseman & Milton, 1998).

In any case, it is important to note that both the peer-reviewed reports such as Radin (1991) and the less detailed ones all rely upon participants who are being counted upon to give somewhat reliable extra-chance results. Thus they all began their averaging procedures on data that showed at least a small hitting tendency at the level of the single item (or in the case of Brier and Tyminski, also used below-chance data to deliberately amplify and transpose the negative effect). The moral holds: Sheer averaging of multiple calls will only serve to distill whatever scoring tendency is in the larger body of data. Overall scoring at a chance level will only result in more reliably chance-level scoring in the averages. A psi-missing trend in the overall data will yield a stronger rate of missing. Thus the bootstrap of repeated sampling is no panacea unless overall scoring rate can be reliably and independently predicted.

Index sampling is a technique developed to try to meet the need of directional prediction. Basically, this involves sampling some of the calling as it proceeds, scoring that, and using the scoring direction of that index sample to predict the scoring rate of the remainder of the participant's work. It was first used in a somewhat intuitive way by Cadoret (1955) and rendered more mathematically precise by Taetzsch (1958, 1962). Brier and Tyminski (1970a, 1970b), already cited, used not only repeated calls but index sampling in their application of ESP predictions to the real world of casino gambling. Results were significantly positive and apparently lucrative. Dean and Taetzsch (1970) reported a suggestively significant replication of the approach. The facts that casinos are still in business and little more research of this sort has been reported suggest that this bootstrap technique often fails as well—as reason says that it should. Index sampling by itself requires that another key assumption be met by the data to be sampled and averaged. This is the requirement that sets of the ESP data (like the runs in standard forced-choice testing) be internally consistent in scoring direction. As Schmeidler (1960) pointed out, and many others have observed, this assumption generally does not hold true for ESP data. Early attempts to find internal consistency, such as split-half reliability with ESP runs, have usually failed. Because of this, the scoring directions of index samples and remaining data are as often opposite as they are the same.

A bit of reflection will make it clear, however, that there is one way in which index sampling can be made to work as intended. Just as the overall mean-direction of a sample must be predicted for majority votes to be useful, another parameter of performance must be reliably predicted for index sampling to be useful. This is the variance of the performance around the chance expectation. If the deviation of a given set of guesses is relatively large,

then the scoring directions of any two parts of the whole set will tend to be the same. In the extreme case, if the whole run of calls is correct at 100% then the scoring direction of a sample will have to be 100%, which will match the 100% scoring rate of the remainder of the calls. Conversely, if the scoring deviation of the whole set is very small, at or very close to chance expectation, then the scoring deviations of any two parts of the whole set will tend to go in opposite directions. There is no magical bootstrap with index sampling, either; but with the assurance of large deviations from chance, or the ability to know when deviations will be large or small, it can be quite useful.

One demonstration of how stunningly well majority vote and index sampling can work was provided by Ryzl (1966). He used the work of a single hypnotically trained participant who repeatedly called the shuffled items of five lists of binary targets. The lists were coded representatives of randomly derived numbers, just as the ones and zeros in my pseudo experiment represented the letters of the word CATS. After sampling and averaging the calls, all five 10-item target lists were identified perfectly. Ryzl and his participant Pavel Stepanek achieved perfect "transmission" of the information in each of five series. It's worth remembering that Stepanek was perhaps the most reliable high scorer in the history of parapsychology. In this case, as in others, his runs of calls showed a persistent tendency to score above chance, and the run scores tended to be large deviations. He fulfilled the requirements of the techniques of majority vote and index sampling, and the flawless results demonstrated this fact.

This brings us to the third means that has been employed to assist in the problem of amplifying psi effects: finding ways to predict scoring trends (and scoring deviations), or ways to evoke them, so that redundancy, sampling, and averaging can be used reliably. Ryzl had at hand his Stepanek to work with and could safely predict that Stepanek would do as he had been doing before. With few Stepaneks around, parapsychologists have spent a great deal of effort in trying to define variables that will predict scoring, separating participants who will perform above chance from those who will perform below, and also trying to determine conditions that will reliably evoke scoring in either direction. Perhaps curiously, little of this effort has been applied to the problem of amplifying efficiency. The question with which I opened this discussion was: Can laboratory psi effects be put to practical use? The results of Ryzl and Stepanek suggest that the answer is yes. But having independent means of predicting or evoking reliable scoring directions and/or reliable scoring deviations is required. Presumably almost any independent predictor of these parameters of performance could be applied to this task.

One line of work aimed at exploring this was carried out by me some 20 years ago. I studied a set of mood adjectives that had been used in other research for describing momentary mood. Combining them with two other variables (a

sheep–goat attitude question and the California F-Scale), I derived a series of scales aimed at predicting run-score variance (sizes of deviations from chance) and hitting vs. missing (overall direction in scoring) in runs of forced-choice ESP testing with binary targets (Carpenter, 1968, 1969, 1983a, 1983b, 1991). The targets used were generally + and O. They were usually derived precognitively after all data were collected, but sometimes the targets were used clairvoyantly and coded to represent other information to be retrieved by all of the participants in that series acting in concert. In each series the participants guessed over and over at the same single list of targets without knowing it. This permitted my studies to test not only the efficiency of my predictive scales, but also my applications of the techniques of repeated guessing, index sampling, and majority votes. The content of the item collections used to predict hitting and variance evolved over the series, as larger and larger bodies of data were used to derive more reliable scales by stepwise multiple regression. I carried out 15 independent series with different groups of participants (usually university psychology students), and generally I met with some success. My final paper in this line of work reported three series in which I attempted to "transmit" by these means two words (represented by Morse code) and one set of octal digits picked by another experimenter and kept hidden from me. All three series showed statistically significant success, and clear amplification of efficiency with the repeated-sampling and averaging procedure. One of the three succeeded in retrieving the coded word PEACE with perfect accuracy.

Two New Studies

I have carried out two further studies along these lines that have not been reported. They were conducted several years ago, but never analyzed correctly until recently. Both of these studies employed revised mood scales that were generated from all previously collected data, including the last series of Carpenter (1991). One scale optimally postdicted scoring direction in the sample, and the other postdicted scoring extremity (run-score variance). In the two new studies these scales were used to predict targets unknown to me, as a demonstration of message-amplification principles to students in the Summer Study Program at the Foundation for Research on the Nature of Man (FRNM) in two consecutive years. Except for different targets, different data-collecting experimenters, and different participants, the two studies were identical. Richard Broughton served as co-experimenter in both studies, picking the targets (with the help of K. R. Rao in Study One), and, most importantly, writing a computer program that permitted an improved way of assigning and shuffling targets across runs and conducting scoring, sampling, and averaging procedures automatically. My previous series had used the same target lists repeatedly within each series, with targets in the same orders, so scoring required a great deal of work to correct

for the stacking effect caused by biases in calling patterns across participants. Broughton's program maintained the identities of targets within a list, shuffled them randomly for each run (avoiding the stacking problem), and carried out all other analyses automatically. Unfortunately, errors in using the program in the rush before scheduled class presentations led to false initial results in both cases. In Study One, extreme-quartile cut-off points intended for the two mood scales (explained in the Method section) were switched, resulting in the inappropriate inclusion and exclusion of much data. In Study Two, one of the predictors of hitting was inadvertently omitted. These problems were later realized, but the data lay in a filing cabinet for a long time before a period of leisure permitted them to be analyzed again, carefully and correctly. The results were interesting enough that I am reporting them now.

Method

These studies aimed to predict real-life events at a designated future time by the use of repeated calling at the same targets by sets of participants, and then analysis of their calls using their mood reports, their scores on the California F-scale, and their responses to the sheep–goat question. It was decided at the outset that the targets would represent the changes over a one-week period in the future of a set of twelve financial entities to be determined by an experimenter not otherwise involved in the procedure. An agreed-upon time was set that would permit the end of the week to coincide with the scheduled lecture at which the demonstration was to be described. The efforts of several parties are involved in this protocol, and after describing the materials used, I will spell out the procedures in terms of the parts played by each.

Materials

A packet of materials was given to each participant to use in self-testing to be done at home alone. The packet consisted of a page of instructions, a California F-scale, the sheep–goat question, and four calling sheets. Each calling sheet had 5 columns of 24 cells in which their guesses were to be recorded. The targets + and O were to be used. On the back of each calling sheet was a list of 57 mood adjectives, most of which were originally used to study the effects of stimulants and sedatives (Nowlis, 1961, 1965). Seven of these items were newly added for these studies for exploratory purposes, but were not used in the planned analyses. Participants were asked to respond to the questions in the packet, and then to pick four times when they could be alone for a few minutes. At these times they were to fill out each column on a given sheet with some order of +'s and O's that they felt would match the targets that would be picked later (no mention was made of coded predictions or repeated guessing). Then they

should immediately turn the sheet over and check the items in a way that would describe their mood at that moment. If an item was left unchecked, that meant it did not at all describe their mood, one check meant that it described their mood somewhat, and two checks meant that it described their mood strongly. The mood adjectives are given in Appendix A.

Based on findings from previous studies, participants were to be divided into two groups in terms of their F-scale scores. Using norms that I have carried over in this research program, those with scores of −31 or lower were held and used in further analyses, others were excluded.[2]

The collections of mood items used as predictors in these studies were as follows: Direction of scoring was predicted by combining responses positively to *amiable*, *fearless, masterful*, *retiring*, and by a *yes* answer to the sheep–goat question; and it was predicted negatively by *adaptable*, and a *no* answer to the sheep–goat question. Extremity of scoring was predicted positively by *close-mouthed* and negatively by *detached* and *witty*.[3]

Experimenter Soliciting Participants

I played this role in Study One and Kathy Dalton did it in Study Two. This person went to some group of potential participants and gave a very brief talk on ESP research and then described the procedure in which people were asked to participate. They were told that they would be asked to try to use ESP to predict targets that would be picked in the future, after all the guesses had been collected, and that this was something like the kinds of predictions that people tried to make in games of chance or gambling. They would also be asked to respond to a questionnaire on some general attitudes, and to check off words to describe their moods at the times in which they did the testing. A date was set for return of the materials, usually about two weeks hence. No payment was offered for participation.

Experimenter Picking Targets

Richard Broughton played this role in both studies, with the help of K. R. Rao in Study One. He picked 12 financial entities the one-week changes in which would generate the targets for the study. This list was kept secret. When the day arrived for the target week to begin, Broughton consulted *The Wall Street Journal* for the baseline values of the entities and recorded them. On the last day of the week, he recorded the values of the same entities and noted the direction of change over the week for each entity. These *gain* or *loss* targets were coded as + and O, respectively, for the ESP test. Then when the predictions generated by the analysis of the guesses were unveiled, he unveiled these targets as well for a check on how the procedure had fared.

Experimenter Analyzing the Calls

After collecting the last of the data at least two weeks prior to the target week, I keyed (or got help in keying) all of the participants' guesses into a spreadsheet along with the participants' F-scale scores, sheep–goat response, the page number and run number of each run, and all mood-item responses. Then I used some high-temperature numbers from that day's local newspaper to pick an entry point into a table of random numbers from which I selected 12 digits. I converted the digits to + if the digit was odd, and O if it was even. These digits were to be used as index targets for the series, to be used to help predict the actual content of the 12 precognitive targets. All of the participants' responses along with the index targets were entered into Broughton's scoring program.

This program maintained the identity of each of the 24 targets used in each run (12 predetermined index targets and 12 unknown precognitive targets) and randomly shuffled the 24 positions anew for each run of calls using a software pseudo-random function. Thus "Index Target One" or "Precognitive Target Five" kept their identities across runs, even though they appeared in different actual run positions in different runs. Then the program scored the mood scales for each page of runs. It excluded the data of high–F-scale participants from further analysis and retained only the low-F cases. It tabulated the mood scale scores for each page of runs and printed them out for the experimenter who then calculated the nearest quartile cut-off points in each scale (only extreme-quartile scores on the scales were used as generating a prediction of direction or extremity. Mid-range scores were omitted from further analysis). These quartile cut-off points were entered back into the program which then used the mood scales to segregate the data into subgroups for two repeated-guessing analyses—one based on directional predictions and the other based on extremity predictions. The logic used was as follows.

Consider first the simpler case of using directional predictions. If a mood-scale score gave a prediction of psi-hitting for its page of ESP runs, then all the calls made to the 12 precognitive targets were tallied as they stood. If the mood scale gave a psi-missing prediction, then all calls were reversed (+ calls became O's, and vice versa) and these reversed calls were tallied. Then all guesses across all pages that had yielded the mood predictions were tallied together for a final set of "votes" for + and for O.

In the case of extremity predictions, the 12 index targets were used in an intermediate step. All index calls in all runs were scored against the index targets. Then in the case of a large-RSV prediction for the page, runs in which index calls scored above chance (7 or more) generated a psi-hitting prediction for the rest of the trials in the run, so the calls on the remaining 12 precognitive targets were entered as they were into a tally. On the other hand, if the index

scoring was below chance (5 or fewer), precognitive calls were reversed and then entered into the tally. Index scores exactly at chance with 6 hits yielded no prediction, and the calls on the precognitive targets were omitted for that run. When the mood scale for the page predicted small RSV, a procedure opposite to that used with large RSV prediction was carried out. Above-chance index scores generated a prediction of psi-missing for the other trials of the run, so they were reversed and tallied, while below-chance index scores generated a psi-hitting prediction for the rest of the run, so those precognitive calls were tallied as they were. Both of these analyses (one using directional mood-scale predictions and the other using extremity predictions) were carried out for all usable data. Then the two sets of tallies were themselves combined for a final set of 12 best guesses as to the identity of the precognitive targets.

At the end of the one-week target period, the target identities predicted by the participants' efforts are revealed and matched against the actual targets determined by the week's financial activity, and a number correct and incorrect can be determined.

Study One

Participants

Volunteers were solicited from Summer Study students and from members of two meditation classes and one yoga class being taught in the community whose instructors were interested in parapsychology. I chose these groups because I expected that they might contain a relatively large proportion of persons with low-authoritarian attitudes. The cut-off points that had separated median groups of university students 10 years earlier (and which I chose to continue to use) had come to select smaller groups of participants over time as attitudes of university students apparently drifted in a more authoritarian direction. This made it inefficient to use this unselected student population. A small sample of 58 participants agreed to take part, of which 25 met the low-F criterion. It was not expected that such a small sample would yield very reliable results, but time did not permit soliciting more participants. Of the 25 low-F participants, 19 were female. The low-F group ranged in age from 18 to 52, with a median age of 21. These low-F participants contributed 495 runs.

Soliciting Experimenter

I played this role in Study One, and maintained what had become my normal routine: a very brief talk about the meaning of ESP testing, a dispassionate statement that the questions involved were interesting but still quite mysterious to science, and a courteous request for their help. My attitude was deliberately neutral and routine, as experimenters typically behave in other psychological

research in which they hope to make their own contributions to the situations as standardized and neutral as possible.

Targets

The one-week behavior of a set of market values and industry group comparisons was chosen to determine the targets. Six were simple comparisons from the beginning day to the last. A rise in value was called a +, a decline was an O. The other six were proportional measures. These were chosen in case a general drift in the market over the week could cause too many positive correlations in behavior and a disproportionate number of targets of the same type. The value of a pair of industry groups was compared at the beginning of the week and then again at the ending day. If the relative value of one over the other was maintained in direction over the week, the trial was called +. If the advantage between them switched to the other over the week, the target was an O. The specific target indices and their actual values and the targets determined are given in Table 3.

Results

The final tally of both repeated-guessing analyses of the low-F data was rather successful. Eleven of the twelve items were predicted correctly. See Table 4.

The overall data of low-F participants scored at a very slightly below-chance rate of 49.9%, which would have made a simple majority vote procedure with these data a waste of time. The votes cast by the two analyses combined were correct at a higher rate of 51.7%, and the decisions made by the votes were more efficient still at 92% correct ($\chi^2 = 8.33$, one degree of freedom, $p = .004$). Only the last item (the behavior of banks central vs. banks west over the period) was called incorrectly. Had actual investment decisions been made at the beginning of the week based on these predictions, the outcome would have been positive for the investor. Of the two mood scales used to generate predictions, the one for scoring direction contributed much more. It correlated positively (but not significantly) with the ESP scores: $r = .15$, $p = .08$. Statistical significance is not always required for practical utility. The extremity scale gave a correlation with run-score variance that was very slightly in the wrong direction: $r = -.05$.

Study Two

Participants

Participants for this study were drawn from classes in acting and creative writing at the University of North Carolina Chapel Hill. This population was

TABLE 3
Economic Indices Used as Precognitive Targets for Study One

Trial	Item	Beg. Value	End Value	Change	Target
	Markets				
1	Price of Gold on CMX (spot)	369.7	364.4	Declined	0
2	Oil, West Texas Int. Crude	22.15	21.5	Declined	0
3	Sugar cane raw, World	.0956	.0989	Increased	+
4	Dow Jones Industrial Avg.	2978.76	2966.23	Declined	0
5	Dollar vs. Yen (preced. day)	.07301	.007273	Declined	0
6	Bond Yield – Corp. Master	510.41	513.03	Increased	+
	Industry Group Comparisons				
7	Banks (money centers) vs. Casinos	B: 195.22 C: 596.95	B: 192.93 C: 568.59	B declined less	+
8	Industrial Tech. vs. Aluminum	I: 290.68 A: 322.23	I: 291.83 A: 322.86	IT gained more	+
9	Energy vs. Financial	E: 281.11 F: 301.77	E: 276.34 F: 297.93	Fin. declined less	0
10	Building Materials vs. Railroads	B: 404.02 R: 371.91	B: 398.15 R: 381.18	R increased, BM decreased	0
11	Office Equipment vs. Diversified Technology	O: 265.99 D: 269.75	O: 262.23 263.75	OE declined less	+
12	Banks Central vs. Banks West	BC: 499.99 BW: 299.36	BC: 493.7 BW: 287.92	BC declined less	+

On **Markets**, if value was higher at the end than at the beginning of the period, the target was coded as 1 (or +), if lower at the end it was coded as 0 (or−). For **Industry Group Comparisons**, if the first item did better relative to the second over the week (improved more or declined less), the target was coded as + (or 1); if the second did better, the target was coded as 0 (or−).

chosen in part to assure a higher proportion of low-authoritarian participants, as in Study One, but also because we thought it would be of interest to see how a group expected to be more psi-productive might respond to this procedure. Previous research had strongly suggested that more creative persons are especially likely to demonstrate psi effects (e.g., Anderson, 1966, Moon, 1974, Moss, 1969). I reasoned that their mood reports might discriminate their scoring patterns especially effectively. The sample for this study was smaller even than the one before: 47 volunteers, of whom 22 were low-F. They contributed 440

TABLE 4
Tallied Votes and Prediction-Decisions in Study One

Cell	Votes +	Votes O	Decision	Target
1	193	<u>208</u>	O	O
2	194	<u>207</u>	O	O
3	<u>206</u>	195	+	+
4	194	<u>207</u>	O	O
5	196	<u>205</u>	O	O
6	<u>203</u>	198	+	+
7	<u>215</u>	186	+	+
8	<u>226</u>	175	+	+
9	194	<u>207</u>	O	O
10	200	<u>201</u>	O	O
11	<u>212</u>	189	+	+
12	<u>193</u>	208	O	+

2,490 votes correct, 2,322 incorrect: 51.7% correct.
Underlining, total votes associated with actual target.
Gray shading, correct decisions. 11 decisions correct 1 incorrect: 92% correct.

runs. Those low-F participants ranged in age from 18 to 58, with a median age of 22.5. Thirteen were female. Time constraints again limited solicitation of participants.

Soliciting Experimenter

Kathy Dalton played this role in Study Two. Although I instructed her briefly in my normal approach to soliciting participants in class groups, discussion with her later made it clear that she also gave some room in the situation to more expression of her own personality. She made it a point to

TABLE 5

Economic Indices Used as Precognitive Targets for Study Two

Trial	Item	Beg. Value	End Value	Change	Target
	Markets				
1	Price of Gold on CMX (spot)	339.30	336.10	Declined	0
2	Oil, West Texas Int. Crude	19.90	20.75	Increased	+
3	Sugar cane raw, World	.0956	.0985	Increased	+
4	Dow Jones Industrial Avg.	3338.77	3333.18	Declined	0
5	Dollar vs. Yen (preced. day)	.007454	.007491	Increased	+
6	Bond Yield – Corp. Master	568.21	568.70	Increased	+
	Industry Group Comparisons				
7	Banks (money centers) vs. Casinos	B: 239.79 C: 650.17	B: 237.41 C: 674.10	Casinos increased	0
8	Industrial Tech. vs. Aluminum	I: 316.32 A: 316.82	I: 322.42 A: 320.08	IT gained more	+
9	Energy vs. Financial	E: 250.13 F: 346.98	E: 260.36 F: 346.00	Energy up, Financial down	+
10	Building Materials vs. Railroads	B: 467.39 R: 476.96	B: 477.32 R: 501.67	Rail did better	0
11	Office Equipment vs. Diversified Technology	O: 307.72 D: 287.98	O: 306.81 D: 282.43	OE declined less	+
12	Banks Central vs. Banks West	BC: 624.5 BW: 384.4	BC: 623.52 BW: 390.67	BW up BC down	+

On **Markets**, if value was higher at the end than at the beginning of the period, the target was coded as 1 (or +), if lower at the end it was coded as 0 (or−). For **Industry Group Comparisons**, if the first item did better relative to the second over the week (improved more or declined less), the target was coded as + (or 1); if the second did better, the target was coded as 0 (or−).

clearly express her interest and enthusiasm and lively faith in the creative and intuitive abilities of the students. She followed the letter of the procedure, but added more lively spirit.

Targets

I asked Broughton to select targets as before, picking some financial indices whose behavior over a one-week period would generate 12 precognitive binary targets. We agreed upon the target week, again a time that would conclude on a

day in which I would be giving a lecture on this subject at the FRNM. Without telling me so, he elected to use the same indices as the year before: six changes in markets and six industry group comparisons. Changes would generate + and O targets as before. See Table 5 for the targets assigned along with the actual financial values recorded at the beginning and end of the target week.

Results

Overall results were unusually positive for this series. In fact, this is the first case among the 17 series conducted in this program of repeated-guessing work, that significant psi-hitting was observed overall. Scoring on the precognitive targets overall (irrespective of F-scale or mood scores) was 5,661 hits where 5,520 were expected by chance: $z = 2.42$, $p =$. 016. On the other hand, in terms of the analyses of most interest, in contrast to Study One the decisions generated by the repeated-guessing procedures were not particularly successful. See Table 6.

TABLE 6
Tallied Votes and Prediction-Decisions in Study Two

Cell	Votes +	Votes O	Decision	Target
1	168	147	+	O
2	153	162	O	+
3	169	146	+	+
4	160	155	+	O
5	167	148	+	+
6	165	150	+	+
7	161	154	+	O
8	157	158	O	+
9	160	155	+	+
10	145	170	O	O
11	164	151	+	+
12	150	165	O	O

1,926 votes correct, 1,854 incorrect: 51% correct. 7 decisions correct, 5 incorrect: 58% correct.
Underlining, total votes associated with actual target.
Gray shading, correct decisions.

The unit majority votes were correct at only a 51% rate, barely above-chance expectation, and only seven majority decisions were correct, with five incorrect ($\chi 2$ = .33, one degree of freedom, p = .56). Thus, the results are on the expected side, but only weakly so. Practical investment decisions made on the basis of these predictions would have beaten sheer chance, but not by very much.

This result would be expected by the rather poor performance of the mood-scale predictors, along with the small number of cases. The cluster intended to predict scoring direction yielded a relationship almost exactly equal to chance: r = .01. This time the cluster predicting extremity performed better, but not significantly: r = .11, p = .155.

An Exploratory Analysis

Overall, psi-hitting was not expected in this study, but I decided to carry out an exploratory majority-vote analysis as if it had been. For this analysis, all

TABLE 7
Study Two Analysis Using No Scales (Simple Majority Vote)

Cell	Votes +	Votes –	Decision	Target
1	427	497	O	O
2	460	464	O	+
3	457	467	O	+
4	458	466	O	O
5	478	446	+	+
6	471	453	+	+
7	456	468	O	O
8	493	431	+	+
9	468	456	+	+
10	453	471	O	O
11	477	447	+	+
12	446	478	O	O

5,684 votes correct, 5,404 incorrect: 51.3% correct. 10 decisions correct, 2 incorrect: 83.3% correct.
Underlining, total votes associated with actual target.
Gray shading, correction decisions.

data across all participants are pooled and tallied into simple votes for + and O for each target position. The results are given in Table 7.

As might be expected from the overall psi-hitting in the data, and with the much larger number of votes, this analysis is more efficient than the last one. The unit majority votes rose in accuracy to 51.3%, and 10 of the 12 majority decisions were correct: $\chi^2 = 5.33$, one degree of freedom, $p = .021$. With an accuracy rate of 83.3%, our investor would be doing better in this case.

Discussion

Some Possible Problems

One problem that bears mentioning hinges on the fact that with almost any "real-world" set of targets, randomness cannot be assumed. For example, there are times when financial entities such as the ones studied here drift up or down in a correlated manner. This has a bearing mainly on simple majority-vote analyses such as the last unplanned analysis done in Study Two. It may be highly unlikely, but it is not inconceivable that, given a high degree of correlation among the targets, some similar excess of calls that happened to be given by participants in the weeks before the target period could cause a spurious relationship between majorities and targets. If most targets happened to be + because of increased value, and most calls happened to be "+" because, say, of some period of exuberance during the testing, then an excess of hits would appear that would not represent ESP. This does not seem to have happened in the case of this particular analysis of Study Two. Only the first six targets are involved in the question, since the last six were determined by relative comparisons, precisely in order to avoid the problem of correlated performance. Among these six targets there may have been some degree of correlation, in that 4 were "+" and 2 were "O." The majorities reached from the participants' calls tended in the opposite direction: 4 "O" and 2 "+." The fact that there were still four hits among the trials was in spite of the two contrary tendencies, not because of them.

While randomness of targets is an understood prerequisite for parapsychological research in general, it is because of the need to be assured that the statistical evaluation of results is appropriate. In the context of wishing to predict real information, statistical evaluation is less important than pragmatic accuracy. Even if the exactly best statistical model for assessing significance cannot be known because of target non-randomness, results may still be practically useful.

Still, the non-randomness of targets does add an unnecessary difficulty in interpreting results, so future studies planning a simple majority vote should eliminate it. This could be done simply on the part of the experimenter who

picks targets by his or her adding one additional step to the procedure. After picking the targets, and observing the actual behavior that was being predicted, before submitting to the second experimenter for scoring, one more pass could be made on the targets by randomly switching about half of them to their opposites. It would be understood at the outset that the "correct target" would be this final, randomly coded set. Participants appear to use their ESP to reach the correct target, whatever it is, without regard to such contingencies, so there seems to be no reason to expect that any problem would arise from adding this step.

The objection might also be raised that the choice of the responses "+" and "O" are too transparently linked to the idea of ups and downs of market performance, and somehow lead to spurious relationships. It is difficult to imagine why this might be so. In this design, participants never know that their responses are yoked to any future outcomes of any sort, only that targets will somehow be selected and they are trying to predict them. In any case, with regard to the primary analyses of these studies involving independent predictors of performance, and not simple majority votes, the manipulations of calls prior to tallying majorities results in about half of the calls being actually rendered into their opposite content before they are used.

Some Implications

These two small studies offer some support for the idea that even relatively weak laboratory psi effects can potentially be put to practical use in predicting unknown future events. The scales of mood items and attitude items used here have shown modest reliability over a number of studies. The findings also underscore the reality that procedures that depend upon such relatively small effects may not be counted upon to always work, particularly when small sample sizes are employed. However, this report is primarily a demonstration of a principle, and it is important to note that such mood items are not the only predictors that could be used in such a way, and they are probably far from the strongest that we might find to use. As they are, they make some psychological sense. Low-authoritarian persons have been found to give more valid self-reports in other settings, as mentioned above. Persons who believe that ESP is possible in the conditions of the study have often been found to score better than those who do not believe that. A factor analysis of the mood items was carried out by Carpenter (1991) and showed that items predicting hitting tended to represent factors of detached relaxation and inward focus, freedom from self-doubt and cognitive analysis, and freedom from anxiety—all things that have been found throughout our literature to effect psi performance (e.g., Carpenter, 2004, Palmer, 1978, 1982, Schmeidler, 1988). Items predicting extremity tended to represent factors that implied a non-analytical and holistic state of

mind along with a freedom from distraction and cognitive work. Such things have been proposed by Carpenter (2004, 2005, 2008) to facilitate a singularity of unconscious intention that is theorized to increase scoring extremity.

Whatever their meaningfulness and predictive power, however, these means are certainly not the only ones for making use of redundant psi data. And they are not the simplest and most direct. The secondary analysis of Study Two illustrates that nicely. Some experimenters (and perhaps some experimental approaches) appear to be more psi-facilitative than others (e.g., Wiseman & Schlitz, 1997), and Kathy Dalton has gone on from her work in this study to do other work that suggests she can be one of those inspiring experimenters, at least with artistic participants (Dalton, 1997, Morris, Dalton, Delanoy & Watt, 1995). At the time this study was done, there was already evidence that persons engaged effectively in creative work can be counted on to score above chance in ESP tests with some reliability, and the ensuing years have also added to that evidence (Dalton, 1997, Morris, Cunningham, McAlpine, & Taylor, 1998, Morris, Summers, & Yim, 2003, Moss, 1969, Schlitz & Honorton, 1992). Thus, while I did not predict above-chance overall scoring in Study Two, I certainly might have done so legitimately. If I had, a straightforward confirmation would have come forth. We should make such direct predictions of scoring when we have reason to.

The main point I wish to make is that any means of predicting scoring direction could be put to work in a majority-vote paradigm. The targets given to participants can be yoked to some "real-world" event which we wish to predict. The yoking seems to present no obstacle to persons demonstrating their psi apprehensions as they normally do, all in the context of our predictors. Similarly, any means at hand of predicting scoring extremity reliably can be put to work as well and used to interpret the scoring implications of index sampling. Will extraversion as measured on the Eysenck Personality Inventory reliably predict nonverbal ESP performance (Roe, Henderson, & Matthews, 2008), or self-rated luckiness reliably predict psi-hitting (Luke, Delanoy, & Sherwood, 2008)? Will self-rated openness to experience show a robust relationship with preference choices linked covertly with psi targets (Luke, Roe, & Davison, 2008)? All of these possibilities are drawn from the most recent journals I have at hand. There are many others. Some will prove to be more robustly reliable than others, and they can all be put to work in the practical ways illustrated in these two studies.

Should psi effects be applied? This is like asking if we wish to have more powerful access to knowledge by any means at all. Should Galileo have figured out the basic equations of refraction and developed the telescope? Should Pascal have built a mechanical calculator? Should the Wright brothers have worked out some basic principles of aerodynamics? In fact, we generally desire new access to knowledge and fear it at once. More knowledge is more power,

and we wonder if our wisdom and humanity will be equal to the challenges of more power. In any case, scientifically derived parapsychological effects await application and will probably be put to work.

Ideas such as "psychic development" and "psychic application" currently tend to be left mostly to practitioners who teach self-development techniques in "mind control" or "intuition", with dubious results. Even if such approaches have some success, they are rather analogous to trying to see greater distances by vision training, or to increase computational power by teaching arithmetical shortcuts, or to solve the problem of human flight by developing the techniques of training high jumpers. It is scientific work that has made the enormous leaps in our capacity in these areas, and it will be scientific work that eventually leads to the reliable application of psi.

Notes

[1] This is a straightforward application of the Law of Large Numbers (Feller, 1968), which holds that the estimate of a population value drawn from averaging samples will more closely approach the true value as the number of samples increases.

[2] In earlier studies I had used the F-scale as a moderating variable, on the assumption that persons lower in authoritarian attitudes were more reliable reporters of their own internal states (Barron, 1953, Scodel & Mussen, 1953, Thayer, 1971) and hence should give more valid mood reports. Lower-F participants were found to give more useful data in my previous ESP studies.

[3] Since these clusters are derived by stepwise multiple regression, which selects for orthogonal contributions of items to a prediction, no unifying conceptual themes would be expected among the items.

References

Anderson, M. (1966). The use of fantasy in testing for extrasensory perception. *Journal of the American Society for Psychical Research, 60,* 150–163.

Barron, F. (1953). Some personality correlates of independence of judgment. *Journal of Personality, 21,* 287–297.

Brier, R. M., & Tyminski, W. V. (1970a). Psi Application: Part I. A preliminary attempt. *Journal of Parapsychology, 34,* 1–25.

Brier, R. M., & Tyminski, W. V. (1970b). Psi Application: Part II. The majority vote technique. *Journal of Parapsychology, 34,* 26–36.

Cadoret, R. J. (1955). The reliable application of ESP. *Journal of Parapsychology, 19,* 203–227.

Carpenter, J. C. (1968). Two related studies on mood and precognition run-score variance. *Journal of Parapsychology, 32,* 75–89.

Carpenter, J. C. (1969). Further study on mood and precognition run-score variance. *Journal of Parapsychology, 33,* 48–56.

Carpenter, J. C. (1983a). Prediction of forced-choice ESP performance: Part I. A mood-adjective scale for predicting the variance of ESP run scores. *Journal of Parapsychology, 47,* 191–216.

Carpenter, J. C. (1983b). Prediction of forced-choice ESP performance: Part II. Application of a mood scale to a repeated guessing technique. *Journal of Parapsychology, 47,* 217–236.

Carpenter, J. C. (1991). Prediction of forced-choice ESP performance: Part III. Three attempts to retrieve coded information using mood reports and repeated-guessing technique. *Journal of Parapsychology, 55,* 227–280.

Carpenter, J. C. (2004). First Sight: Part One: A model of psi and the mind. *Journal of Parapsychology, 68,* 217–254.

Carpenter, J. C. (2005). First Sight: Part Two: Elaborations of a model of psi and the mind. *Journal of Parapsychology, 69,* 63–112.

Carpenter, J. C. (2008). Relations between ESP and memory in light of the first sight model of psi. *Journal of Parapsychology, 72,* 47–76.

Dalton, K. (1997). Exploring the links: Creativity and psi in the ganzfeld. *Proceedings of Presented Papers: The Parapsychological Association 40th Annual Convention,* 119–134.

Dean, D., & Taetszch, R. (1970). Psi in the casino: Taetzsch method. *Proceedings of the Parapsychological Association, 7,* 14–15.

Feller, W. (1968). Laws of Large Numbers. In: *An Introduction to Probability Theory and Its Applications, Vol. 1, 3rd ed.* (Chapter 10, pp. 228–247). New York: Wiley.

Fisk, G. W., & West, D. J. (1956). ESP and mood: Report of a "mass" experiment. *Journal of the Society for Psychical Research, 38,* 1–7.

Fisk, G. W., & West, D. J. (1957). Towards accurate predictions from ESP data. *Journal of the Society for Psychical Research, 39,* 157–162.

Foster, A. A. (1943). ESP tests with American Indian children. *Journal of Parapsychology, 7,* 94–103.

Granhag, P. A. (1997). Realism in eyewitness confidence as a function of type of event witnessed and repeated recall. *Journal of Applied Psychology, 82,* 599–613.

Harary, K. (1992). The goose that laid the silver eggs: A criticism of psi and silver futures forecasting. *The Journal of the American Society for Psychical Research, 86,* 375–410.

Harary, K., & Targ, R. (1985). A new approach to forecasting commodity futures. *Psi Research, 4,* 79–88.

Kasian, S. J. (2004). Remote viewing of real estate: Entrepreneurial explorations. *Journal of Indian Psychology, 22,* 34–43.

Kennedy, J. E. (1979). Redundancy in psi information. *Journal of Parapsychology, 43,* 290–314.

Luke, D. P., Delanoy, D., & Sherwood, S. (2008). Psi may look like luck: Perceived luckiness and beliefs about luck in relation to precognition. *Journal of the Society for Psychical Research, 72,* 193–207.

Luke, D. P., Roe, C. A., & Davison, J. (2008). Testing for forced-choice precognition using a hidden task: Two replications. *Journal of Parapsychology,72,* 133–154.

Lyons, A., & Truzzi, M. (1991). *The Blue Sense: Psychic Detectives and Crime.* New York: Warner Books.

Marks, D. (2002). *The Psychology of the Psychic.* New York: Prometheus Books.

May, E. C. (1998). Response to "Experiment One of the SAIC remote viewing program: A critical re-evaluation." *Journal of Parapsychology, 62,* 309–318.

McMoneagle, J. (2002). *The Stargate Chronicles: Memoirs of a Psychic Spy.* Charlottesville, VA: Hampton Roads Publishing.

Michie, D., & West, D. J. (1957). A mass ESP test using television. *Journal of the Society for Psychical Research, 39,* 113–133.

Moon, M. (1974). *Extrasensory Perception and Art Experience.* [Unpublished doctoral dissertation]. University of Pennsylvania.

Morris, R. L., Cunningham, S., McAlpine, S., & Taylor, R. K. (1998). Toward replications and extension of autoganzfeld results [abstract]. In *Research in Parapsychology 1993,* edited by N. L. Zingrone, M. Schlitz, C. S. Alvarado, & J. Milton. Lanham, MD: Scarecrow Press. p. 308.

Morris, R., Dalton, K., Delanoy, D., & Watt, C. (1995). Comparison of the sender/no sender condition in the ganzfeld. *Proceedings of Presented Papers, The Parapsychological*

Association 38th Annual Convention (pp. 244–259). [Convention in Durham, NC.]

Morris, R. W., Summers, J., & Yim, S. (2003). Evidence of anomalous information transfer with a creative population in ganzfeld stimulation. *Journal of Parapsychology, 67,* 256–257 [abstract].

Moss, T. (1969). ESP effects in "artists" compared with "non-artists." *Journal of Parapsychology, 33,* 57–69.

Nowlis, V. (1961). Methods for studying mood changes produced by drugs. *Revue de Psychologie Applique, 11,* 373–386.

Nowlis, V. (1965). Research with the mood-adjective check list. In: S. S. Tomkins & C. E. Izard (Eds.), *Affect, Cognition and Personality* (pp. 352–389). New York: Springer Publishing.

Palmer, J. (1978). Extrasensory perception: Research findings. In: S. Krippner (Ed.), *Advances in Parapsychological Research 2* (pp. 59–243). New York: Plenum Press.

Palmer, J. (1982). ESP research findings: 1976–1978. In: S. Krippner (Ed.), *Advances in Parapsychological Research 3* (pp. 41–82). New York: Plenum Press.

Puthoff, H. E. (1985). Calculator assisted psi amplification. In: R. A. White & J. (Eds.), *Research in Parapsychology 1984* (pp. 48–51). Metuchen, NJ: Scarecrow Press.

Puthoff, H. E., May, E. C., & Thomson, M. J. (1986). Calculator assisted psi amplification II: Use of the sequential-sampling technique as a variable-length majority-vote code. In: D. H. Weiner & D. I. Radin (Eds.), *Research in Parapsychology 1985* (pp. 73–77). Metuchen, NJ: Scarecrow Press.

Radin, D. I. (1991). Enhancing effects in psi experiments with sequential analysis: A replication and extension. In: L. A. Henkel & G. R. Schmeidler (Eds.), *Research in Parapsychology 1990* (pp. 21–25). Metuchen, NJ: Scarecrow Press.

Roe, C. A., Henderson, S. J., & Matthews, J. (2008). Extraversion and performance at a forced-choice ESP task with verbal stimuli: Two studies. *Journal of the Society for Psychical Research, 72,* 208–221.

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

Schlitz, M. J., & Honorton, C. (1992). Ganzfeld ESP performance within an artistically gifted population. *Journal of the American Society for Psychical Research*, 86, 83–98.

Schmeidler, G. R. (1960). The accuracy of parapsychological information. *Indian Journal of Parapsychology, 2,* 169–173.

Schmeidler, G. R. (1988). *Parapsychology and Psychology: Matches and Mismatches.* Jefferson, NC: McFarland.

Schultheiss, O. C., Scott, S. H., & Schad, D. (2008). The reliability of a Picture Story Exercise measure of implicit motives: Estimates of internal consistency, retest reliability, and ipsative stability. *Journal of Research in Personality, 42,* 1560–1571.

Schwartz, S. (2001). *The Alexandria Project.* Bloomington, IN: iUniverse.

Scodel, A., & Mussen, P. (1953). Social perceptions of authoritarians and nonauthoritarians. *Journal of Abnormal and Social Psychology, 48,* 181–184.

Scott, C. (1960). An appendix to "The repeated guessing technique." *International Journal of Parapsychology, 2,* 37–46.

Taetzsch, R. (1958). Application of statistical quality control techniques to statistical psi control problems (Abstract). *Journal of Parapsychology, 22,* 304.

Taetzsch, R. (1962). Design of a psi communication system. *International Journal of Parapsychology, 4,* 35–70.

Targ, R. E. (1988). ESP on Wall Street. *The Explorer, 4*(2), 1–2. Society for Scientific Exploration.

Targ, R. E., & Puthoff, H. (1977). *Mind Reach: Scientists Look at Psychic Abilities.* New York: Delacourt.

Thayer, R. E. (1971). Personality and discrepancies between verbal reports and physiological measures of private emotional experiences. *Journal of Personality, 39,* 57–69.

Thouless, R. H. (1960). The repeated guessing technique. *International Journal of Parapsychology, 2,* 21–36.

Utts, J. (1996). An assessment of the evidence for psychic functioning. *Journal of Scientific Exploration, 10,* 3–30.

Wiseman, R., & Milton, J. (1998). Experiment One of the SAIC remote viewing program: A critical re-evaluation. *Journal of Parapsychology, 62,* 297–308.

Wiseman, R., & Schlitz, M. (1997). Experimenter effects and the remote detection of staring. *Journal of Parapsychology, 61,* 197–208.