Journal of Scientific Exploration, Vol. 5, No. 1, pp. 61–83, 1991
Pergamon Press plc. Printed in the USA.

0892-3310/91 $3.00+.00

Effects of Consciousness on the Fall of Dice: A Meta-Analysis

DEAN I. RADIN

Intelligent Systems Laboratory, Contel Technology Center, 15000 Conference Center Drive, Chantilly, VA 22021

DIANE C. FERRARI

Department of Psychology, Princeton University

Abstract—This article presents a meta-analysis of experiments testing the hypothesis that consciousness (in particular, mental intention) can cause tossed dice to land with specified targets face up. Seventy-three English language reports, published from 1935 to 1987, were retrieved. This literature describes 148 studies reported by a total of 52 investigators, involving more than 2 million dice throws contributed by 2,569 subjects. The full database indicates the presence of a physical bias that artifactually inflated hit rates when higher dice faces (e.g., the "6" face) were used as targets. Analysis of a subset of 59 homogeneous studies employing experimental protocols that controlled for these biases suggests that the experimental effect size is independently replicable, significantly positive, and not explainable as an artifact of selective reporting or differences in methodological quality. The estimated effect size for the full database lies more than 19 standard deviations from chance while the effect size for the subset of balanced, homogeneous studies lies 2.6 standard deviations from chance. We conclude that this database provides weak cumulative evidence for a genuine relationship between mental intention and the fall of dice.

Introduction

Iacta alea est. (The die is cast).

—Julius Caesar (100–44 B.C.)

Forty thousand years ago, primitive peoples believed that destiny could be revealed by casting bones, or influenced by sacrifice and prayer (Radin, 1957; Watson, 1988, p. 13). It appears that astragalomancy (divination by dice) was universally employed in ancient times, with evidence for "casting lots" ranging from African tribes, to the Inuit, to the Mayans. The related concepts of chance and destiny play significant roles in the beliefs of early peoples, as reflected, for example, in Siva, the Hindu god of a thousand

Acknowledgments. We thank two anonymous referees as well as Ephraim Schechter, Ed May, Chuck Honorton, and Roger Nelson for their valuable comments on an earlier version of this article.

names, who is portrayed in some statues as determining the fate of mankind by throwing dice (Harvie, 1973).

Today, sophisticated men and women still "roll the bones" in casinos, and still fervently wish for favorable destinies. The language, garb, and technology have changed, but the primeval urge to know and control fate remains the same.

Is it possible to mentally control how "the die is cast"? Many believe that the answer must be "no," for gambling casinos generally enjoy huge profits. However, casino profits are calculated by subtracting gamblers' wins from losses. With gambling wagers running in the hundreds of millions of dollars a year in many casinos, and gamblers losing more than winning on average, the gambling industry can still record a healthy profit. Gamblers who tend to win or lose consistently undoubtedly differ in innate mathematical abilities and memory skills, but one wonders whether some of the consistent winners occasionally violate chance expectation (assuming a fair game).

Beginning in 1935, researchers' have tested the hypothesis that the fall of dice may be influenced by mental intention. Over the next half century, 52 investigators[2] published (in English) the results of 148 such studies. The basic experiment is deceptively simple: A die face is prespecified, then a die (or group of dice) is tossed while the subject "wills" that face to turn up. If the subject's intention matches the resulting die face, a "hit" is scored. If more hits are obtained than expected by chance, this is taken as evidence for the hypothesis.

The relevant literature has been criticized and reviewed in detail numerous times (Edge et al., 1986; Girden & Girden, 1985; Girden, 1962, Girden et al., 1964; Murphy, 1962; Rhine, 1944), but in spite of continuing experiments and reviews, no clear consensus has emerged on the state of the hypothesis. Much of the ongoing controversy is undoubtedly the result of poor replicability of the hypothesized effect, and the fact that behind this straightforward task lies a wide array of methodological pitfalls (Barber, 1976), any one of which can legitimately cast doubt on experimental results.

Still another factor perpetuating the controversy is that previous reviews of this literature have generally focused more on the presence of real or imagined methodological flaws than on the empirical data itself. Studies examining how referees judge the adequacy of experimental reports have shown that prior beliefs about a hypothesis correlate highly with assessments of methodological adequacy (e.g., Nisbett & Ross, 1980). Thus, in general, skeptics tend to judge these experiments as methodologically flawed, while proponents tend to see the same set of experiments as methodologically adequate. This interaction between belief and assessment of flaws is important because skeptics characteristically dismiss any experiment they consider to be flawed, and hence, end up dismissing most if not all of the positive evidence. The

---

[1] Largely stimulated by the work of J. B. Rhine, L. Rhine, and their colleagues.
[2] "Investigators" are defined here as the combined number of unique authors.

present report uses the techniques of meta-analysis to address the question of methodological quality in a way that is more explicit and objective than the narrative review (cf. Glass, McGaw, & Smith, 1981; Rosenthal, 1984).

Method

Meta-Analysis

The quantitative literature review, also called meta-analysis, has become a valuable tool in the behavioral and social sciences. Meta-analysis is analogous to well-established procedures used in the physical sciences to determine physical parameters and constants. The technique assesses the issue of replication of a hypothesized effect within a body of studies by examining the distribution of "effect sizes" (Bangert-Drowns, 1986). In the present context, the null hypothesis (no mental influence on the fall of dice) specifies an expected mean effect size of zero. A homogeneous distribution of effect sizes with nonzero mean indicates replication of an effect, and the size of the deviation of the mean from its expected value estimates the magnitude of the effect.

Meta-analyses assume that effects being compared are similar across different experiments, that is, that all studies seek to estimate the same population parameters. Thus, the scope of a quantitative review must be strictly delimited to ensure appropriate commonality across the different studies that are combined (Glass, 1978). This can present a nontrivial problem because replication studies typically investigate a number of variables in addition to those studied in the original experiments. In the present case, because different subjects and experimental protocols were employed within the reviewed literature, some heterogeneity attributable to these factors was expected in the obtained distribution of effect sizes. However, the circumscription for the review required that every study in the database have the same primary goal (mental influence of the fall of dice) and, hence, would estimate the same underlying effect.

Circumscription of Meta-Analysis

This meta-analysis examined studies testing whether mental intention could influence the fall of dice and, in particular, cause a prespecified die face to land face up after being tossed. The usual measure of this "influence" is a statistical test that compares the number of obtained hits to chance expectation. Because this test is almost always reported as a standard normal deviate (i.e., z score) in the reviewed literature, we determined *effect size* as a z score normalized by the square root of the sample size (N), $e = z/\sqrt{N}$, where N was the total number of individual random events.

Excluded from this analysis were experiments where the goal was to cause dice to fall in specific locations (e.g., "placement" studies), or where the hypothesis involved looking at patterns of die falls (e.g., "quarter decline" and similar studies). Given this restrictive circumscription, it should be noted that J. B. Rhine (and well as other parapsychologists) considered that the *best* evidence for a mental effect on dice were consistent patterns observed in dice data rather than hits on prespecified die faces. Nevertheless, the present study examined only the die face studies in an attempt to evaluate an elementary hypothesis in a group of conceptually simple experiments. If the present meta-analysis showed no evidence for the elementary hypothesis, there would be less motivation to study more complicated hypotheses and experiments.

The Literature

The following journals were surveyed for studies fitting the above circumscription: *Journal of Parapsychology, European Journal of Parapsychology, Journal of the American Society for Psychical Research,* and *Journal of the Society for Psychical Research.* In addition, through the use of a specialized bibliographic database,[3] we retrieved other relevant studies from the following sources: *Research Letter* (of the University of Utrecht's Parapsychology Laboratory), *Newsletter of the Parapsychology Foundation, Proceedings of the Society for Psychical Research, Proceedings of the First International Conference of Parapsychological Studies, Journal of Experimental Psychology, Parapsychological Journal of South Africa,* and a book, *The Algonquin Experiments* (see the references for details). Abstracts of the proceedings of the annual conventions of the Parapsychological Association, as reported in the volumes *Research in Parapsychology* (for the years 1980–1987), were also examined.[4]

Unit of Study

A "unit of study" was defined as the largest possible aggregation of non-overlapping data collected under a single target (i.e., a predefined die face). For each unit of study, the following data were recorded: The number of subjects, the intended die face target, the aim (i.e., to hit or to miss the target), the number of dice tossed at once, the total number of dice thrown in the study (N), the number of hits (H), the probability of a single event (p),[5] and a series of quality criteria (described below). From H, N, and p, a standard normal deviate (z score, corrected for continuity) and an effect size ($e = z/\sqrt{N}$), were calculated. The sign of the z score (and e) was established according to whether the observed result matched the intended aim. Thus, a nega-

[3] Parapsychological Sources of Information Center, 2 Plane Tree Lane, Dix Hills, NY 11746.

[4] We may have missed a few studies reported in *Research* in *Parapsychology* for the years 1972–1979.

[5] The usual hit probability was $p = 1/6$, but in a few cases the target was more than one die face, such as faces 4, 5, or 6, in which case the hit probability was adjusted accordingly.

tive z score obtained under intention to "aim low" was recorded as a positive score.[6]

Control studies were defined as experiments performed under conditions in which (a) dice were tossed with no specified target aim, and (b) the condition was defined as such in the published report. This definition excluded some studies in which control data was collected by replicating an experiment with a different aim (e.g., aiming low for the same target, or aiming for faces 1 through 6 an equal number of times). These studies were coded as "protocol" controls.

Mean *Effect* Size

Given the per study effect size e, described above, the overall weighted mean effect size was calculated as $\bar{e} = \sum \omega_i e_i / \sum \omega_i$, where $\omega_i = N_i$, and i ranges from 1 to K studies. The standard error of $\bar{e}$ is $s_e = 1/\sqrt{\sum \omega_i}$. A quality-weighted effect size was determined as $\bar{e}_Q = \sum (Q_i \omega_i e_i) / \sum (Q_i \omega_i)$, where Q_i is the quality score (described below) assessed for study i. The standard error associated with $\bar{e}_Q$ is $se_Q = \sqrt{\sum (Q_i^2 \omega_i)/(\sum (Q_i \omega_i)^2}$.

Quality Criteria

The following quality criteria cover virtually all published criticisms of the present experiments, including the issues raised in Girden's (1962) critical review of dice experiments. The criteria were coded as being either present (1) or absent (0) in each report:

(A) Automatic recording—Die faces automatically recorded onto permanent medium (e.g., photographed onto film). This helped eliminate human recording errors, and reduced the possibility of data selection.
(B) Independent recording—Independent (human) data recorder, other than the experimenter. If present, this criterion helped increase the trustworthiness of the recorded data, and provided a separate record for later comparison and double checking.
(C) Data selection prevented—This refers to designs with sequential-frame photographic data recordings, or studies in which data was kept in bound record books, or some other method of ensuring that all data were used in the final analysis.
(D) Data double checked—Data was manually or automatically double checked for accuracy.
(E) Witnesses present—Witnesses were present during data recording to help reduce the possibility of mistakes or fraud.
(F) Control noted—A control study was mentioned, but no details were

[6] Six of 148 of experimental studies were low aim, i.e., the prescribed intention was to miss the target face.

published. This criterion reflects the fact that the author(s) recognized the importance of performing a control test.

(G) Local control—Control data obtained under the same conditions as the experiment, using the same subject(s) and same conditions, but with no specific mental effort applied to the dice. A local control is usually conducted at the same time as the experiment, sometimes alternated with experimental conditions.

(H) Protocol control—Study designed in a manner such that controls are inherently a part of the experiment (e.g., controls for dice bias, equal number of throws for each die face, equal number of throws for a single die face in both high- and low-aim conditions, etc.).

(I) Calibration control—Long-term randomness test; usually before and after an experimental series. This differs from a local control by its longer duration, and by not being conducted under the same setup and protocols as the experiment itself.

(J) Fixed run lengths—Optional stopping ruled out by prespecified design.

(K) Formal study—Prespecified methodology; analyses prespecified in advance of experimentation.

In addition to the above factors, the following criteria were coded with (arbitrarily assigned) partial credit:

(L) Dice toss *method*—
0.00 = dice tossed by hand,
0.50 = dice tossed by hand but bounced against a back wall,
1.00 = dice tossed by a cup or chute (semimechanical release),
2.00 = tossed automatically by machine.

(M) Subject type—
0.00 = persons selected on the basis of prior performance or individuals claiming special abilities,[7]
0.50 = experimenter(s) as sole subject,
0.75 = experimenter participated in experiment along with subjects,
1.00 = unselected subjects.

The maximum score a study could receive is the sum of the present/absent codes, plus 2 for the dice toss method and 1 for unselected subjects, for a total of 14. This relatively simple additive method has been shown to be an excellent predictor of overall methodological quality (Dawes, 1979). Note that this quality measure is dependent on the details that researchers describe in their publications, thus, given that some particulars may be left out of shorter

[7] This criterion is assigned zero credit to reflect the skeptical suspicion that persons claiming special abilities may be more induced to attempt fraud to back up their claim. Recall that these criteria are designed to reflect a worse-case assessment of experimental methodology. It could be argued that persons selected on the basis of prior performance should be assigned higher credit.

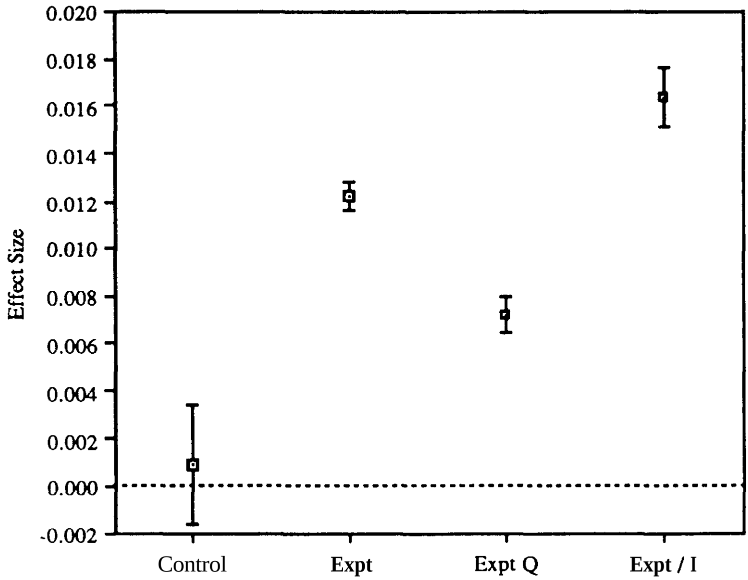

Fig. 1. Effect size means for controls, experiments, experiments weighted by quality, and for 24 experimenters who performed three or fewer studies. Error bars are one standard error.

reports and abstracts, the present quality measure is expected to underestimate "true" methodological quality.

Results

Overall Analysis

The literature search retrieved 73 relevant publications, representing the efforts of 52 investigators (40 different first authors) from 1935 to 1987. Over this 53-year period, a total of 2,569 subjects attempted to "influence" 2,592,817 dice-casts in 148 experimental studies, and 153,288 dice-casts in 31 control studies. Sample size (i.e., the number of dice tossed) per study ranged from 60 to 240,000 (median = 5,500); the number of subjects per study ranged from 1 to 393 (median = 3). Of the 148 studies, 44% were individually significant at the 5% level.

Figure 1 summarizes the overall results: a nonzero effect for experimental studies ($\bar{e}$ = .01220 a .00062),[8] and a control effect within chance limits ($\bar{e}$ = .00093 ± .00255, p = .715).[9] The overall level of statistical significance

[8] Weighted mean effect size estimate, plus and minus one standard error.

[9] Probabilities are estimates calculated with the "STAT PAC" inverse normal distribution function on the Hewlett-Packard HP41CX programmable calculator. Probabilities associated with effect sizes are determined from $z = e/se$, where se is the standard error.

(Stouffer z method) for experimental studies is $z = 18.2$, and for control studies, $z = 0.18$. Experimental effect size weighted by methodological quality is also nonzero ($\bar{e} = .00723 \pm .00071$). Note that the quality weighted effect size is significantly smaller than the nonweighted effect size. More about this later.

The number of studies conducted per investigator[10] ranged from 1 to 21, with the majority of investigators (64%) reporting one, two, or three studies. To examine the possibility that a minority of investigators reporting four or more studies may have skewed the mean effect size with exceptionally large scores, we calculated mean effect size for 25 investigators reporting three or fewer studies (n = 42 studies). The result was $\bar{e} = .01640 \pm .00126$, thus the overall effect size does not appear to be due to a few exceptional investigators.

Filedrawer Analysis

In the behavioral and social sciences it is well known that experiments with significant results tend to be published more often than nonsignificant studies (Hedges, 1984; Iyengar & Greenhouse, 1987; McNemar, 1960). In a meta-analysis, this reporting bias may inflate the estimate of effect size and overall statistical significance. This is called the "filedrawer effect." One way of assessing the effect of this presumed hidden filedrawer of nonsignificant studies is by calculating a "failsafe" number (see Rosenthal, 1984, p. 108). This is the number required to reduce the observed overall mean z score to a nonsignificant level (i.e., just below $z = 1.645$).

In the present analysis, the failsafe number is 17,974. This means it would take almost 18,000 additional studies averaging a null effect to reduce the observed results to a nonsignificant level. This is a ratio of 121 unretrieved, nonsignificant studies to each study retrieved in the meta-analysis. Rosenthal (1984, p. 110) suggests that an effect is robust with respect to the filedrawer if this ratio is 5:1 or greater. A filedrawer of 18,000 studies would require, for example, each of the 52 investigators involved in these dice experiments to have conducted one unpublished, nonsignificant study per month, every month, for 28 years.

There is another way to address the filedrawer problem. If investigators (or journal editors) tend to favor publication of statistically significant reports, we might expect the distribution of reported z scores to show a depressed number of studies reported just before the point of conventional significance, or a discontinuous "spike" starting at $z = 1.645$. Figure 2 plots the distribution of z scores; a suspected filedrawer "notch" is evident in the range $1.0 \leq z \leq 2.0$. This notch gives us a way to estimate the size of the "real" filedrawer.

---

[10] "Investigator" here refers to the first author of a publication; there were 40 different first authors in the present database.

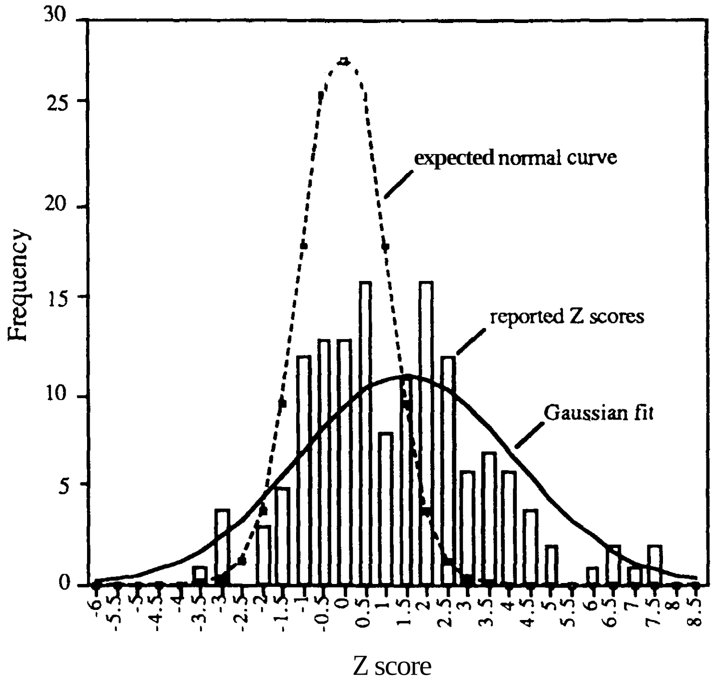

Fig. 2. Distribution of observed z scores, expected normal curve based on 148 studies, and best Gaussian fit to the observed distribution.

We begin by assuming that the positive tail of the curve (scores $z \geq 1.645$) does not hide a filedrawer effect. Since this portion is, by definition, 5% of a normal distribution, we can reconstruct the magnitude of the "entire" distribution. The number of studies in the positive tail is 65, thus the total distribution is 20×65, or 1,300 studies. Subtracting the 148 observed studies from this figure leaves our estimate for the size of the "true" filedrawer at 1,152, for a ratio of nine filedrawer studies for each observed study. Now we append to the reported z-score distribution an additional 1,152 z scores, selected at random from a normal distribution and bounded within the range $-1.645 \leq z \leq 1.645$. The Stouffer z score resulting from this filedrawer-adjusted database is 7.557 ($p < 10^{-13}$).

We can also add an additional 1,152 nonsignificant effect sizes to the original distribution to determine a new, filedrawer-weighted effect size. To form the nonsignificant effect sizes, we used the randomly selected z scores generated for the above filedrawer-adjusted database, as well as the same sample sizes observed in the original data. The procedure was to take the sample sizes from the 148 original studies and replicate that set eight times (subtracting 32 extra data points) until 1,152 samples sizes were generated. Then, effect sizes were calculated in the usual way ($z/\sqrt{N}$ for each study). The resulting overall mean effect size was $\bar{e} = .00165 \pm .00021$ ($p < 10^{-14}$). This distribution consisted of an overall sample size of some 23 million (simulated) dice tosses. Thus, it seems that the filedrawer problem cannot adequately explain the mean effect size or the overall level of statistical significance in this database.

It must be noted, however, that the above assessment of the filedrawer effect makes the implicit assumption that significantly negative studies are as likely to be reported as significantly positive studies. In fact, only eight studies were reported as significantly negative (compared to 65 positive), implying that significantly negative studies truly occurred some eight times less often than significantly positive studies, or that there was a reporting (or editorial) bias favoring the publication of positive studies. Which is a more plausible conclusion?

One way of addressing this issue is to compare studies published before and after the mid-1970s, for this is when the *Journal of Parapsychology* adopted the policy of publishing all studies, positive or negative, significant or nonsignificant (Broughton, 1987). The number of studies published after say, 1975, was 16. Of these, one was significantly negative and three were significantly positive. If we assume that reporting of post-1975 studies is *not* biased according to the outcome of the study (recognizing the fact that the post-1975 sample size is small), this would imply that a genuine "mental effect" may produce something like a 3:1 ratio of significantly positive to negative studies. If we reduce the full database ratio of 8:1 by three, we are still left with a 5:1 ratio, implying that the pre-1975 database indeed contains a reporting bias favoring studies reporting positive significance. Therefore, we must consider the overall (1935–1987) database as suspect with respect to the filedrawer problem.

Another way of examining the potential biasing effects of selective reporting is to examine the relationship between reported z scores and the year of publication. This relationship, shown in Figure 3, reveals that studies reported from the 1940s through the 1960s were substantially more significant than later studies. This suggests that (a) earlier studies were conducted in some significantly different fashion than later studies, (b) that the effect declined over the years, or (c) that the apparently greater magnitude of the earlier studies reflect an artifact of selective reporting. We believe that selective reporting is the more parsimonious explanation of this decline. (However, it is interesting to note that effect size does not significantly decline over time; slope $= -.001$, $t_{147} = 1.691$, $p = .093$.)

Homogeneity Analysis

If replication experiments were exact, and if the hypothetical ability were uniform in the population, the distribution of effect sizes might be expected to be homogeneous. However, such distributions are rarely homogeneous because replications are rarely intended to be exact. Instead, most experiments are designed to examine the same or similar phenomena under many

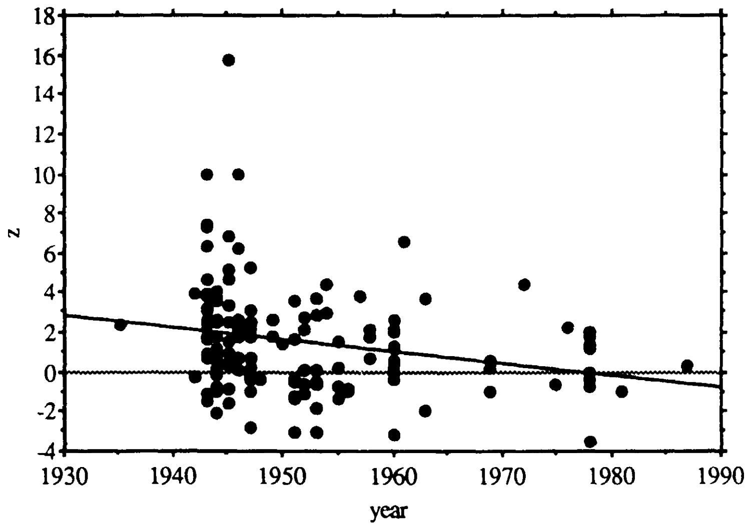

Fig. 3. Reported z scores declined over time (slope = $-.059$, $t_{147} = 3.151$, $p = .002$).

different conditions. Thus, to adjust for the heterogeneity introduced by different experimental conditions, and to reduce possibility that the overall mean effect size (and significance levels) may have been spuriously enlarged by extreme values, we deleted potential "outlier" studies as follows: If the homogeneity statistic[11] for all studies combined was significant (at the $p < .05$ level), the study that produced the largest reduction in this statistic was deleted. This procedure was repeated, reducing the database one study at a time, until the homogeneity statistic had become nonsignificant.

Results showed that out of the total of 148 studies, it was necessary to delete 52 (35%) to produce a homogeneous distribution of effect sizes. This may be compared with exemplary physical and social science reviews, where it is sometimes necessary to discard as many as 45% of the studies to achieve a homogeneous distribution of effect sizes (Hedges, 1987). Of the discarded studies, 33 had positive effect sizes (22% of the entire database) and 19 had negative effect sizes (13%). The mean weighted effect size with the remaining 96 studies is $\bar{e} = 0.00459 \pm .00087$. The combined significance level (Stouffer z) of these 96 studies is $z = 5.336$ ($p < 10^{-7}$). This homogeneity analysis suggests that the experimental effect is independently replicable.

Note that a filedrawer failsafe number is based on the assumption that a distribution contains all retrievable studies. When some studies are eliminated as statistical outliers, as in a homogenized distribution, we have by

[11] A test for homogeneity for K estimates of e_i is given by $H_K = \sum \omega_i(e_i - \bar{e})^2$, where H_K has a chi-square distribution with $K - 1$ degrees of freedom.

definition created a selected subset of studies, so one of the underlying assumptions of the filedrawer effect is violated. We assume that if the filedrawer effect is robust with respect to a full distribution, then it is also robust with respect to the homogenized form of that distribution. (However, in the present case, since the pre-1975 database is uncertain with respect to the filedrawer problem, the homogeneous subset is suspect as well.)

Temporal Analysis

Some critics assert that each new generation of parapsychologists starts from scratch, without acknowledging or benefitting from previous researchers' efforts (e.g., Marks, 1986). The assertion is part of an argument alleging that parapsychology is a pseudoscience, because it lacks a research tradition, like a "real" science. The assertion can be tested by examining assessed experimental quality over time (see Fig. 4). The trend is significantly positive, indicating that later researchers did not take note of earlier methodological criticisms. The critical allegation is not supported.

Quality Analysis

Critics have long maintained that experimental effects in parapsychology will decline as methodological quality improves, thereby reflecting a regression to the true (null hypothesis) mean effect size of zero. The relationship

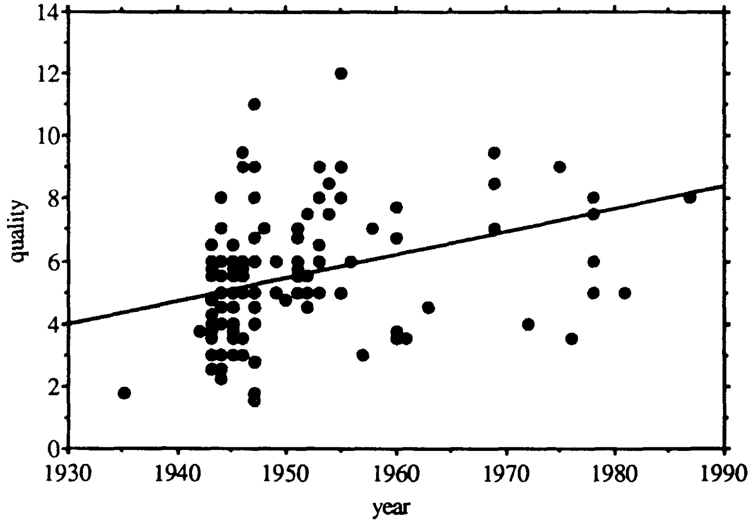

Fig. 4. Methodological quality improved over time (slope = .07, $t_{147} = 5.12$, $p < 10^{-5}$). [Some of the points in this and following graphs may overlap, thus the number of plotted points may appear to be fewer than indicated by the degrees of freedom. In all graphs it was confirmed that all data points are in fact plotted.]

between assessed methodological quality and experimental effect size in the full database is nearly significantly negative (slope = −.002, t_{145} = 1.845, p = .0672), but this decline is less apparent in the homogeneous database (Fig. 5).

However, the slopes from these linear regressions do not reflect the fact that we should have greater confidence in effect size estimates from studies with larger sample sizes and less confidence in studies with smaller sample sizes. Therefore, a more appropriate test would examine the slope of a weighted linear regression, where the weight reflects our confidence in the individual effect sizes. We use the inverse of the standard error as the weighting factor for each study (i.e., weight = $\sqrt{N_i}$, where N is the sample size from study i).

The slope (± 1 standard error) of a weighted linear regression for the entire database is −0.0029 ± 0.0010 (N = 148, z = slope/se = −2.874, p = .004). Thus, in the overall database, there is indeed evidence that effect size decreases as quality increases. This is also reflected in the fact that the quality-weighted effect size is about half the unweighted effect size (as shown in Fig. 1). However, this relationship is not significant in the homogeneous database: slope = −0.0006 + 0.0007 (N = 96, z = −0.8927, p = .372). Based upon the assumption that a homogeneous database provides a better basis upon which to assess the magnitude of an effect size, it appears that the effect is not affected by methodological quality. Still, it is prudent to keep in mind that the slope observed in the overall database supports the critics' assertion.

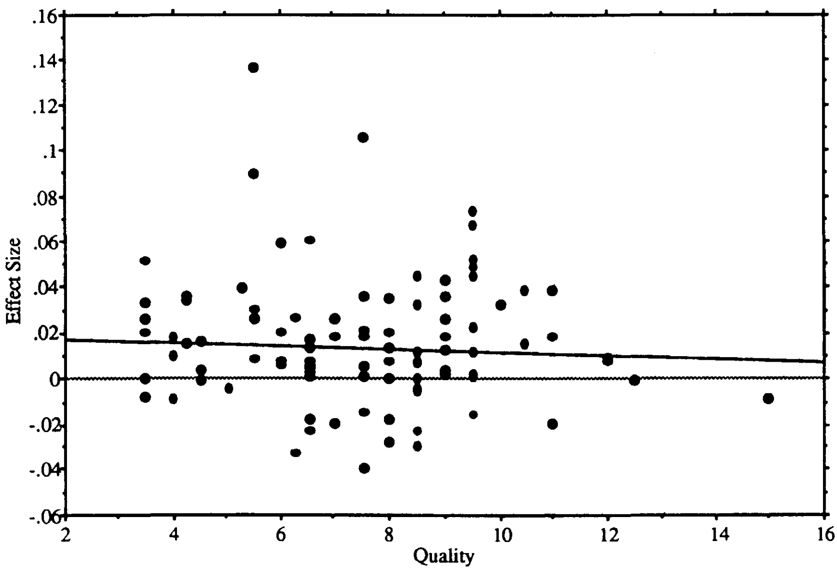

Fig. 5. **In the homogeneous database, effect size is essentially constant across methodological quality (slope** = −.001, t_{95} = .802, **p** = .4244).

Distributional Analysis

Some critics have claimed that some psi studies involving dice are flawed because the a priori probabilities of die faces are rarely equal (Scarne, 1974).[12] These biases arise because die faces on most common dice are marked by scooping out small bits of material. For example, the "6" face has six scoops removed which cause it to have slightly less mass than the other die faces. On any random toss, this bias makes the 6 most likely to land face up, followed by the 5, 4, 3, 2, and 1 face. Thus, an experiment exclusively using the 6 face as the target may be flawedbecause we could not tell whether the reported results were due to a consciousness-mediated influence, or the (potentially) higher probability of a 6.

To see whether this suspected dice bias was present in the current database, we examined all reports where published data allowed the separate calculation of effect size for each die face tossed under experimental and control conditions. The biases (Fig. 6) are found in these studies, as indicated by a high correlation between the experimental and control curves ($r = 0.826$), as well as a significantly positive slope of a linear regression between the two curves (slope = 0.514, $t_4 = 2.935$, $p = .0426$).

However, the observed experimental effect sizes also suggest the presence of an anomaly beyond the artifacts caused by dice biases. For example, in Figure 6 we see that for each die face the experimental effect size is larger than the control; and after a Bonferroni adjustment, we find that die face 6 is significantly larger ($z = 3.765$, $p = .0002$, two-tailed[13]).

But, regardless of this suggestive difference, the presence of any artifact in this group of studies raises doubts about the overall database. Therefore, we examined a subset of studies that controlled for these dice biases—studies employing design protocols where die faces were equally distributed among the six targets. We refer to such studies as the "balanced protocol subset.

Balanced Protocol Analysis

Sixty-nine experimental studies employed protocols in which targets were evenly balanced among the six die faces. These studies (Table 1) show (a) evidence for the "mental influence" hypothesis, both in terms of effect size and overall level of statistical significance; (b) an effect size that is relatively constant across different measures of methodological quality; and (c) a file-drawer requiring a 20:1 ratio of unretrieved, nonsignificant studies for each observed study.

However, as mentioned in the analysis of the entire database, it is observed that of these 69 studies, 23 were reported as significantly positive while only

---

[12] It should be noted that J. B. Rhine and his colleagues were among the first to recognize and take into account the dice bias problem.

[13] The statistic used to determine significance of the difference between means was $z = (e_E - e_C)/\sqrt{se_E^2 + se_C^2}$.

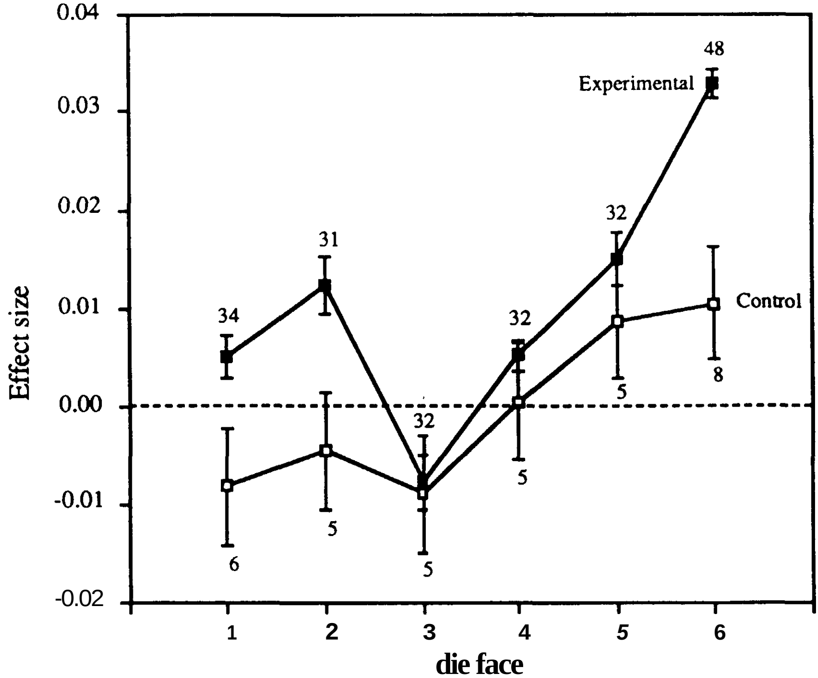

Fig. **6.** Relationship between die face and effect size for experimental and control conditions. The numbers indicate the number of studies contributing to the effect size estimate. Error bars are one standard errors.

one was reported as significantly negative. The large excess of significant positive studies raises the possibility of a reporting bias, so to examine this issue, we calculated the relationship between reported *z* scores and time of publication. A linear regression results in a virtually flat slope (slope $= -.002$, $t_{67} = .136$, p $= .892$); a regression between effect size and time is also nonsignificant (slope $= .0003$, $t = 1.109$, p $= .272$). Therefore, there is reason

TABLE 1
Analysis of subset of studies using balanced target protocols. The quality vs. effect size slope is calculated using a weighted linear regression, as discussed above in the section on Quality Analysis

		Probability (Two-tailed)
Number of studies	**69**	
Effect size	.00861 ± .00110	$<10^{-14}$
Quality-weighted effect size	.00706 ± .00115	$<10^{-9}$
Quality vs. effect size slope	−.00081 ± .00007	.279
Stouffer *z*	**7.617**	$<10^{-14}$
Filedrawer failsafe number	**1410**	

TABLE 2
Analysis of homogeneous subset of studies using balanced target protocols

		Probability (Two-tailed)
Number of studies	59	
Effect size	.00315 ± .00121	.009
Quality-weighted effect size	.00294 ± .00126	.020
Slope for quality vs. effect size	−.00013 ± .00024	.603
Stouffer z	3.188	.001

to believe that this subset of the entire database does not appear to be seriously compromised by a selective reporting artifact.

A similar analysis was performed on a homogenized, balanced distribution (achieved after discarding 10, or 14% of the studies). Results summarized in Table 2 show (a) evidence for the hypothesized effect in a homogeneous subset of experiments, suggesting that the effect is replicable; and (b) the effect size is essentially constant across different measures of methodological quality.

Efect of Different Quality Weighting Schemes

One might argue that the method used to assess experimental quality was nonoptimal because it gave the same (unit) weight to the different criteria used to rate quality. For example, one might assert that the "balanced protocol" quality criterion should be weighted more heavily than, say, the "data doubled-checked" criterion, because without a balanced protocol one could not be sure how much of the final effect was due to inherent dice biases. The same goes for the other quality criteria—some are arguably more important than others, and the overall quality assessment should take such differences into account.

To judge the effects of different weighting schemes, we asked 11 scientists[14] to assign importance weights to each of the quality criteria. A scale of 1 to 5 was used, where 1 indicated that a criterion was of relatively low importance in establishing a credible dice experiment, and 5 indicated critical importance. Table 3 shows the means and standard deviations of their combined responses, listed in descending order by weight.

On observing the values in Table 3, notice that there is relatively good agreement on which criteria are judged to be important in establishing a credible dice experiment, and which criteria are judged as not so important.

[14] This included seven members of the Parapsychological Association, two members of the Society for Scientific Exploration, an experimental physicist, and a computer scientist. All were familiar with standard experimental methodologies.

TABLE 3
Means and standard deviations for quality criteria weighting factors

Criterion	Mean Weight	SD
Dice-machine	4.73	0.65
Protocol control	4.45	0.69
Fixed run lengths	4.27	0.79
Automatic recording	4.09	0.83
Formal study	4.00	1.18
No data selection	3.82	1.17
Local control	3.82	1.40
Double check data	3.82	1.40
Independent recording	3.64	1.03
Dice-cup	3.45	1.13
Witnesses present	3.36	1.36
Calibration	3.18	1.40
Subjects—unselected	2.73	1.35
Subjects—experimenter and unselected	2.64	1.12
Subjects—selected	2.55	1.29
Subjects—experimenter only	2.27	1.42
Dice—bounced	2.00	1.34
Control noted	1.73	1.35
Dice—by hand	1.00	0.63

This agreement is reflected in the fact that more extreme weights generally resulted in smaller standard deviations. To reflect these agreements, we created a simplified weighting scheme in which criteria rated with a mean of 4 or above were assigned a 5, and the rest were assigned a 1. These were called *extreme* weights, whereas the means from the survey were called ***opinion*** weights.

To assess the effect of these new weighting schemes on the meta-analysis, we produced three arrays of quality scores for all 148 studies: one array for the original method, and one for each of the new weighting schemes. Then we examined the correlation matrix among these arrays. Table 4 shows that these three methods produced similar overall scores. This confirms the expectation that a simple combination of unit weights on various elementary variables can be as good as (and sometimes better than) so-called expert opinion (Dawes, 1979).

TABLE 4
Correlation matrix for three methods of weighting experimental quality (N = 148)

	Unit	Opinion	Expert
Unit	1	—	
Extreme	.872	1	—
Opinion	.916	.957	1

Potential Moderating Variables

Further analysis revealed a possible moderator variable: effect size in the homogeneous database (N = 96) differed according to the type of subjects tested (Fig. 7). This suggests that unselected subjects did not perform as well as persons selected on the basis of prior testing. Honorton & Ferrari (1989) reported similar findings in a meta-analysis of forced-choice precognition tests.

Discussion

Reviews of dice experiments have raised numerous arguments why the evidence for a "mental influence" hypothesis should not be taken seriously (e.g., Girden, 1962, 1985). Besides raising an assortment of methodological and statistical criticisms that can affect all experiments, several dice-specific criticisms are often mentioned. One is the fact that the *a priori* probabilities of die faces are rarely equal (Scarne, 1974); a related problem is that die face probabilities tend to change as dice are used (Scarne, 1974). In addition, the effects of environmental variables on dice behavior (such as temperature and humidity) can be difficult to predict. Nevertheless, by controlling for these and other factors that can potentially introduce statistical biases, a rigorous dice experiment can be performed.

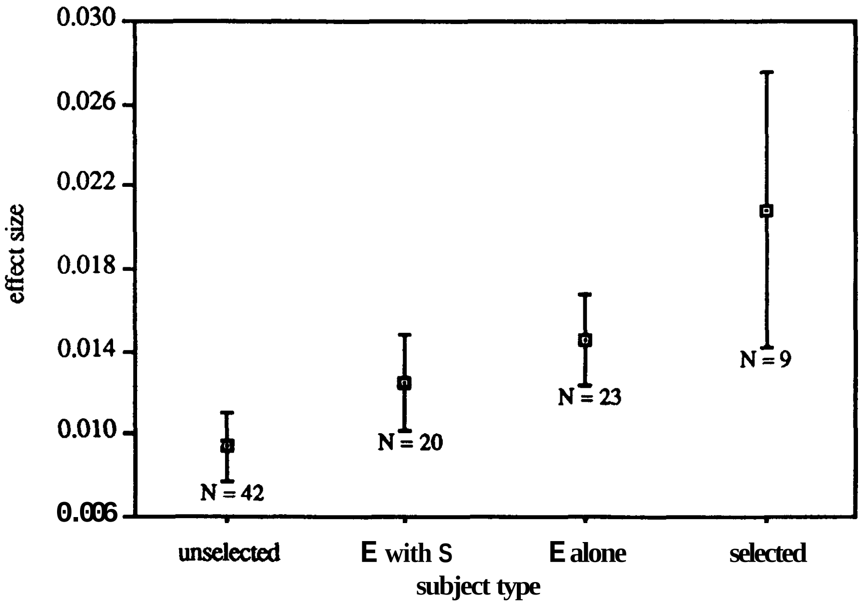

Fig. 7. Effect size by the type of subject, with standard error bars. "Unselected" were volunteers; "E with S" means that experimenters participated as subjects in the experiment along with other subjects; "E alone" means that experimenters were the sole subjects; and "Selected" were subjects chosen on the basis of prior testing." N is the number of studies used to determine the effect size.

In analyzing the full database, an overall experimental effect was observed to be more than 19 standard errors from chance[15] while the control effect size was within 1 *se*. After partialing out those first authors who had reported three or fewer studies, the experimental effect size was found to be 13 *se*, thus the effect could not be explained as an artifact of a few experimenters who had reported exceptionally significant results. A filedrawer analysis indicated that nearly 18,000 additional nonsignificant studies would be required to nullify the observed results, thus selective reporting would not be a reasonable explanation for the effect.

Next, we addressed the possibility that the results were primarily caused by studies that can be identified as statistical outliers. This was examined by creating and analyzing a homogeneous distribution of effect sizes. This reduced the overall effect size from 13 to 5 *se*. This 5 *se* homogeneous result suggested that the anomaly is both genuine and replicable.

The next step was to see whether effect sizes would decline as experimental quality improved—the "regression-to-the-mean" criticism. If such a relationship were observed, it would suggest that the "perfect" experiment would consistently demonstrate null results. Analysis of the full database, using a linear regression weighted by experimental sample size, did lend support to the regression-to-the-mean hypothesis, however analysis of the homogeneous subset did not.

The next analysis examined the data for evidence of biases that can be attributed to slight differences of mass among the different dice faces. Evidence for this bias was found, prompting an analysis of a subset of the database in which such biases were controlled for by experimental protocol. In this subset of 69 "balanced protocol" studies, the effect size was some 7 *se*, the effect did not significantly covary with quality, and the filedrawer effect required 20 nonsignificant studies for each observed study. To achieve a homogenized subset of balanced studies, it was necessary to discard 10 outliers. This lowered the effect to 2.6 *se*, but no significant relationship with quality was observed.

The last analysis addressed the question of whether a different weighting scheme might have provided quality assessment scores that better reflect expert opinion. This analysis indicated that weights based upon expert opinion, an extreme weighting scheme, and unit weighting all produced similar scores, thus the original quality assessment method was deemed to be an adequate measure of experimental quality.

Conclusion

This meta-analysis examined a straightforward hypothesis: Mental intention is correlated with the fall of specified die faces. Based primarily on the analysis of a homogeneous subset of balanced protocol studies, we conclude

[15] Hereafter, the phrase "standard errors from chance" is designated as "*se*."

that the aggregate evidence suggests the presence of a weak, genuine mental effect.

Many parapsychologists consider the best evidence for a mental intention effect on dice to be consistent patterns observed in the data rather than simple counts of direct hits on die faces. Given that we believe the present meta-analysis provides some evidence for a mental effect on dice, we recommend that future meta-analyses of dice studies concentrate on examining these claimed patterns.

The possible mental effect evaluated in this study resembles similar consciousness-related effects observed in both microscopic (Radin & Nelson, 1989; Schmidt, 1987) and macroscopic (Dunne, Nelson, & Jahn, 1988) random physical systems.

References

Bangert-Drowns, R. L. (1986). Review of developments in meta-analytic method. *Psychological Bulletin, 99,* 388–399.

Barber, T. X. (1976). *Pitfalls in human research: Ten pivotal points.* Elmsford, NY: Pergamon.

Broughton, R. S. (1987). Publication policy and the *Journal of Parapsychology.* Journal of Parapsychology, ***51,*** 21–32.

Dawes, R. M. (1979). The robust beauty of improper linear models in decision making. *American Psychologist, 34,* 571–582.

Dunne, B. J., Nelson, R. D., & Jahn, R. G. (1988). Operator-related anomalies in a random mechanical cascade. *Journal of Scientific Exploration, 2,* 155–180.

Edge, H. L., Morris, R. L., Palmer, J., & Rush, J. H. (1986). *Foundations ofparapsychology.* New York: Routledge & Kegan Paul.

Girden, E. (1962). A review of psychokinesis (PK). *Psychological Bulletin,* 59, 353–388.

Girden, E., & Girden, E. (1985). Psychokinesis: Fifty years afterward. In P. Kurtz (Ed.), *A skeptic's handbook ofparapsychology* (pp. 129–146). Buffalo, NY: Prometheus Press.

Girden, E., Murphy, G., Beloff, J., Flew, A., Rush, J. H., Schmeidler, **G.,** & Thouless, R. H. (1964). A discussion of "A review of psychokinesis (PK)." *International Journal of Parapsychology, 6,* 26–137.

Glass, G. V. (1978). In defense of generalization. *Behavioral and Brain Sciences, 3,* 394–395.

Glass, G. V., McGaw, B., & Smith, M. L. (1981). *Meta-analysis in social research.* Beverly Hills, CA: Sage.

Harvie, R. (1973). Probability and serendipity. In A. Hardy et al., (Eds.), *The challenge of chance.* London: Hutchinson.

Hedges, L. V. (1984). Estimation of effect size under nonrandom sampling: The effects of censoring studies yielding statistically insignificant mean differences. *Journal of Educational Statistics, 9,* 61–86.

Hedges, L. V. (1987). How hard is hard science, how soft is soft science? *American Psychologist, 42,* 443–455.

Honorton, C., & Ferrari, D. *C.* (1989). "Future telling": A meta-analysis of forced-choice precognition experiments, 1935–1987. In J. Palmer (Ed.), *Proceedings of Presented Papers of the Parapsychological Association 32nd Annual Convention,* San Diego, CA, pp. 110–121.

Iyengar, S., & Greenhouse, J. B. (1987). Selection models and the file-drawer problem. *Technical Report 394,* Department of Statistics, Carnegie-Mellon University.

Marks, D. F. (1986). Investigating the paranormal. *Nature, 320,* 119–124.

McNemar, **Q.** (1960). At random: Sense and nonsense. *American Psychologist,* ***15,*** 295–300.

Nisbett, R., & Ross, L. (1980). *Human inference: Strategies and shortcomings of social judgem nt.* Englewood Cliffs, NJ: Prentice-Hall.

Murphy, G. (1962). Report on paper by Edward Girden on psychokinesis. *Psychological Bulletin, 59,* 520–528.

Radin, D. I., & Nelson, R. D. (1989). Consciousness-related effects in random physical systems. *Foundations of Physics, 19,* 1499–1514.

Radin, P. (1957). *Primitive religion.* New York: Dover Publications.
Rhine, J. B. (1944). "Mind over matter" or the PK effect. *Journal of the American Society for Psychical Research, 38,* 185–201.
Rosenthal, R, (1984). *Meta-analytic procedures for social research.* Beverly Hills, CA: Sage.
Scarne, J. (1974). *Scarne on dice.* Harrisburg, PA: Stackpole Books.
Schmidt, H. (1987). The strange properties of psychokinesis. *Journal of Scientific Exploration, 1,* 103–118.
Utts, J. M. (1988). Successful replication vs. statistical significance. *Journal of Parapsychology, 4,* 305–320.
Watson, L. (1988). *Beyond supernature.* New York: Bantam Books.

Appendix

Publications Used in the Meta-Analysis

Averill, R. L., & Rhine, J. B. (1945). The effect of alcohol upon performance in PK tests. *Journal of Parapsychology, 9,* 32–41. [2][16]
Breederveld, H. (1976). Towards reproducible experiments in psychokinesis: I. Experiments with dice. *Research Letter, 7,* 1–5a. [1]
Carington, W. (1935). Preliminary experiments in precognitive guessing. *Journal of the Society for Psychical Research, 29,* 86–104. [1]
Cox, W. E. (1951). The effect of PK on the placement of falling objects. *Journal of Parapsychology, 15,* 40–48. [6]
Dale, L. A. (1946). The psychokinetic effect: The first A.S.P.R. experiment. *Journal of the American Society for Psychical Research, 40,* 123–151. [1]
Dale, L. A., & Woodruff, J. L. (1947). The psychokinetic effect: Further A.S.P.R. experiments. *Journal of the American Society for Psychical Research, 41,* 65–82. [3]
Feather, S. R., & Rhine, L. E. (1969). PK experiments with same and different targets. *Journal of Parapsychology, 33,* 213–227. [3]
Fisk, G. W., & West, D. J. (1957). Psychokinetic experiments with a single subject. *Newsletter of the Parapsychology Foundation, 4,* 3–7. [1]
Fisk, G. W., & West, D. J. (1958). Dice-casting experiments with a single subject. *Journal of the Society for Psychical Research, 39,* 277–287. [2]
Forwald, H. (1961). A PK experiment with die faces as targets. *Journal of Parapsychology, 25,* 1–12. [1]
Forwald, H. (1963). An experiment in guessing ESP cards by throwing a die. *Journal of Parapsychology, 27,* 16–22. [2]
Gatling, W., & Rhine, J. B. (1946). Two groups of PK subjects compared. *Journal of Parapsychology, 10,* 120–125. [1]
Gibson, E. P. (1947). Note on an impromptu experiment in psychokinesis. *Journal of the American Society for Psychical Research, 41,* 22–28. [2]
Gibson, E. P. (1948). An exploratory PK experiment based upon throwing twenty-four dice. *Journal of Parapsychology, 12,* 289–295. [1]
Gibson, E. P., & Rhine, J. B. (1943). The PK effect: III. Some introductory series. *Journal of Parapsychology, 7,* 118–134. [4]
Gibson, E. P., Gibson, L. H., & Rhine, J. B. (1943). A large series of PK tests. *Journal of Parapsychology, 7,* 228–237. [2]
Gibson, E. P., Gibson, L. H., & Rhine, J. B. (1944). The PK effect: Mechanical throwing of three dice. *Journal of Parapsychology, 8,* 95–109. [3]
Greene, F. M. (1960). The feeling of luck and its effect on PK. *Journal of Parapsychology, 24,* 129–141. [2]
Herter, C. J., & Rhine, J. B. (1945). An exploratory investigation of the PK effect. *Journal of Parapsychology, 9,* 17–25. [1]

[16] The number of experiments reported in the reference is shown in brackets.

Hilton, H., Baer, G., & Rhine, J. B. (1943). A comparison of three sizes of dice in PK tests. *Journal of Parapsychology, 7,* 172–190. [3]

Hilton, H., & Rhine, J. B. (1943). A second comparison of three sizes of dice in PK tests. *Journal of Parapsychology, 7,* 191–206. [2]

Humphrey, B. M. (1947). Help-hinder comparison in PK tests. *Journal of Parapsychology, 11,* 4–13. [1]

Humphrey, B. M. (1947). Simultaneous high and low aim in PK tests. *Journal of Parapsychology, 11,* 160–174. [2]

Humphrey, B. M., & Rhine, J. B. (1945). PK tests with two sizes of dice mechanically thrown. *Journal of Parapsychology, 9,* 124–132. [1]

Hyde, D. H. (1945). A report on some English PK trials. *Proceedings of the Society for Psychical Research, 47,* 293–296. [1]

Kelly, E. F., & Kanthamani, B. K. (1972). A subject's efforts toward voluntary control. *Journal of Parapsychology, 36,* 185–197. [1]

Knowles, E. A. G. (1949). Report on an experiment concerning the influence of mind over matter. *Journal of Parapsychology, 13,* 186–196. [2]

Lyndon, R., & Rose, R. (1951). Psi experiments with Australian Aborigines, *Journal of Parapsychology, 15,* 122–131. [1]

Mangan, G. L. (1954). A PK experiment with thirty dice released for high- and low-face targets. *Journal of Parapsychology, 18,* 209–218. [2]

McConnell, R. A. (1953). Some unpublished evidence regarding the occurrence of psychokinesis. *Proceedings of the First International Conference of Parapsychological Studies,* p. 12–13. [1]

McConnell, R. A. (1955). Remote night tests for PK. *Journal of the American Society for Psychical Research, 49,* 99–108. [2]

McConnell, R. J., Snowdon, R. J., & Powell, K. F. (1955). Wishing with dice. *Journal of Experimental Psychology, 50,* 269–275. [2]

Mitchell, A. M. J., & Fisk, G. W. (1953). The application of differential scoring methods to PK tests. *Journal of the Society for Psychical Research, 37,* 45–61. [4]

Nash, C. B. (1944). PK tests of a large population. *Journal of Parapsychology, 8,* 304–310. [2]

Nash, C. B. (1946). Position effects in PK tests with twenty-four dice. *Journal of Parapsychology, 10,* 51–57. [1]

Nash, C. B. (1956). An exploratory analysis for displacement in PK. *Journal of the American Society for Psychical Research, 50,* 151–157. [2]

Nash, C. B. (1981). High- and low-dice psychokinesis with hidden targets. *Parapsychological Journal of South Africa, 2,* 11–17. [1]

Nash, C. B. (1987). A blind-target PK die test with future targets. *Journal of the Society for Psychical Research, 54,* 257–260. [1]

Nash, C. B., & Richards, A. (1947). Comparison of two distances in PK tests. *Journal of Parapsychology, 11,* 269–282. [1]

Nicol, J. F., & Carington, W. (1947). Some experiments in willed die-throwing. *Proceedings of the Society for Psychical Research, 48,* 164–175. [5]

Osis, K. (1953). A test of the relationship between ESP and PK. *Journal of Parapsychology, 17,* 298–309. [2]

Parsons, D. (1945). Experiments on PK with inclined plane and rotating cage. *Proceedings of the Society for Psychical Research, 47,* 296–300. [5]

Pegram-Reeves, M., & Rhine, J. B. (1943). The PK effect: I. *Journal of Parapsychology, 7,* 76–93. [1]

Pegram-Reeves, M., & Rhine, J. B. (1943). The PK effect: II. A study of declines. *Journal of Parapsychology, 7,* 118–134. [1]

Pegram-Reeves, M., & Rhine, J. B. (1945). The PK effect: The first doubles experiment, *Journal of Parapsychology, 9,* 42–51. [1]

Pope, D. H. (1946). Bailey's comparison of a coin and a die in PK tests, *Journal of Parapsychology, 10,* 213–215. [1]

Pratt, J. G., & Woodruff, J. L. (1946). An exploratory investigation of PK position effects. *Journal of Parapsychology, 10,* 197–207. [1]

Price, M. M., & Rhine, J. B. (1944). The subject-experimenter relation in the PK test. *Journal of Parapsychology, 8,* 177–186. [2]

Rae, J. P. (1978). *The Algonquin experiments.* Hazeldean, Ontario: Canadian Institute of Parapsychology. [12]
Ratte, R. J. (1960). Comparison of game and standard PK testing techniques under competitive and noncompetitive conditions. *Journal of Parapsychology, 24,* 235–244. [6]
Ratte, R. J., & Greene, F. M. (1960). An exploratory investigation of PK in a game situation. *Journal of Parapsychology, 24,* 159–170. [5]
Rhine, J. B. (1943). Dice thrown by cup and machine in PK tests. *Journal of Parapsychology,* 7, 207–217. [3]
Rhine, J. B. (1944). The PK effect: Early singles tests. *Journal of Parapsychology,* 8,287–303. [2]
Rhine, J. B. (1945). Early PK tests: Sevens and low-dice series. *Journal of Parapsychology, 9,* 106–115. [1]
Rhine, J. B. (1946). Hypnotic suggestion in PK tests. *Journal of Parapsychology, 10,* 126–140. [2]
Rhine, J. B. (1946). The Schwartz PK experiment. *Journal of Parapsychology, 10,* 208–212. [1]
Rhine, J. B., & Humphrey, B. M. (1943). The PK effect: The McDougall one-die series. *Journal of Parapsychology, 7,* 252–263. [1]
Rhine, J. B., & Humphrey, B. M. (1944). PK tests with six, twelve, and twenty-four dice per throw. *Journal of Parapsychology, 8,* 139–157. [4]
Rhine, J. B., & Humphrey, B. M. (1945). Position effects in the six-by-six series of PK tests. *Journal of Parapsychology, 9,* 296–302. [1]
Rhine. J. B., & Humphrey, B. M. (1945). The PK effect with sixty dice per throw. *Journal of Parapsychology, 9,* 203–218. [1]
Rhine, J. B., Humphrey, B. M., & Averill, R. L. (1945). An exploratory experiment on the effect of caffeine upon performance in PK tests. *Journal of Parapsychology,* 9, 80–91. [2]
Rhine, L. E., & Rhine, J. B. (1943). The psychokinetic effect: I. The first experiment. *Journal of Parapsychology, 7,* 20–43. [3]
Rose, R. (1950). Some notes on a preliminary PK experiment with six dice. *Journal of Parapsychology, 14,* 116–126. [1]
Rose, R. (1952). Experiments in ESP and PK with Aboriginal subjects. *Journal of Parapsychology, 16,* 219–220. [2]
Rose, R. (1955). A second report on psi experiments with Australian Aborigines. *Journal of Parapsychology, 19,* 92–98. [1]
Stanford, R. G. (1969). "Associative activation of the unconscious" and "visualization" as methods for influencing the PK target. *Journal of American Societyfor Psychical Research, 63,* 338–351. [1]
Steilberg, B. J. (1975). "Conscious concentration" versus "visualization" in PK tests. *Journal of Parapsychology, 39,* 12–20. [1]
Thouless, R. H. (1951). A report on an experiment on psycho-kinesis with dice, and a discussion of psychological factors favouring success. *Proceedings of the Society for Psychical Research, 49,* 107–130. [3]
Van De Castle, R. L. (1958). An exploratory study of some personality correlates associated with PK performance. *Journal of the American Society for Psvchical Research. 52,* 134–150. [1]
Vasse, P., & Vasse, C. (1951). A comparison of two subjects in PK. *Journal of Parapsychology,* 15, 263–270. [1]
Woodruff, J. L., & Rhine, J. B. (1942). An experiment in precognition using dice. *Journal of Parapsychology, 6,* 243–262. [2]