Journal of Parapsychology, Vol. 53, December 1989

"FUTURE TELLING": A META-ANALYSIS OF FORCED-CHOICE PRECOGNITION EXPERIMENTS, 1935–1987

BY CHARLES HONORTON AND DIANE C. FERRARI

ABSTRACT: We report a meta-analysis of forced-choice precognition experiments published in the English-language parapsychological literature between 1935 and 1987. These studies involve attempts by subjects to predict the identity of target stimuli selected randomly over intervals ranging from several hundred milliseconds to one year following the subjects' responses. We retrieved 309 studies reported by 62 investigators. Nearly two million individual trials were contributed by more than 50,000 subjects. Study outcomes are assessed by overall level of statistical significance and effect size. There is a small, but reliable overall effect ($z = 11.41$, $p = 6.3 \times 10^{-25}$). Thirty percent of the studies (by 40 investigators) are significant at the 5% significance level. Assessment of vulnerability to selective reporting indicates that a ratio of 46 unreported studies averaging null results would be required for each reported study in order to reduce the overall result to nonsignificance. No systematic relationship was found between study outcomes and eight indices of research quality. Effect size has remained essentially constant over the survey period, whereas research quality has improved substantially. Four moderating variables appear to covary significantly with study outcome: Studies using subjects selected on the basis of prior testing performance show significantly larger effects than studies using unselected subjects. Subjects tested individually by an experimenter show significantly larger effects than those tested in groups. Studies in which subjects are given trial-by-trial or run-score feedback have significantly larger effects than those with delayed or no subject feedback. Studies with brief intervals between subjects' responses and target generation show significantly stronger effects than studies involving longer intervals. The combined impact of these moderating variables appears to be very strong. Independently significant outcomes are observed in seven of the eight studies using selected subjects, who were tested individually and received trial-by-trial feedback.

Precognition refers to the noninferential prediction of future events. Anecdotal claims of "future telling" have occurred throughout human history in virtually every culture and period. Today such

This work was funded by SRI International and the John E. Fetzer Foundation. We wish to thank our PRL colleague George P. Hansen, who is primarily responsible for retrieving the studies used in the meta-analysis. We are grateful to Edwin C. May, Jessica Utts, and to five anonymous reviewers at SRI for valuable comments on an earlier draft of this report. Valuable comments were also made by Ephraim Schechter and by three anonymous referees. The division of authorship responsibility is as follows: Honorton is responsible for the design of the meta-analysis, definition of study coding criteria, the actual analyses, and the report itself. Ferrari coded the individual research reports in consultation with Honorton and/or Hansen.

claims are generally believed to be based on factors such as delusion, irrationality, and superstitious thinking. The concept of precognition runs counter to accepted notions of causality and appears to conflict with current scientific theory. Nevertheless, over the past half-century a substantial number of experiments have been reported claiming empirical support for the hypothesis of precognition. Subjects in forced-choice experiments, according to many reports, have correctly predicted to a statistically significant degree the identity (or order) of target stimuli randomly selected at a later time.

We performed a meta-analysis of forced-choice precognition experiments published in the English-language research literature between 1935 and 1987. Four major questions were addressed through this meta-analysis: (1) Is there overall evidence for accurate target identification (above-chance hitting) in experimental precognition studies? (2) What is the magnitude of the overall precognition effect? (3) Is the observed effect related to variations in methodological quality that could allow a more conventional explanation? (4) Does precognition performance vary systematically with potential moderating variables, such as differences in subject populations, stimulus conditions, experimental setting, knowledge of results, and time interval between subject response and target generation?

Delineating the Domain

Retrieval of Studies

Parapsychological research is still academically taboo, and it is unlikely that there have been many dissertations and theses in this area that have escaped publication. Our retrieval of studies for this meta-analysis is therefore based on the published literature. The studies include all forced-choice precognition experiments appearing in the peer-reviewed English-language parapsychology journals: *Journal of Parapsychology, Journal* (and *Proceedings) of the Society for Psychical Research, Journal of the American Society for Psychical Research, European Journal of Parapsychology* (including the *Research Letter* of the Utrecht University Parapsychology Laboratory), and abstracts of peer-reviewed papers presented at Parapsychological Association meetings published in *Research in Parapsychology.*

Criteria for Inclusion

Our review is restricted to fixed-length studies in which significance levels and effect sizes based on direct hitting can be calcu-

lated. Studies using outcome variables other than direct hitting, such as run-score variance and displacement effects, are included only if the report provides relevant information on direct hits (i.e., number of trials, hits, and probability of a hit). Finally, we exclude studies conducted by two investigators, S. G. Soal and Walter J. Levy, whose work has been unreliable.

Many published reports contain more than one experiment or experimental unit. In experiments involving multiple conditions, significance levels and effect sizes are calculated for each condition.

Outcome Measures

Significance level. Significance levels (z scores) were calculated for each study from the reported number of trials, hits, and probability of success using the normal approximation to the binomial distribution with continuity correction. Positive z scores indicate above-chance scoring, and negative z scores reflect below-chance scoring.

Effect size. Because most parapsychological experiments, particularly those in the older literature, have used the trial rather than the subject as the sampling unit, we use a trial-based estimator of effect size. The effect size (*ES*) for each study is the z score divided by the square root of the number of trials in the study.[1]

General Characteristics of the Domain

We located 309 studies in 113 separate publications. These studies were contributed by 62 different senior authors and were published over a 53-year period, between 1935 and 1987. Considering the half-century time-span over which the precognition experiments were conducted, it is not surprising that the studies are very diverse.

The database comprises nearly two million individual trials and more than 50,000 subjects. Study sample sizes range from 25 to 297,060 trials (median = 1,194). The number of subjects ranges from 1 to 29,706 (median = 16). The studies use a variety of methodologies, ranging from guessing ESP cards and other card symbols to automated random number generator experiments. The domain encompasses diverse subject populations: the most frequently used

[1] Elsewhere (Honorton, 1985), we have used the effect size index Cohen's h (Cohen, 1977), and one referee has asked that we explain why we are now using $z/N^{1/2}$. The answer is that h and $z/N^{1/2}$ yield virtually identical results, and $z/N^{1/2}$ is computationally simpler. For the present sample of 309 precognition studies, the mean difference between the two indices is .00047, and the standard deviation of the difference is .026; $t(308) = 0.312$, $p = .756$, two-tailed. The correlation between the two indices is .97.

TABLE 1
OVERALL SIGNIFICANCE LEVEL AND EFFECT SIZE

	z	*ES*
Mean	0.65	0.020
SD	2.68	0.100
Lower 95% confidence estimate	0.40	0.011

Combined $z = 11.41$, $p = 6.3 \times 10^{-25}$

"Fail-safe *N*" = 14,268

$t(ES) = 3.51$, 308 *df*, $p = .00025$

population is students (in approximately 40% of the studies); the least frequently used populations are the experimenters themselves and animals (each used in about 5% of the studies).

Though a few studies tested subjects through the mail, more typically subjects were tested in person, either individually or in groups. Target selection methods included no randomization at all (studies using "quasi-random" naturalistic events), informal methods including manual card-shuffling or dice-throwing, and formal methods, primarily random number tables or random number generators. The time interval between the subjects' responses and target generation varied from less than one second to one year.

OVERALL CUMULATION

Evidence for an overall effect is strong. As shown in the top part of Table 1, the overall results are highly significant.[2] Lower bound (one-tailed) 95% confidence estimates of the mean *z* score and *ES* are displayed in the bottom portion of Table 1.

Ninety-two studies (30%) show significant hitting at the 5% level, and significant outcomes are contributed by 40 different investigators. The *z* scores correlate significantly with sample size: $r(307) = .156$, $p = .003$. The mean number of trials for significant studies is 34% larger than the mean number of trials for nonsignificant studies.

[2] The statistical analyses presented here were performed using SYSTAT (Wilkinson, 1988). When *t* tests are reported on samples with unequal variances, they are calculated using the separate variances within groups for the error and degrees of freedom following Brownlee (1965). Unless otherwise specified, *p* levels are one-tailed. Combined z's are based on Stouffer's method (Rosenthal, 1984).

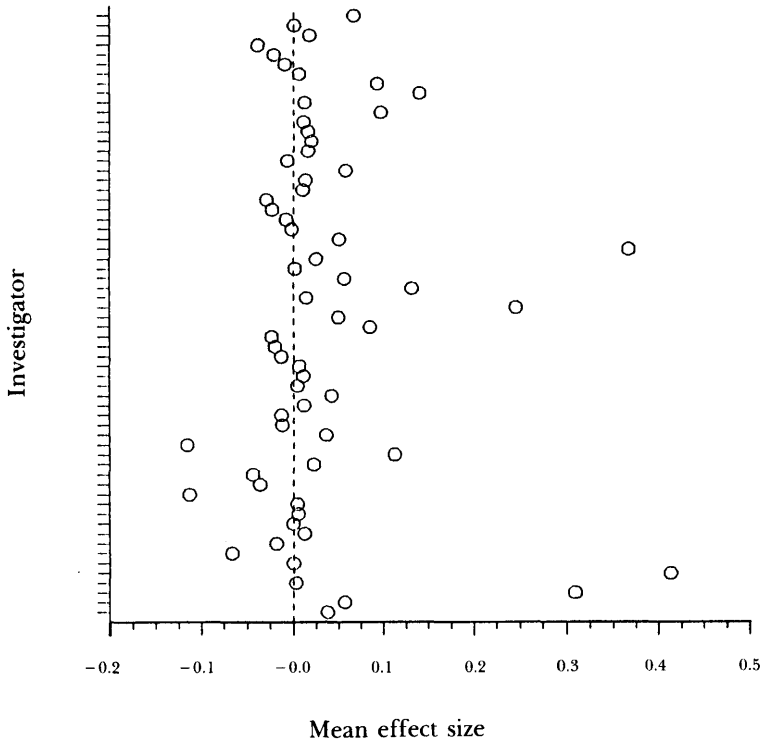

Figure 1. Mean effect size by investigator. $N = 62$ investigators.

Replication Across Investigators

Virtually the same picture emerges when the cumulation is by investigator rather than study as the unit of analysis; the combined z is 12.13, and 23 of the 62 investigators (37%) have overall outcomes significant at the 5% level. The mean (investigator) effect size is 0.033 ($SD = .093$).

There is a significant difference in the mean *ES* across investigators, but it is surprisingly small: Kruskal-Wallis one-way ANOVA by ranks, $\chi^2(61) = 82.71$, $p = .034$. The effect is clearly not due to a few major contributors. If investigators contributing more than three studies are eliminated, leaving 33 investigators, the combined z is still 6.00 ($p = 1.25 \times 10^{-9}$) and the mean *ES* is .028 ($SD = .091$). Figure 1 shows the mean effect sizes by investigator.

These results indicate substantial cross-investigator replicability and directly contradict the claim of critics such as Akers (1987) that

successful parapsychological outcomes are achieved by only a few investigators.

The Filedrawer Problem

A well-known reporting bias exists throughout the behavioral sciences favoring publication of "significant" studies (e.g., Sterling, 1959). The extreme view of this "filedrawer problem" is that "the journals are filled with the 5% of the studies that show Type I errors, while the filedrawers back at the lab are filled with the 95% of the studies that show nonsignificance..." (Rosenthal, 1984, p. 108). Recognizing the importance of this problem, the Parapsychological Association in 1975 adopted an official policy against selective reporting of positive results.[3] Examination of the parapsychological literature shows that nonsignificant results are frequently published, and, in the precognition database, 70% of the studies have reported nonsignificant results. Nevertheless, 75% of the precognition studies were published before 1975, and we must ask to what extent selective publication bias could account for the cumulative effects we observe.

The central section of Table 1 uses Rosenthal's (1984) "fail-safe *N*" statistic to estimate the number of unreported studies with *z* scores averaging zero that would be necessary to reduce the known database to nonsignificance. The filedrawer estimate indicates that over 46 unreported studies must exist for each reported study to reduce the cumulative outcome to a nonsignificant level.

A different approach to the filedrawer problem is described by Dawes, Landman, and Williams (1984; personal communication from Dawes to Honorton, July 14, 1988). Their truncated normal curve analysis, like Rosenthal's "fail-safe *N*," is based on normal curve assumptions. Their null hypothesis is that *z* scores above some critical level (e.g., $z = 1.65$, 1.96, etc.) are randomly sampled from $N(0,1)$ above that critical level. The alternative to the null hypothesis is that, because there is some real effect, the distribution of *z*'s is shifted to the right of 0 and the *z*'s will be larger than predicted by the null. For a critical level of $z = 1.65$, the expected mean *z* is 2.06 and the variance is .14. In the precognition database, there are 92 studies with *z*'s > 1.65. Their average is 3.61, not 2.06 as predicted

[3] Analyses indicate no significant difference in the magnitude of reported study outcomes before and after 1975. The mean *ES* for studies prior to 1975 is 0.021 (*SD* = .099), and for studies reported thereafter the mean is 0.017 (*SD* = .106): $t(307) = 0.28$, $p = .782$, two-tailed.

by the null hypothesis. Since the variance of the normal truncated above 1.65 is .14, the *test* z (using the Central Limit Theorem) comparing 3.61 to 2.06 is 39.84 [1.55 divided by $(.14/92)^{1/2}$]. Here, p is virtually zero. Similar results are found with cut points of 1.96, 2.33, and 2.58.

On the basis of these analyses, we conclude that the cumulative significance of the precognition studies cannot satisfactorily be explained by selective reporting.

Outlier Reduction

Although the overall z scores and effect sizes cannot reasonably be attributed to chance, inspection of the standard deviations in Table 1 indicates that the study outcomes are extremely heterogeneous. Given the diversity of methods, subject populations, and other study features that characterize this research domain, this is not surprising.

The study outcomes are in fact extremely heterogeneous. Although a major objective of this meta-analysis is to account for the variability across studies by blocking on differences in study quality, procedural features, and sampling characteristics, the database clearly contains extreme outliers. The z scores range from -5.1 to 19.6, a 25-sigma spread! The standardized index of kurtosis (g_2) is 9.47, suggesting that the tails of the distribution are much too long for a normal distribution.

We eliminated the extreme outliers by performing a "10 percent trim" on the study z scores (Barnett & Lewis, 1978). This involves eliminating studies with z scores in the upper and lower 10% of the distribution, and results in an adjusted sample of 248 studies. The trimmed z scores range from -2.24 to 3.21 ($g_2 = -1.1$). The revised z scores and effect sizes are presented in Table 2.

Elimination of extreme outliers reduces the combined z scores by approximately one half, but the outcomes remain highly significant. Twenty-five percent of the studies (62/248) show overall significant hitting at the 5% level. Lower bound confidence estimates show that the mean z's and effect sizes are above 0 at the 95% confidence level.

Elimination of outliers reduces the total number of investigators from 62 to 57, but the results remain basically the same when the analyses are based on investigators rather than studies. The combined z is 6.84; 18 of the 57 investigators (31.6%) have overall sig-

TABLE 2
SIGNIFICANCE LEVEL AND EFFECT SIZE FOR TRIMMED SAMPLE

	z	*ES*
Mean	0.38	0.012
SD	1.45	0.065
Lower 95% confidence estimate	0.23	0.005

Combined $z = 6.02$, $p = 1.1 \times 10^{-9}$

$t(ES) = 2.90$, 247 df, $p = .002$

nificant outcomes at the 5% level. The mean (investigator) *ES* is 0.020 (*SD* = .05).

For the trimmed sample, the difference in *ES* across investigators is *not* significant: Kruskal-Wallis one-way ANOVA by ranks, $\chi^2(56) = 59.34$, $p = .355$. If investigators contributing more than three studies are eliminated, leaving 37 investigators, the combined *z* is still 5.00 ($p = 3.0 \times 10^{-7}$) and the mean *ES* is 0.022 (*SD* = .056). Figure 2 shows the mean effect size by investigator.

Thus, elimination of the outliers does not substantially affect the conclusions drawn from our analysis of the database as a whole. There clearly is a nonchance effect. In the remainder of this report, we use the trimmed sample to examine covariations in effect size and a variety of methodological and other study features.

STUDY QUALITY

Because target stimuli in precognition experiments are selected only after the subjects' responses have been registered, precognition studies are usually not vulnerable to sensory leakage problems. Other potential threats to validity must, however, be considered. The problem of variations in research quality remains a source of controversy in meta-analysis. Some meta-analysts advocate eliminating low quality studies whereas others recommend empirically accessing the impact of variations in quality on study outcome. Rosenthal (1984) points out that the practice of discarding studies is equivalent to assigning them weights of zero, and he recommends weighting study *z* scores in relation to ratings of research quality.

Study Quality Criteria

Ideally, the assessment of study quality should be performed by knowledgeable specialists who are blind to the study outcomes. In

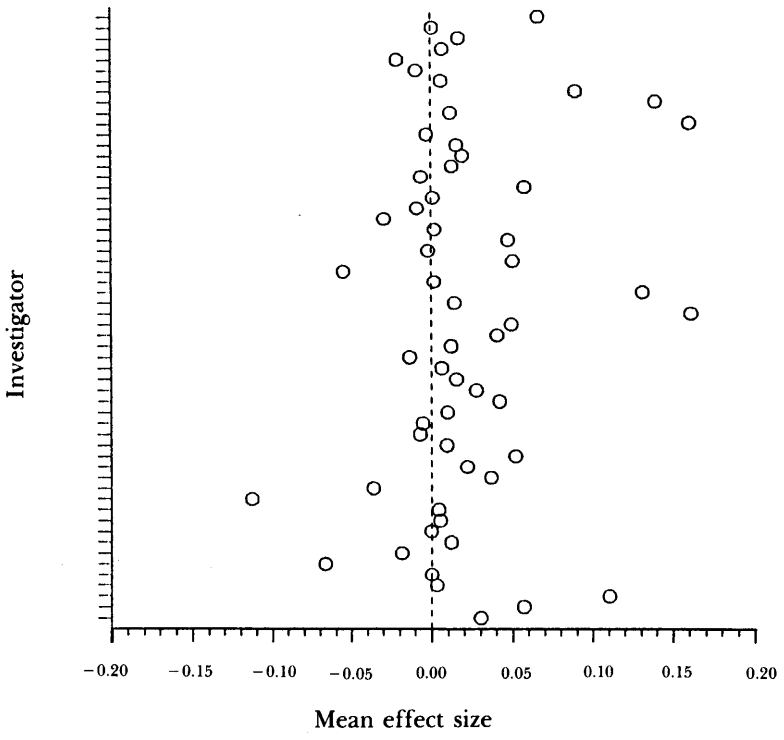

Figure 2. Mean effect size by investigator for trimmed sample. $N = 57$ investigators.

practice, this is usually not feasible, particularly when, as in the present case, large numbers of studies are involved. For our analysis of study quality, statistical and methodological variables are defined and coded in terms of procedural descriptions (or their absence) in the research reports. This approach was used in an earlier meta-analysis of psi ganzfeld research (Honorton, 1985), and it led to study quality ratings that were generally in agreement, $r(26) = .766$, $p = 10^{-6}$, with independent "flaw" ratings by an outside critic (Hyman, 1985).

One point is given (or withheld) for each of the following eight criteria:

Specification of sample size. Does the investigator preplan the number of trials to be included in the study or is the study vulnerable to the possibility of optional stopping? Credit is given to reports that explicitly specify the sample size. Studies involving group testing, in which it is not feasible to specify the sample size precisely, are also

given credit. No credit is given to studies in which the sample size is either not preplanned or not addressed in the experimental report.

Preplanned analysis. Is the method of statistical analysis, including the outcome (dependent variable) measure, preplanned? Credit is given to studies explicitly specifying the form of analysis and the outcome measure. No credit is given to those not explicitly stating the form of the analysis or those in which the analysis is clearly post hoc.

Randomization method. Credit is given for use of random number tables, random number generators, and mechanical shufflers. No credit is given for failure to randomize (i.e., use of "quasi-random naturalistic events") or for informal methods such as hand-shuffling, die-casting, and drawing lots.

Controls. Credit is given to studies reporting randomness control checks, such as random number generator (RNG) control series and empirical cross-check controls.

Recording. One point is allotted for automated recording of targets and responses, and another for duplicate recording.

Checking. One point is allotted for automated checking of matches between target and response, and another for duplicate checking of hits.

Study Quality Analysis

Each study received a quality weight between 0 and 8 (mean = 3.3, SD = 1.8). We find no significant relationship between study quality and *ES:* $r(246) = .081$, $p = .202$, two-tailed. This tendency for study outcomes to correlate *positively* with study quality has the consequence that the quality-weighted z score of 6.26 is slightly *larger* than the unweighted z of 6.02. Table 3 shows the correlations between effect size and each of the eight individual quality measures.[4] The mean effect sizes by quality level are displayed graphically in Figure 3.

[4] The correlation between *ES* and study quality is also nonsignificant for the untrimmed sample of 309 studies: $r(307) = -.060$, $p = .289$. The quality-weighted z score is 7.38: $p = 2.32 \times 10^{-13}$. However, three of the individual quality measures are significantly related to performance. Controls and duplicate checking correlate significantly positively with *ES*, and randomization correlates significantly negatively with *ES*. These correlations appear to be due to a few studies with z scores that are extreme outliers ($z > 7$). When the 10 studies with $z > 7$ are eliminated, the significant correlations between quality and *ES* disappear.

TABLE 3
CORRELATIONS BETWEEN EFFECT SIZE AND QUALITY MEASURES

Quality measure	$r(246)$
Sample size specified in advance	−.100
Preplanned analysis	−.001
Randomization	−.011
Controls	.058
Automated recording	.169
Duplicate recording	.047
Automated checking	.136
Duplicate checking	.078

Quality Extremes

Is there a tendency for extremely weak studies to show larger effects than exceptionally "good" studies? Analysis on the extremes of the quality ratings indicates that this is not the case.

This analysis, based on the untrimmed sample of 309 studies, uses studies with quality ratings outside the interquartile range of the rating distribution (median = 4, $Q_1 = 2$, $Q_3 = 5$). There are 56 "low-quality" studies (ratings of 0–1) and 35 "high-quality" studies (ratings of 6–8). The high-quality studies have effect sizes that are not significantly lower than the low-quality studies; the *ES* means are 0.017 ($SD = 0.063$) and 0.037 ($SD = 0.137$), for the low- and high-quality studies, respectively: $t(82) = -.92$, $p = .358$, two-tailed.

Quality Variation in Publication Sources

Precognition *ES* is not significantly related to source of publication: Kruskal-Wallis one-way ANOVA, $\chi^2(4) = 0.78$, $p = .942$. However, the sources of publication differ significantly in study quality: Kruskal-Wallis one-way ANOVA, $\chi^2(4) = 17.19$, $p = .002$. This is due largely to the lower quality of studies published in the *Journal of the Society for Psychical Research* and in *Research in Parapsychology*.

Study Quality in Relation to Year of Publication

Precognition effect size has remained constant over a half-century of research, even though the methodological quality of the re-

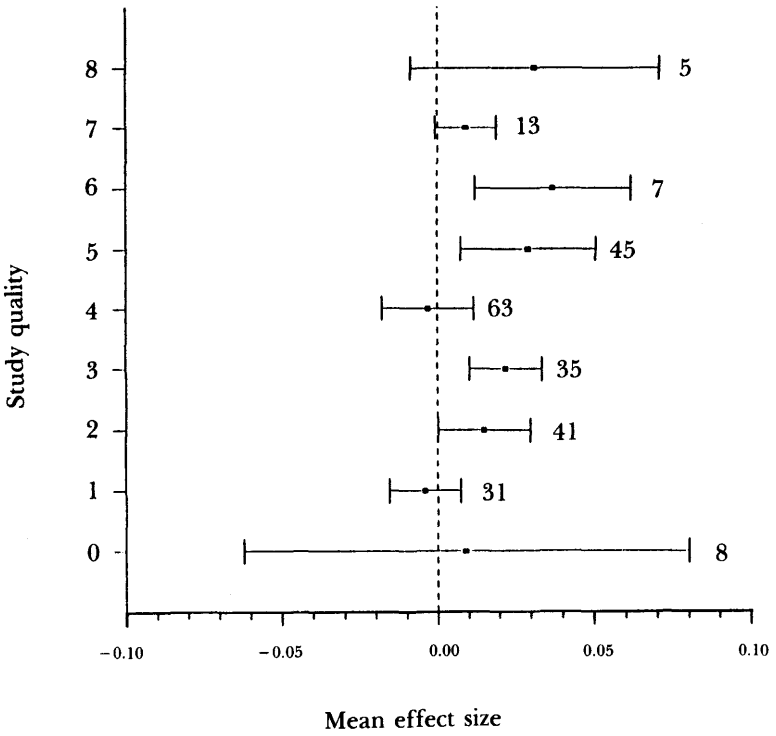

Figure 3. Precognition effect size in relation to study quality, with 95% confidence limits. $N = 248$ studies.

search has improved significantly during this period. The correlation between *ES* and year of publication is $-.071$: $t(307) = -1.25$, $p = .213$, two-tailed. Study quality and year of publication are, however, positively and significantly correlated: $r(246) = .282$, $p = 2 \times 10^{-7}$, two-tailed.

Critics of parapsychology have long believed that evidence for parapsychological effects disappears as the methodological rigor increases. The precognition database does not support this belief.

"Real-Time" Alternatives to Precognition

Investigators have long been aware of the possibility that precognition effects could be modeled without assuming either time reversal or backward causality. For example, outcomes from studies with

targets based on indeterminate random number generators (RNGs) could be due to a causal influence on the RNG—a psychokinetic (PK) effect—rather than information acquisition concerning its future state. In experiments with targets based on prepared tables of random numbers, the possibility exists that the experimenter or other randomizer may be the actual psi source, unconsciously using "real-time" ESP combined with PK to choose an entry point in the random number sequence that will significantly match the "subject's" responses. While the latter possibility may seem far-fetched, it cannot be logically eliminated if one accepts the existing evidence for contemporaneous ESP and PK, and it has been argued that it is less far-fetched than the alternative of "true" precognition.

Morris (1982) discusses models of experimental precognition based on "real-time" psi alternatives and methods for testing "true" precognition. In general terms, these methods constrain the selection of the target sequence so as to eliminate nonprecognitive psi intervention. In the most common procedure, attributed to Mangan (1955), dice are thrown to generate a set of numbers that are mathematically manipulated to obtain an entry point in the random number table. This procedure is sufficiently complex "as to be apparently beyond the capacities of the human brain, thus ruling out PK because the 'PKer' would not know what to do even via ESP" (Morris, 1982, p. 329).

Two features of precognition study target determination procedures were coded to assess "real-time" psi alternatives to precognition: method of determining random number table entry point and use of Mangan's method.

Methods of eliminating "real-time" psi alternatives have not been used in studies with random number generators and have only been used in a small number of studies involving randomization by hand-shuffling. These analyses are therefore restricted to studies using random number tables (N = 138).

Method of Determining RNT Entry Point

The reports describe six different methods of obtaining entry points in random number tables. If the study outcomes were due to subjects' precognitive functioning rather than to alternative psi modes on the part of the experimenter or the experimenter's assistants, there should be no difference in mean effect size across the various methods used to determine the entry point. Indeed, our analysis indicates that the study effect sizes do not vary systemati-

cally as a function of method of determining the entry point: Kruskal-Wallis one-way ANOVA by ranks: $\chi^2(5) = 7.32$, $p = .198$.

Use of Mangan's Method

We find no significant difference in *ES* between studies using complex calculations of the type introduced by Mangan to fix the random number table entry point and those that do not use such calculations: $t(45) = 0.38$, $p = .370$, two-tailed.

Moderating Variables

The stability of precognition study outcomes over a 50-year period, which we described earlier, is also bad news. It shows that investigators in this area have yet to develop sufficient understanding of the conditions underlying the occurrence (or detection) of these effects to reliably increase their magnitude. We have identified four variables that appear to covary systematically with precognition *ES*: (1) selected versus unselected subjects, (2) individual versus group testing, (3) feedback level, and (4) time interval between subject response and target generation.

The analyses use the raw study *z* scores and effect sizes; we found that this results in uniformly more conservative estimates of relationships with moderating variables than when the analyses are based on quality-weighted *z* scores and effect sizes.

Selected Versus Unselected Subjects

Our meta-analysis identifies eight subject populations: unspecified subject populations, mixtures of several different populations, animals, students, children, "volunteers," experimenter(s), and selected subjects.

Effect size magnitude does not vary significantly across these eight subject populations: Kruskal-Wallis one-way ANOVA, $\chi^2(7) = 10.90$, $p = .143$. Effect sizes by subject population are displayed in Figure 4.

However, studies using subjects selected on the basis of prior performance in experiments or pilot tests show significantly larger effects than studies using unselected subjects. As shown in Table 4, 60% of the studies with selected subjects are significant at the 5% level. The mean *z* score for these studies is 1.39 ($SD = 1.40$). The *ES* is significantly higher for selected-subjects studies than for stud-

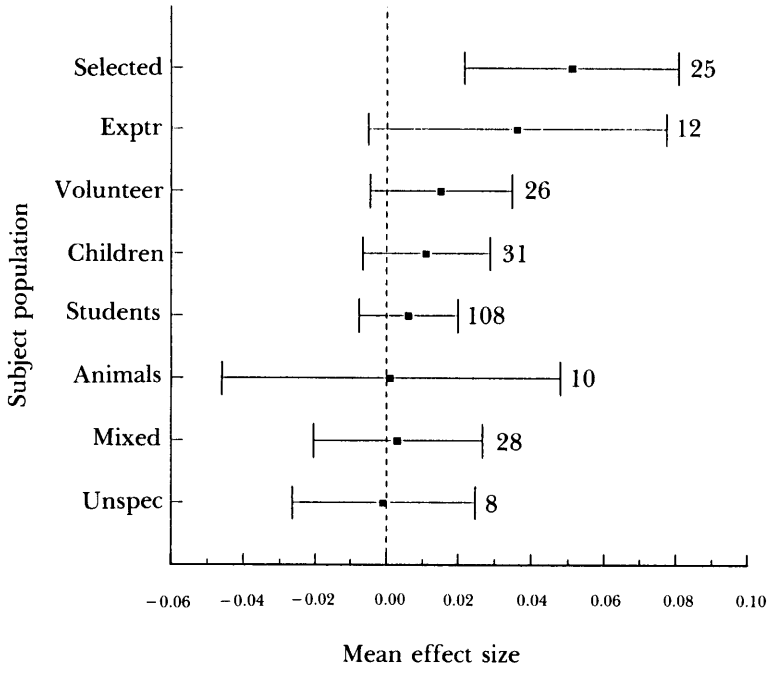

Figure 4. Precognition effect size by subject population, with 95% confidence limits. $N = 248$ studies.

ies with unselected subjects. The t test of the difference in mean ES is equivalent to a point-biserial correlation of .198.

Does this difference result from less stringent controls in studies with selected subjects? The answer appears to be "No." The average quality of studies with selected subjects is *higher* than studies using

TABLE 4
SELECTED VERSUS UNSELECTED SUBJECTS

	Selected	Unselected
N studies	25	223
Combined z	6.89	4.04
Studies with $p < .05$	60%	21%
Mean ES	.051	.008
SD_{ES}	.075	.063
$t(246) = 3.16, p = .001$		

TABLE 5
INDIVIDUAL VERSUS GROUP TESTING

	Individual	Group
N studies	97	105
Combined *z*	6.64	1.29
Studies with $p < .05$	30%	19%
Mean *ES*	.021	.004
SD_{ES}	.060	.066
$t(200) = 1.89, p = .03$		

unselected subjects: $t(27) = 1.51$, $p = .142$, two-tailed. This result appears to reflect a general tendency toward increased rigor and more detailed reporting in studies with selected subjects.

Individual Versus Group Testing

Subjects were tested in groups, individually, or through the mail. Studies in which subjects were tested individually by an experimenter have a significantly larger mean *ES* than studies involving group testing (Table 5).

The *t* test of the difference is equivalent to a point-biserial correlation of .132, favoring individual testing. Of the studies with subjects tested individually, 30% are significant at the 5% level.

The methodological quality of studies with subjects tested individually is significantly higher than that of studies involving group testing: $t(137) = 3.08$, $p = .003$, two-tailed. This result is consistent with the conjecture that group experiments are frequently conducted as "targets of opportunity" and may often be carried out hastily in an afternoon without the preparation and planning that go into a study with individual subjects that may be conducted over a period of weeks or months.

Thirty-five studies were conducted through the mail. In these studies, subjects completed the task at their leisure and mailed their responses to the investigator. These correspondence studies yield outcomes similar to those involving individual testing. The combined *z* score is 2.66, with a mean *ES* of 0.018 ($SD = .082$). Ten correspondence studies (25.7%) are significant at the 5% level.

Eleven studies are unclassifiable with regard to experimental setting.

TABLE 6
FEEDBACK RECEIVED BY SUBJECTS

	Feedback of Results			
	None	Delayed	Run score	Trial-by-trial
N studies	15	21	21	47
Combined *z*	−1.30	2.11	4.74	6.98
Studies with $p < .05$	0.0%	19.0%	33.3%	42.6%
Mean *ES*	−.001	.009	.023	.035
SD_{ES}	.028	.036	.048	.072

Feedback

A significant positive relationship exists between the degree of feedback subjects receive about their performance and precognitive effect size (Table 6).

Subject feedback information is available for 104 studies. These studies fall into four feedback categories: no feedback, delayed feedback (usually notification by mail), run-score feedback, and trial-by-trial feedback. We gave these categories numerical values between 0 and 3. Precognition effect size correlates .231 with feedback level (102 *df*, $p = .009$). Of the 47 studies involving trial-by-trial feedback, 20 (42.6%) are significant at the 5% level. None of the studies without subject feedback are significant.

Feedback level correlates positively though not significantly with research quality: $r(102) = .173$, $p = .082$, two-tailed. Inadequate randomization is the most plausible source of potential artifacts in studies with trial-by-trial feedback. We performed a separate analysis on the 47 studies in this group. Studies using formal methods of randomization do not differ significantly in mean *ES* from those with informal randomization: $t(15) = 0.67$, $p = .590$, two-tailed. Similarly, studies reporting randomness control data do not differ significantly in *ES* from those not including randomness controls: $t(42) = 0.79$, $p = .436$, two-tailed.

Time Interval

The interval between the subject's response and target selection ranges from less than one second to one year. Information about the time interval is available for 144 studies. This information, how-

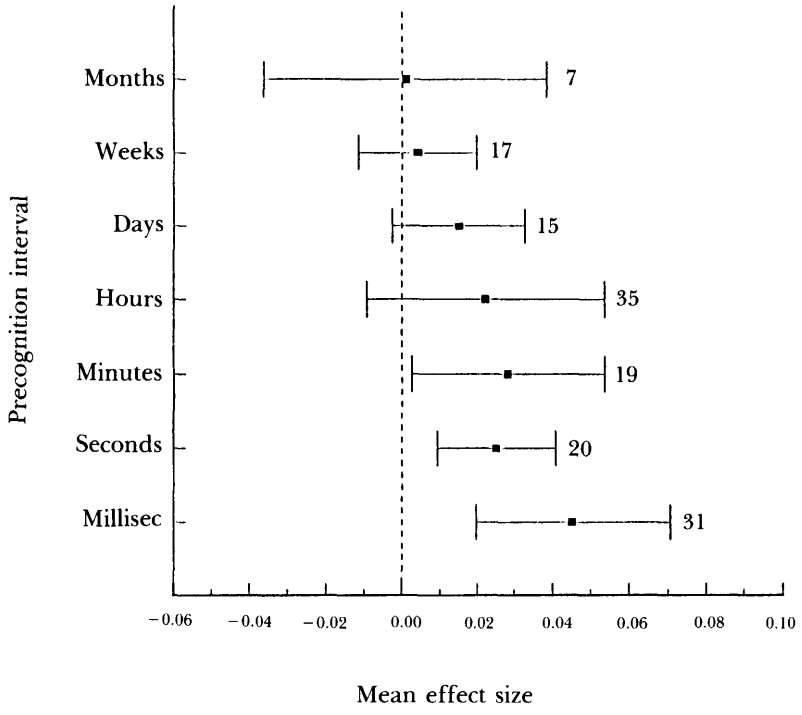

Figure 5. Effect size by precognition interval, with 95% confidence limits. N = 144 studies.

ever, is often imprecise. Our analysis of the relationship between precognitive *ES* and time interval is therefore limited to seven broad interval categories: milliseconds, seconds, minutes, hours, days, weeks, and months. (Effect sizes by precognition interval are displayed in Figure 5.)

Although it is confounded with degree of feedback, there is a significant decline in precognition *ES* over increasing temporal distance: $r(142) = -.199$, $p = .017$, two-tailed. The largest effects occur over the millisecond interval: $N = 31$ studies, combined $z = 6.03$, mean $ES = 0.045$, $SD = .073$. The smallest effects occur over periods ranging from a month to a year: $N = 7$, combined $z = 0.53$, mean $ES = 0.001$, $SD = .049$.

Interestingly, the decline of precognition performance over increasing temporal distances results entirely from studies using *un-*

selected subjects: $r(122) = -.235$, $p = .009$, two-tailed. Studies with selected subjects show a nonsignificant positive relationship between *ES* and time interval: $r(18) = .077$, $p = .745$, two-tailed. Although the difference between these two correlations is not significant ($z = 1.24$), this suggests that the origin of the decline over time may be motivational rather than the result of some intrinsic physical boundary condition. The relationship between precognition *ES* and feedback also supports this conjecture. Nevertheless, any finding suggesting potential boundary conditions on the phenomenon should be vigorously pursued.

Influence of Moderating Variables in Combination

The above analyses examine the impact of each moderating variable in isolation. In this final set of analyses, we explore their joint influence on precognition performance. For this purpose, we identify two subgroups of studies. One subgroup is characterized by the use of selected subjects tested individually with trial-by-trial feedback. We refer to this as the *Optimal* group ($N = 8$ studies). The second group is characterized by the use of unselected subjects tested in groups with no feedback. We refer to this as the *Suboptimal* group ($N = 9$ studies).

The Optimal studies are contributed by four independent investigators and the Suboptimal studies are contributed by two of the same four investigators. All of the Optimal studies involve short precognition time intervals (millisecond interval); the Suboptimal studies involve longer intervals (intervals of weeks or months). All of the Optimal studies and 5 of the 9 Suboptimal studies use RNG methodology. The two groups do not differ significantly in average sample size. The mean study quality for the Optimal group is significantly higher than that of the Suboptimal studies: Optimal mean = 6.63, $SD = 0.92$; Suboptimal mean = 3.44, $SD = 0.53$; $t(10) = 8.63$, $p = 3.3 \times 10^{-6}$, two-tailed.

The combined impact of the moderating variables appears to be quite strong (Table 7). Seven of the 8 Optimal studies (87.5%) are independently significant at the 5% level, whereas none of the Suboptimal studies are statistically significant. All four investigators contributing studies to the Optimal group have significant outcomes.[5]

[5] In the untrimmed sample of 309 studies, there are a total of 17 Optimal studies. The mean *ES* is 0.117 ($SD = .154$), and the combined z is 15.84. The percentage of independently significant studies is virtually the same as it is in the trimmed sample: 15 of the 17 studies (88.2%) are significant.

TABLE 7
IMPACT OF MODERATORS IN COMBINATION

	"Optimal" studies	"Suboptimal" studies
N studies	8	9
Combined z	6.14	−1.29
Studies with $p < .05$	87.5%	0.0%
Mean ES	.055	.005
SD_{ES}	.045	.035
	$t(15) = 2.61, p = .01$	
	$r = .559$	

These results are quite striking and suggest that future studies combining these moderators should yield especially reliable effects.

SUMMARY AND CONCLUSIONS

Our meta-analysis of forced-choice precognition experiments confirms the existence of a small but highly significant precognition effect. The effect appears to be replicable; significant outcomes are reported by 40 investigators using a variety of methodological paradigms and subject populations.

The precognition effect is statistically very robust: it remains highly significant despite elimination of studies with z scores in the upper and lower 10% of the z-score distribution and when a third of the remaining investigators—the major contributors of precognition studies—are eliminated.

Estimates of the "filedrawer" problem and consideration of parapsychological publication practices indicate that the precognition effect cannot plausibly be explained on the basis of selective publication bias. Analyses of precognition effect sizes in relation to eight measures of research quality fail to support the hypothesis that the observed effect is driven to any appreciable extent by methodological flaws; indeed, several analyses indicate that methodologically superior studies yield stronger effects than methodologically weaker studies.

Analyses of parapsychological alternatives to precognition, although limited to the subset of studies using random number tables, provide no support for the hypothesis that the effect results from

the operation of contemporaneous ESP and PK at the time of randomization.

Although the overall precognition effect size is small, this does not imply that it has no practical consequences. It is, for example, of the same order of magnitude as effect sizes leading to the early termination of several major medical research studies. In 1981, the National Heart, Lung, and Blood Institute discontinued its study of propranolol because the results were so favorable to the propranolol treatment that it would be unethical to continue placebo treatment (Kolata, 1981); the effect size was 0.04. More recently, The Steering Committee of the Physicians' Health Study Research Group (1988), in a widely publicized report, terminated its study of the effects of aspirin in the prevention of heart attacks for the same reason. The aspirin group suffered significantly fewer heart attacks than a placebo control group; the associated effect size was 0.03.

The most important outcome of the meta-analysis is the identification of several moderating variables that appear to covary systematically with precognition performance. The largest effects are observed in studies using subjects selected on the basis of prior test performance, who are tested individually, and who receive trial-by-trial feedback. The outcomes of studies combining these factors contrast sharply with the null outcomes associated with the combination of group testing, unselected subjects, and no feedback of results. Because the two groups of studies were conducted by a subset of the same investigators, it is unlikely that the observed difference in performance is due to experimenter effects. Indeed, these outcomes underscore the importance of carefully examining differences in subject populations, test setting, and so forth, before resorting to facile "explanations" based on psi-mediated experimenter effects or the "elusiveness of psi."

The identification of these moderating variables has important implications for our understanding of the phenomena and provides a clear direction for future research. The existence of moderating variables indicates that the precognition effect is not merely an unexplained departure from a theoretical chance baseline, but rather is an effect that covaries with factors known to influence more familiar aspects of human performance. It should now be possible to exploit these moderating factors to increase the magnitude and reliability of precognition effects in new studies.

References

Akers, C. (1987). Parapsychology is science, but its findings are inconclusive. *Behavioral and Brain Sciences*, **10**, 566–568.

BARNETT, V., & LEWIS, T. (1978). *Outliers in statistical data.* New York: John Wiley & Sons.

BROWNLEE, K. A. (1965). *Statistical theory and methodology in science and engineering.* New York: John Wiley & Sons.

COHEN, J. (1977). *Statistical power analysis for the behavioral sciences.* New York: Academic Press.

DAWES, R. M., LANDMAN, J., & WILLIAMS, J. (1984). Reply to Kurosawa. *American Psychologist,* **39,** 74–75.

HONORTON, C. (1985). Meta-analysis of psi ganzfeld research: A response to Hyman. *Journal of Parapsychology,* **49,** 51–92.

HYMAN, R. (1985). The ganzfeld psi experiment: A critical appraisal. *Journal of Parapsychology,* **49,** 3–50.

KOLATA, G. B. (1981). Drug found to help heart attack survivors. *Science,* **214,** 774–775.

MANGAN, G. L. (1955). Evidence of displacement in a precognition test. *Journal of Parapsychology,* **19,** 35–44.

MORRIS, R. L. (1982). Assessing experimental support for true precognition. *Journal of Parapsychology,* **46,** 321–336.

ROSENTHAL, R. (1984). *Meta-analytic procedures for social research.* Beverly Hills, CA: Sage.

STEERING COMMITTEE OF THE PHYSICIANS' HEALTH STUDY RESEARCH GROUP. (1988). Preliminary report: Findings from the aspirin component of the ongoing Physicians' Health Study. *New England Journal of Medicine,* **318,** 262–264.

STERLING, T. D. (1959). Publication decisions and their possible effects on inferences drawn from tests of significance—or vice versa. *Journal of the American Statistical Association,* **54,** 30–34.

WILKINSON, L. (1988). *SYSTAT: The system for statistics.* Evanston, IL: SYSTAT.

CHRONOLOGICAL LISTING OF STUDIES IN META-ANALYSIS

CARINGTON, W. (1935). Preliminary experiments in precognitive guessing. *Journal of the Society for Psychical Research,* **29,** 86–104.

RHINE, J. B. (1938). Experiments bearing on the precognition hypothesis: I. Pre-shuffling card calling. *Journal of Parapsychology,* **2,** 38–54.

RHINE, J. B., SMITH, B. M., & WOODRUFF, J. L. (1938). Experiments bearing on the precognition hypothesis: II. The role of ESP in the shuffling of cards. *Journal of Parapsychology,* **2,** 119–131.

HUMPHREY, B. M., & PRATT, J. G. (1941). A comparison of five ESP test procedures. *Journal of Parapsychology,* **5,** 267–293.

RHINE, J. B. (1941). Experiments bearing upon the precognition hypothesis: III. Mechanically selected cards. *Journal of Parapsychology,* **5,** 1–57.

STUART, C. E. (1941). An analysis to determine a test predictive of extra-chance scoring in card-calling tests. *Journal of Parapsychology,* **5,** 99–137.

HUMPHREY, B. M., & RHINE, J. B. (1942). A confirmatory study of salience in precognition tests. *Journal of Parapsychology,* **6,** 190–219.

RHINE, J. B. (1942). Evidence of precognition in the covariation of salience ratios. *Journal of Parapsychology,* **6,** 111–143.

NICOL, J. F., & CARINGTON, W. (1947). Some experiments in willed die-throwing. *Proceedings of the Society for Psychical Research,* **48,** 164–175.

THOULESS, R. H. (1949). A comparative study of performance in three psi tasks. *Journal of Parapsychology,* **13,** 263–273.

BASTIN, E. W., & GREEN, J. M. (1953). Some experiments in precognition. *Journal of Parapsychology,* **17,** 137–143.

MCMAHAN, E. A., & BATES, E. K. (1954). Report of further Marchesi experiments. *Journal of Parapsychology,* **18,** 82–92.

MANGAN, G. L. (1955). Evidence of displacement in a precognition test. *Journal of Parapsychology,* **19,** 35–44.

OSIS, K. (1955). Precognition over time intervals of one to thirty-three days. *Journal of Parapsychology,* **19,** 82–91.

NIELSEN, W. (1956). An exploratory precognition experiment. *Journal of Parapsychology,* **20,** 33–39.

NIELSEN, W. (1956). Mental states associated with success in precognition. *Journal of Parapsychology,* **20,** 96–109.

FAHLER, J. (1957). ESP card tests with and without hypnosis. *Journal of Parapsychology,* **21,** 179–185.

MANGAN, G. L. (1957). An ESP experiment with dual-aspect targets involving one trial a day. *Journal of Parapsychology,* **21,** 273–283.

ANDERSON, M., & WHITE, R. (1958). A survey of work on ESP and teacher-pupil attitudes. *Journal of Parapsychology,* **22,** 246–268.

NASH, C. B. (1958). Correlation between ESP and religious value. *Journal of Parapsychology,* **22,** 204–209.

ANDERSON, M. (1959). A precognition experiment comparing time intervals of a few days and one year. *Journal of Parapsychology,* **23,** 81–89.

ANDERSON, M., & GREGORY, E. (1959). A two-year program of tests for clairvoyance and precognition with a class of public school pupils. *Journal of Parapsychology,* **23,** 149–177.

NASH, C. B. (1960). Can precognition occur diametrically? *Journal of Parapsychology,* **24,** 26–32.

FREEMAN, J. A. (1962). An experiment in precognition. *Journal of Parapsychology,* **26,** 123–130.

RHINE, J. B. (1962). The precognition of computer numbers in a public test. *Journal of Parapsychology,* **26,** 244–251.

RYZL, M. (1962). Training the psi faculty by hypnosis. *Journal of the Society for Psychical Research,* **41,** 234–252.

SANDERS, M. S. (1962). A comparison of verbal and written responses in a precognition experiment. *Journal of Parapsychology,* **26,** 23–34.

FREEMAN, J. (1963). Boy-girl differences in a group precognition test. *Journal of Parapsychology,* **27,** 175–181.

RAO, K. R. (1963). Studies in the preferential effect: II. A language ESP test involving precognition and "intervention." *Journal of Parapsychology,* **27,** 147–160.

FREEMAN, J. (1964). A precognition test with a high-school science club. *Journal of Parapsychology,* **28,** 214–221.

FREEMAN, J., & NIELSEN, W. (1964). Precognition score deviations as related to anxiety levels. *Journal of Parapsychology,* **28,** 239–249.

SCHMEIDLER, G. (1964). An experiment on precognitive clairvoyance: Part I. The main results. *Journal of Parapsychology,* **28,** 1–14.

FREEMAN, J. A. (1965). Differential response of the sexes to contrasting arrangements of ESP target material. *Journal of Parapsychology,* **29,** 251–258.

OSIS, K., & FAHLER, J. (1965). Space and time variables in ESP. *Journal of the American Society for Psychical Research,* **59,** 130–145.

FAHLER, J., & OSIS, K. (1966). Checking for awareness of hits in a precognition experiment with hypnotized subjects. *Journal of the American Society for Psychical Research,* **60,** 340–346.

FREEMAN, J. A. (1966). Sex differences and target arrangement: High-school booklet tests of precognition. *Journal of Parapsychology,* **30,** 227–235.

ROGERS, D. P. (1966). Negative and positive affect and ESP run-score variance. *Journal of Parapsychology,* **30,** 151–159.

ROGERS, D. P., & CARPENTER, J. C. (1966). The decline of variance of ESP scores within a testing session. *Journal of Parapsychology,* **30,** 141–150.

BRIER, B. (1967). A correspondence ESP experiment with high-I.Q. subjects. *Journal of Parapsychology,* **31,** 143–148.

BUZBY, D. E. (1967). Subject attitude and score variance in ESP tests. *Journal of Parapsychology,* **31,** 43–50.

BUZBY, D. E. (1967). Precognition and a test of sensory perception. *Journal of Parapsychology,* **31,** 135–142.

FREEMAN, J. A. (1967). Sex differences, target arrangement, and primary mental abilities. *Journal of Parapsychology,* **31,** 271–279.

HONORTON, C. (1967). Creativity and precognition scoring level. *Journal of Parapsychology,* **31,** 29–42.

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

DUVAL, P., & MONTREDON, E. (1968). ESP experiments with mice. *Journal of Parapsychology,* **32,** 153–166.

FEATHER, S. R., & BRIER, R. (1968). The possible effect of the checker in precognition tests. *Journal of Parapsychology,* **32,** 167–175.

FREEMAN, J. A. (1968). Sex differences and primary mental abilities in a group precognition test. *Journal of Parapsychology,* **32,** 176–182.

NASH, C. S., & NASH, C. B. (1968). Effect of target selection, field dependence, and body concept on ESP performance. *Journal of Parapsychology,* **32,** 248–257.

RHINE, L. E. (1968). Note on an informal group test of ESP. *Journal of Parapsychology,* **32,** 47–53.

RYZL, M. (1968). Precognition scoring and attitude toward ESP. *Journal of Parapsychology,* **32,** 1–8.

RYZL, M. (1968). Precognition scoring and attitude. *Journal of Parapsychology,* **32,** 183–189.

CARPENTER, J. C. (1969). Further study on a mood adjective check list and ESP run-score variance. *Journal of Parapsychology,* **33,** 48–56.

DUVAL, P., & MONTREDON, E. (1969). Precognition in mice: A confirmation. *Journal of Parapsychology,* **33,** 71–72.

FREEMAN, J. A. (1969). The psi-differential effect in a precognition test. *Journal of Parapsychology,* **33,** 206–212.

FREEMAN, J. A. (1969). A precognition experiment with science teachers. *Journal of Parapsychology,* **33,** 307–310.

JOHNSON, M. (1969). Attitude and target differences in a group precognition test. *Journal of Parapsychology,* **33,** 324–325.

MONTREDON, E., & ROBINSON, A. (1969). Further precognition work with mice. *Journal of Parapsychology,* **33,** 162–163.

SCHMIDT, H. (1969). Precognition of a quantum process. *Journal of Parapsychology,* **33,** 99–108.

BENDER, H. (1970). Differential scoring of an outstanding subject on GESP and clairvoyance. *Journal of Parapsychology,* **34,** 272–273.

FREEMAN, J. A. (1970). Sex differences in ESP response as shown by the Freeman picture-figure test. *Journal of Parapsychology,* **34,** 37–46.

FREEMAN, J. A. (1970). Ten-page booklet tests with elementary-school children. *Journal of Parapsychology,* **34,** 192–196.

FREEMAN, J. (1970). Shift in scoring direction with junior-high-school students: A summary. *Journal of Parapsychology,* **34,** 275.

FREEMAN, J. A. (1970). Mood, personality, and attitude in precognition tests. *Journal of Parapsychology,* **34,** 322.

HARALDSSON, E. (1970). Subject selection in a machine precognition test. *Journal of Parapsychology,* **34,** 182–191.

HARALDSSON, E. (1970). Precognition of a quantum process: A modified replication. *Journal of Parapsychology,* **34,** 329–330.

NIELSEN, W. (1970). Relationships between precognition scoring level and mood. *Journal of Parapsychology,* **34,** 93–116.

SCHMIDT, H. (1970). Precognition test with a high-school group. *Journal of Parapsychology,* **34,** 70.

BELOFF, J., & BATE, D. (1971). An attempt to replicate the Schmidt findings. *Journal of the Society for Psychical Research,* **46,** 21–31.

HONORTON, C. (1971). Automated forced-choice precognition tests with a "sensitive." *Journal of the American Society for Psychical Research,* **65,** 476–481.

MITCHELL, E. D. (1971). An ESP test from Apollo 14. *Journal of Parapsychology,* **35,** 89–107.

SCHMIDT, H., & PANTAS, L. (1971). Psi tests with psychologically equivalent conditions and internally different machines. *Journal of Parapsychology,* **35,** 326–327.

STANFORD, R. G. (1971). Extrasensory effects upon "memory." *Journal of the American Society for Psychical Research,* **64,** 161–186.

STEILBERG, B. J. (1971). Investigation of the paranormal gifts of the Dutch sensitive Lida T. *Journal of Parapsychology,* **35,** 219–225.

THOULESS, R. H. (1971). Experiments on psi self-training with Dr. Schmidt's pre-cognitive apparatus. *Journal of the Society for Psychical Research,* **46,** 15–21.

HONORTON, C. (1972). Reported frequency of dream recall and ESP. *Journal of the American Society for Psychical Research,* **66,** 369–374.

JOHNSON, M., & NORDBECK, B. (1972). Variation in the scoring behavior of a "psychic" subject. *Journal of Parapsychology,* **36,** 122–132.

KELLY, E. F., & KANTHAMANI, B. K. (1972). A subject's efforts toward voluntary control. *Journal of Parapsychology,* **36,** 185–197.

SCHMIDT, H., & PANTAS, L. (1972). Psi tests with internally different machines. *Journal of Parapsychology,* **36,** 222–232.

CRAIG, J. G. (1973). The effect of contingency on precognition in the rat. *Research in Parapsychology 1972,* 154–156.

FREEMAN, J. A. (1973). The psi quiz: A new ESP test. *Research in Parapsychology 1972,* 132–134.

ARTLEY, B. (1974). Confirmation of the small-rodent precognition work. *Journal of Parapsychology,* **38,** 238–239.

HARRIS, S., & TERRY, J. (1974). Precognition in a water-deprived Wistar rat. *Journal of Parapsychology,* **38,** 239.

RANDALL, J. L. (1974). An extended series of ESP and PK tests with three English schoolboys. *Journal of the Society for Psychical Research,* **47,** 485–494.

EYSENCK, H. J. (1975). Precognition in rats. *Journal of Parapsychology,* **39,** 222–227.

HARALDSSON, E. (1975). Reported dream recall, precognitive dreams, and ESP. *Research in Parapsychology 1974,* 47–48.

HONORTON, C., RAMSEY, M., & CABIBBO, C. (1975). Experimenter effects in extrasensory perception. *Journal of the American Society for Psychical Research,* **69,** 135–149.

KANTHAMANI, H., & RAO, H. H. (1975). Response tendencies and stimulus structure. *Journal of Parapsychology,* **39,** 97–105.

LEVIN, J. A. (1975). A series of psi experiments with gerbils. *Journal of Parapsychology,* **39,** 363–365.

TERRY, J. C., & HARRIS, S. A. (1975). Precognition in water-deprived rats. *Research in Parapsychology 1974,* 81.

DAVIS, J. W., & HAIGHT, J. (1976). Psi experiments with rats. *Journal of Parapsychology,* **40,** 54–55.

JACOBS, J., & BREEDERVELD, H. (1976). Possible influences of birth order on ESP ability. *Research Letter* (Parapsychology Laboratory, University of Utrecht), No. 7, 10–20.

NEVILLE, R. C. (1976). Some aspects of precognition testing. *Research in Parapsychology 1975,* 29–31.

DRUCKER, S. A., DREWES, A. A., & RUBIN, L. (1977). ESP in relation to cognitive development and IQ in young children. *Journal of the American Society for Psychical Research,* **71,** 289–298.

HARALDSSON, E. (1977). ESP and the defense mechanism test (DMT): A further validation. *European Journal of Parapsychology,* **2,** 104–114.

SARGENT, C. L. (1977). An experiment involving a novel precognition task. *Journal of Parapsychology,* **41,** 275–293.

BIERMAN, D. J. (1978). Testing the "advanced wave" hypothesis: An attempted replication. *European Journal of Parapsychology,* **2,** 206–212.

BRAUD, W. (1979). Project Chicken Little: A precognition experiment involving the SKYLAB space station. *European Journal of Parapsychology,* **3,** 149–165.

HARALDSSON, E., & JOHNSON, M. (1979). ESP and the defense mechanism test (DMT) Icelandic study No. III: A case of the experimenter effect? *European Journal of Parapsychology,* **3,** 11–20.

O'BRIEN, J. T. (1979). An examination of the checker effect. *Research in Parapsychology 1978,* 153–155.

CLEMENS, D. B., & PHILLIPS, D. T. (1980). Further studies of precognition in mice. *Research in Parapsychology 1979,* 156.

HARALDSSON, E. (1980). Scoring in a precognition test as a function of the frequency of reading on psychical phenomena and belief in ESP. *Research Letter* (Parapsychology Laboratory, University of Utrecht), No. 10, 1–8.

SARGENT, C., & HARLEY, T. A. (1981). Three studies using a psi-predictive trait variable questionnaire. *Journal of Parapsychology,* **45,** 199–214.

WINKELMAN, M. (1981). The effect of formal education on extrasensory abilities: The Ozolco study. *Journal of Parapsychology,* **45,** 321–336.

NASH, C. B. (1982). ESP of present and future targets. *Journal of the Society for Psychical Research,* **51,** 374–377.

THALBOURNE, M., BELOFF, J., & DELANOY, D. (1982). A test for the "extraverted sheep versus introverted goats" hypothesis. *Research in Parapsychology 1981,* 155–156.

CRANDALL, J. E., & HITE, D. D. (1983). Psi-missing and displacement: Evidence for improperly focused psi? *Journal of the American Society for Psychical Research,* **77,** 209–228.

SCHWARTZ, S. A., & DEMATTEI, R. J. (1983). The Mobius Psi-Q test: Preliminary findings. *Research in Parapsychology 1982,* 103–105.

JOHNSON, M., & HARALDSSON, E. (1984). The Defense Mechanism Test as a predictor of ESP scores: Icelandic studies IV and V. *Journal of Parapsychology,* **48,** 185–200.

TEDDER, W. (1984). Computer-based long-distance ESP: An exploratory examination (RB/PS). *Research in Parapsychology 1983,* 100–101.

HESELTINE, G. L. (1985). PK success during structured and nonstructured RNG operation. *Journal of Parapsychology,* **49,** 155–163.

HARALDSSON, E., & JOHNSON, M. (1986). The Defense Mechanism Test (DMT) as a predictor of ESP performance: Icelandic studies VI and VII. *Research in Parapsychology 1985,* 43–44.

VASSY, Z. (1986). Experimental study of complexity dependence in precognition. *Journal of Parapsychology,* **50,** 235–270.

HONORTON, C. (1987). Precognition and real-time ESP performance in a computer task with an exceptional subject. *Journal of Parapsychology,* **51,** 291–320.

Psychophysical Research Laboratories
P. O. Box 569
Plainsboro, NJ 08536