Meta-analysis of mind-matter interaction experiments: 1959 to 2000

Dean Radin & Roger Nelson

Boundary Institute, Los Altos, California
Princeton Engineering Anomalies Research, Princeton University

Laboratory experiments examining the possibility of direct mind-matter interactions have been reported for over a century. Two classes of such experiments reported most frequently include tossing dice while maintaining the intention for certain die faces to appear, and mental influence of truly random bits generated by electronic random number generators (RNG). Earlier meta-analyses of publications reporting dice and RNG experiments published up to 1987 provided strong statistical evidence for mind-matter interaction phenomena. We conducted an update of the RNG experiment literature to see if the evidence persists.

The updated RNG review covered all known studies from the first published in 1959 to the most recent published in mid-2000. We found a total of 515 experiments published in 216 articles by 91 different first authors, of which 423 were published through 1987, and 92 published after 1987. The magnitude of the overall effect size per experiment is small, on average less than the equivalent of 1% for binary RNGs, but statistically the overall effect is more than 16 standard errors from chance. The average z score for studies published up to 1987 is $\bar{z}$ = .73 and for studies published after 1987 is $\bar{z}$ = .61. The difference in average z scores is not significant (p = 0.48), indicating that the meta-analytic evidence for mind-matter interaction effects persists.

A conservative estimate of the effect of selective reporting practices (the "filedrawer problem") indicates that to reduce the observed statistical outcome to chance, each of the 91 researchers would have had to conduct but not report 29 additional, nonsignificant experiments. Variations in methodological quality did not correlate with experimental results (r = 0.03, p = .26), but quality did significantly improve over time (r = 0.50, p = 10^{-34}). We conclude that the RNG experiments continue to provide persuasive statistical evidence for independently repeatable mind-matter interaction effects observed under controlled conditions.

Introduction

Underlying the concept of "distant healing" is an assumption that goes beyond the present understanding of mind-body interactions within a given body, as in the discipline of psychoneuroimmunology. Distant healing requires some form of mind-matter interaction (MMI) at a distance. While not widely known to the medical and scientific mainstreams, a substantial empirical literature has addressed whether MMI is possible in principle by studying the effects of mental intention on inanimate physical systems. These studies, conducted by researchers around the world for nearly a century, have examined MMI effects on, for example, morphological changes in thin strips of metal (Randall & Davis, 1982; Sasaki, Shigemi & Yasuo, 1982; Hasted & Robertson, 1980), the distribution of metallic and plastic balls (Forward, 1977; Cox, 1974; Nelson, et al. 1983), temperature changes in well-shielded environments (Schmeidler, 1973, 1984), the statistical behavior of spinning coins (Thouless, 1945; Binski, 1958), latencies in radioactive decay (Chauvin & Genthon, 1965; Ollmar & Tengstrand, 1976), and perturbations in sensitive magnetometers (Puthoff & Targ, 1976) and interferometers (Jahn, 1982).

While many of these studies produced interesting results, by far two classes of MMI experiments have been conducted most often: The first involves tossing dice while intending certain die faces to appear, the second involves mental interaction with random numbers. In dice experiments, individuals typically toss one or more dice while wishing for pre-specified faces to appear. In most of the published studies the dice were held in cups, or were tossed by a machine to avoid the possibility of manual manipulation. The statistical evaluation was based on the number of resulting matches to the target face compared to the number of dice tossed. A meta-analysis found 73 relevant experiments published from 1935 to 1987, reflecting the efforts of 52 different principal investigators (Radin & Ferrari, 1991). Most (74%) of these studies were conducted in the 1940s and 1950s. The publications describe a total of 2.6 million dice-throws in 148 experiments, and just over 150,000 dice-throws in 31 studies where no mental influence was applied to the dice (as a control test).

This set of dice experiments produced a small overall effect (an average of 1.2% over chance expectation), but statistically this was more than 18 standard errors from chance. The control results were well within chance expectation. Examination of ten factors of methodological quality showed that variations in reported study designs were not correlated with the outcomes, and that experimental quality significantly improved with time. In addition, the "filedrawer" or selective reporting problem could not plausibly explain the results.

A second class of MMI experiments began in 1959. These involved mental influence of random numbers generated by truly random number generators (RNG). RNGs are electronic circuits designed to produce a sequence of random bits at the press of a button. After generating a sequence of say, 100 bits, the number of 1's in the sequence might be provided as feedback. A single "run" in such an experiment might consist of an observer being asked to mentally intend an RNG to generate, in two successive button presses, a high number (i.e., the number of 1 bits was greater than chance expectation of 50), and then a low number (less than 50). This may be followed by a control condition in which no directional intention is applied. An experiment might consist of a group of individuals, each contributing a hundred such runs, or one individual contributing several thousand runs. Outcomes of these experiments are often statistically expressed in terms of *z* scores, or standard normal deviates from chance. For meta-analytic purposes, outcomes in these studies were evaluated one-tailed, i.e., looking for deviations from chance in the direction of the pre-defined mental intention. Thus reported *z* scores are recorded for trials defined as high-aim, whereas sign-reversed *z* scores are recorded for trials defined as low-aim.

A 1989 meta-analysis of RNG experiments located 597 experiments and 235 control studies. These were described in 152 articles published from 1959 to 1987 by 68 different principal investigators (Radin & Nelson, 1989). Of these 597 experiments, 258 were reported in a long-term RNG study from the Princeton University PEAR laboratory, which also reported 127 of the control studies.

The overall statistical outcome was small in magnitude (the equivalent of a 0.9% shift of the 50% chance expected ratio of 1s to 0s in a binary RNG), but this shift of the mean was more than 12 standard errors from chance. As in the dice studies, no significant effects could be attributed to variations in experimental quality, and the filedrawer problem could not plausibly explain these results. The present paper updates the RNG literature and compares pre and post-1987 results.

Results

A literature review found 64 new publications describing 176 RNG experiments that were not retrieved in the earlier meta-analysis (for convenience this will be referred to as "MA-1989").[1] Of these 176 experiments, 84 were reported up to 1987 and 92 after 1987. The new publications included a description of the 20-year PEAR RNG program, thus the 258 PEAR lab experiments reported separately in MA-1989 were collapsed into a single datapoint for the purposes of the present (MA-2000) analysis. This resulted in combining 339 non-PEAR experiments from the MA-1989 database along with 176 new studies, for a total of 515 studies reported by 91 different principal investigators in 216 publications. Figure 1 graphs the number of RNG experiments reported annually from 1959 to 2000.

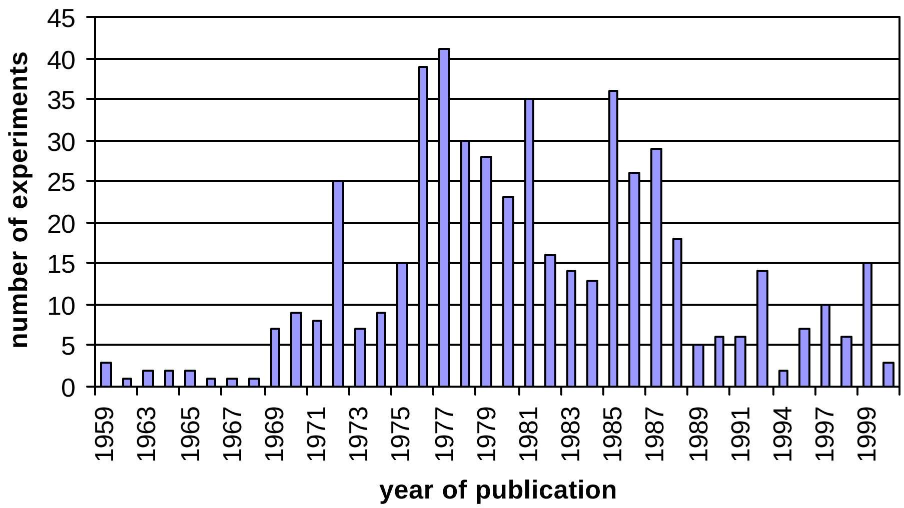

Figure 1. Number of RNG experiments published annually.

The RNG experiments in MA-1989 were independently assessed for methodological rigor by the first and second author on a quantitative 16-point scale[2]. For MA-2000 methodological quality was judged based on a subjective 5-point scale by the second author. To allow a pooled examination of the effects of experimental quality across all experiments, the MA-1989 16-point quality scores were transformed into equivalent 5-point scores.

Besides the quality score, for each experiment the total number of random samples (referred to as "N") and the statistical outcome were recorded (in the form of *z* scores). Experiments reported only as "nonsignificant" without further details were assigned *z* scores of 0. The value of N across all RNG studies ranged from a low of 312 individual random events to a high of 390 million random events. The total N was just over 1.4 billion.

---

[1] See the Appendix for a description of the new analysis.
[2] These included factors such as the presence or absence of control tests, use of failsafe equipment, redundant data recording, data permanently archived, data selection prevented by protocol or equipment, fixed run lengths, and tamper resistant hardware.

The average effect size per random event over these 515 studies, expressed in terms of a percentage over chance expectation assuming a binary RNG, was 0.7%. Overall this cumulated to 16.1 standard errors from chance ($p \ll 10^{-50}$). Figure 2 shows the chronological history of this cumulative *z* score.[3]

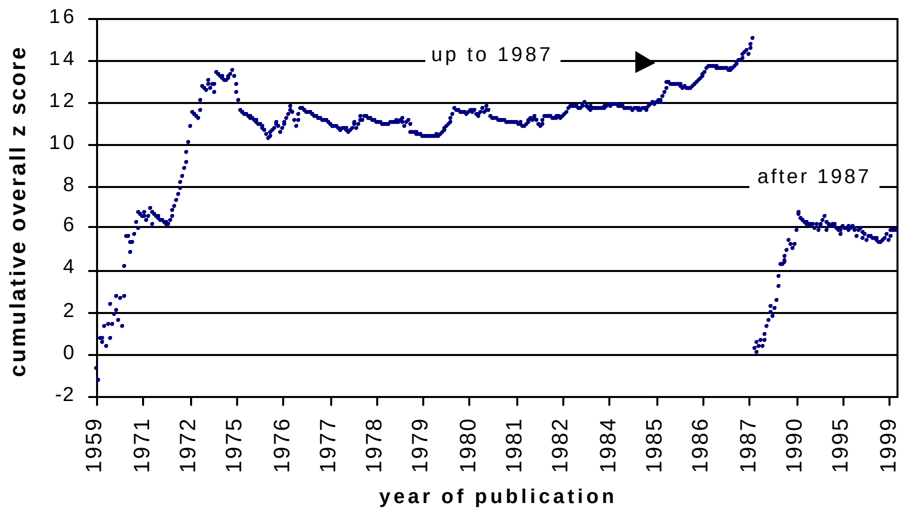

Figure 2. Cumulative *z* score for RNG experiments pre- and post-1987.

Figure 3 shows the cumulative mean *z* score for 423 studies reported up to and including 1987 ($\bar{z}$ = 0.73, combined *z* = 15.1), and for the 92 post-1987 studies ($\bar{z}$ = 0.61, combined *z* = 5.86). A t-test for the difference in the terminal mean *z* scores is not significant (p = 0.48), indicating that experiments published after MA-1989 continue to provide similar evidence for MMI effects.

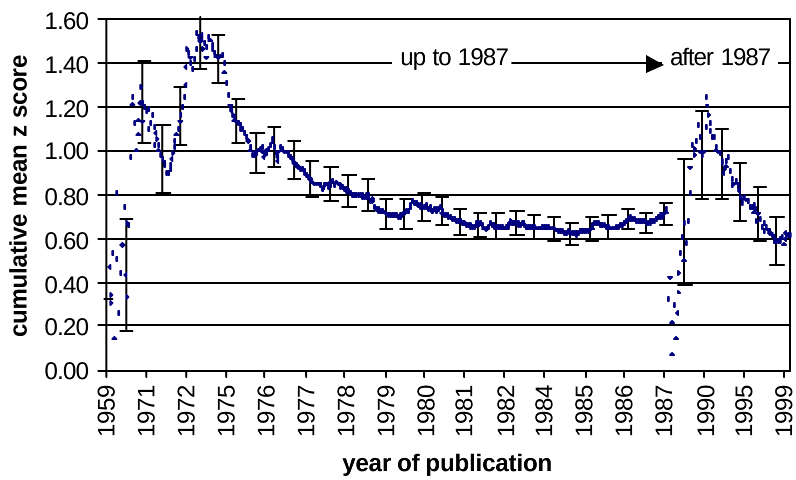

Figure 3. Chronological mean *z* scores for RNG experiments pre- and post-1987, with one standard error bars.

[3] *z* scores were combined by Stouffer's method.

Approximately 50% of the 515 studies (259) were reported by 10 of the 91 investigators. Their average z score was $\bar{z} = 0.96$ (combined $z = 15.5$). For the remaining 81 investigators, $\bar{z} = 0.46$ (combined $z = 7.3$). Of that group, 23 authors reported a single RNG study, $\bar{z} = 0.65$ (combined $z = 3.1$). This suggests that the overall non-chance result was not attributable to peculiar methods employed by a small group of investigators.

Estimating the filedrawer effect

While a 16 standard error effect is impressively significant, it is likely that some of this effect can be attributed to selective reporting practices. This is because authors tend not to write-up and editors tend not to publish experiments reporting statistically nonsignificant results. This publication bias inflates a meta-analytic assessment of a set of studies, as a meta-analysis is based upon published (or otherwise retrievable) reports.

In the worst case, published articles are composed only of significant studies, i.e., where the outcomes are reported as $p < 0.05$ or $z > 1.645$. In this case, the filedrawer distribution would be composed only of nonsignificant studies, i.e., where $p > 0.05$ or $z < 1.645$ (Scargle, 2000). To assess the effects of this worst-case model, a set of 10,000 z scores were generated[4] and the mean of the distribution below $z = 1.645$ was determined. This value was $\bar{z}' = -0.105$. From this we can determine how many studies averaging $\bar{z}'$ would be required to bring the combined statistical outcome down from the observed 16.1 standard errors to just above significance, i.e., to $z < 1.645$.

The answer is 2,610 studies. This is a ratio of 5.1 unpublished, nonsignificant, and slightly negative studies to each experiment retrieved in the meta-analysis. Following Rosenthal (1991), we may call the observed meta-analytic result "robust" with respect to the selective reporting problem.[5] Note that this worst-case scenario for the filedrawer estimate assumes that *each* of the 91 authors reporting at least one RNG study had also conducted an additional 29 nonsignificant studies averaging $\bar{z} = -0.105$.

A more realistic model for the filedrawer effect may be formed by examining the actual distribution of 515 z scores, as shown in Figure 4. The anomalous spike of observed studies at $z = 0$ is due to the fact that some studies reported only as "nonsignificant" were conservatively assigned $z = 0$ in the meta-analysis.

[4] Using the Microsoft Excel 2000 random number generator routine for a normal distribution.
[5] Rosenthal recommended a 5 to 1 ratio for calling a combined effect robust.

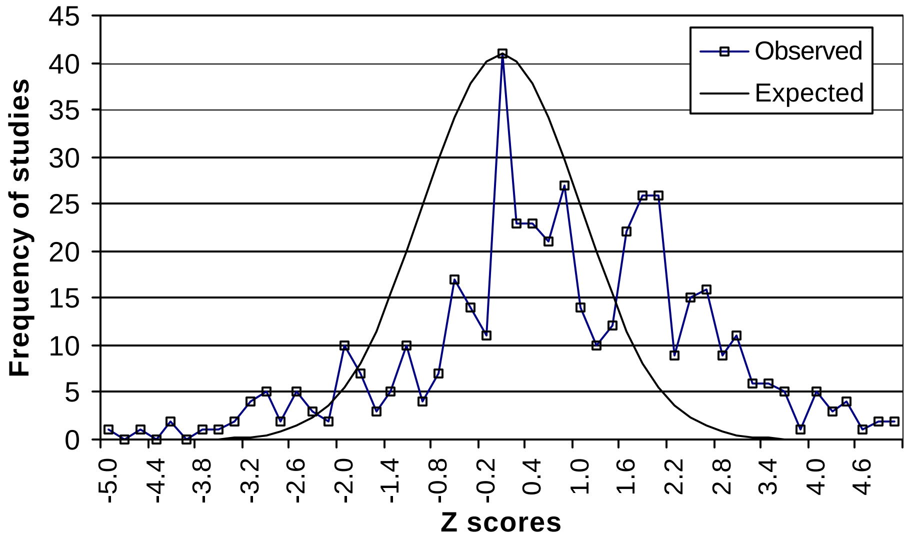

Figure 4. Distribution of observed and chance-expected RNG experiments.

Figure 4 also shows a substantial surplus of z scores starting at $z = 1.65$, which is precisely what we would expect from a selective reporting bias. However, we also see a surplus of z scores starting at $z = -2$. This indicates that when experiments proved to be significantly *negative* at $p < .05$, two-tailed (i.e., at $z < -1.96$), that they also tended to be published. This is supported by the fact that at $z \leq -2$ the observed number of studies (39) is far more than the expected number, $p \approx 10^{-10}$. This suggests that a more realistic model for the filedrawer is a z score distribution bounded by the ranges $-1.96 < z < 1.65$.

The average z for a distribution truncated in this fashion, determined by examining a distribution of 10,000 random z scores, is $\bar{z}' = -0.046$. Again we ask, How many studies averaging $\bar{z}'$ would be required to bring the overall statistical outcome down to $z < 1.645$? The answer is 5,240, or a ratio of 10.2 unpublished to each published experiment. This would have required each of the 91 authors to have conducted approximately 58 studies, averaging $\bar{z} = -0.046$, and not to have published any of them. This seems implausible.

Another way to assess the filedrawer effect is to examine the combined statistical result of observed studies falling within the presumed filedrawer range of $-1.96 < z < 1.65$. There were 320 such studies, with a combined statistical effect of 2.59 standard errors from chance ($p < 0.005$). In other words, the cumulative effect of all published *nonsignificant* studies is in itself significantly positive. This again indicates that the distribution of observed studies is genuinely shifted beyond chance and that the overall result is not due to a selective reporting problem. The next question is whether design flaws might have been responsible for the observed result.

Quality assessment

If there were a negative correlation between measures of methodological quality and experimental outcomes, it would imply that apparently significant results were actually due to flaws in experimental design. Alternatively, significant studies tend to be reported in more detail than nonsignificant studies, in which case we might expect to see a positive correlation between study quality and outcomes. If instead no correlation were observed between quality and study outcomes, then it would suggest that methodological quality was not a significant factor in producing the non-chance results.

Figure 5 plots the correlation between experimental quality and reported *z* scores. Table 1 shows that this correlation is nonsignificant (p = 0.255). The same correlation versus effect size (defined as the percentage over chance expectation of 50% for a binary RNG) is also nonsignificant (p = 0.109). Thus, the outcomes of the RNG studies were not due to variations in design quality. Figure 6 shows that experimental quality substantially improved with time.

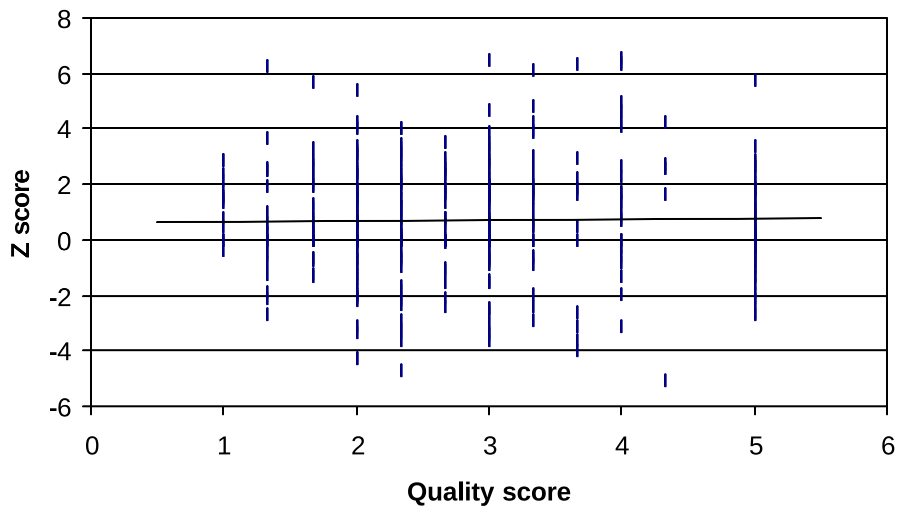

Figure 5. Relationship between quality and *z* score.

	Quality vs z	Quality vs effect size per bit
r	0.029	-0.056
t	0.659	-1.234
N	515	490
p	0.255	0.109

Table 1. Relationships between study quality and *z* scores, and quality vs. effect size. There are 490 studies listed for the correlation with effect size, as 25 experiments were insufficiently detailed to determine sample size and could not be included in the correlation.

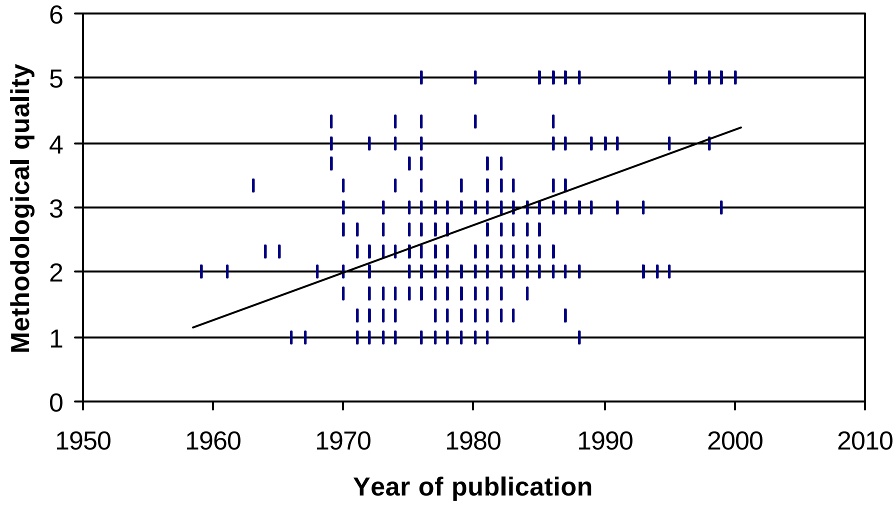

Figure 6. Relationship between year of publication and experimental quality, r = 0.502, p = 10^{-34}.

Discussion

The experiments discussed here were considered in the context of evaluating claims of distant healing effects. The results of this meta-analysis support the existence of genuine mind-matter interaction effects, which in turn supports the plausibility of distant healing.

A common assumption about RNG experiments is that MMI effects "operate" on individual random events. If this assumption were true, then the statistical results would increase proportionally with $\sqrt{N}$, where N is the number of bits per experiment. That is, the correlation between $\sqrt{N}$ and z should be positive. Indeed, the number of random events used per experiment has significantly increased over the years (r = 0.35, p = 10^{-15}), reflecting both an increase in the speed with which bits can be generated in modern RNG circuits, and the expectation that more bits per experiment would lead to more significant results .

However, the observed correlation between $\sqrt{N}$ and z is r = -0.015, p = 0.36. This means the statistical effects observed in these experiments are effectively independent of sample size, and cannot be explained as simple, linear, force-like mechanisms. Thus, if MMI effects in RNG experiments legitimately model similar effects in living systems, then it is likely that distant healing effects also cannot be explained as simple causal process.

Further indication that a novel approach will be required to explain these effects are experiments strongly resembling RNG studies, but involving pre-recorded random bits rather than bits generated in real-time. Those studies show significant cumulative results similar to those reported here (Bierman, 1998). This implies that some MMI effects, perhaps including those claimed for distant healing, may involve acausal processes.

Conclusions

Meta-analysis of 515 RNG experiments conducted by 91 researchers over a span of 41 years indicates the presence of a small magnitude, but statistically highly significant and repeatable mind-matter interaction effect. The overall results cannot be attributed to chance, or selective reporting problems, or variations in design quality. These studies indicate that there are ways in which mind and matter interact that support the plausibility of distant intentional healing. Because modern RNG experiments can be conducted under tightly controlled laboratory conditions at relatively low cost, they may serve as a convenient model to help us better understand the relevant conditions and mechanisms of distant healing.

References

Bierman, D.J. (1998). Do psi phenomena suggest radical dualism? In: Stuart R. Hameroff, Alref W. Kaszniak, and Alwyn C. Scott (Eds), *Toward a science of Consciousness II*, The MIT Press, Cambridge MA, pp. 709-714.

Binski, S. (1958). Performance by a single subject in exploratory PK experiments. *Zeitschrift fur Parapsychologie und Grenzgebiete der Psychologie*, 2(1), 30-40.

Chauvin, R., and Genthon, J-P. (1965). An investigation of the possibility of PK experiments with uranium and a geiger-counter. *Zeitschrift fur Parapsychologie und Grenzgebiete der Psychologie*, 8 (3), 140-147.

Cox, W. E. (1974). PK tests with a thirty-two channel balls machine. *Journal of Parapsychology*, 38(1), 56-68.

Forwald, H. (1977). A PK experiment with a single ball rolling on a decline. *European Journal of Parapsychology*, 2(1), 4-14

Hasted, J.B., and Robertson, D. (1980). Paranormal action on metal and its surroundings. *Journal of the Society for Psychical Research*, 50(784), 379-398

Jahn, R. G. (1982). The persistent paradox of psychic phenomena: An engineering perspective. *Proceedings of the IEEE*, **70**, 136-170.

Nelson, R. D., Dunne, B. J., and Jahn, R. G. "A Psychokinesis Experiment with a Random Mechanical Cascade," Technical Note PEAR 83002, Princeton Engineering Anomalies Research, Princeton University, School of Engineering/Applied Science, June 1983.

Ollmar, S. and Tengstrand, G. (1976) PK on radioactive decay. *Research Letter*, No. 7, 23-26

Puthoff, H. E. & Targ, R. (1976). A perceptual channel for information transfer over kilometer distances: Historical perspective and recent research, *Proceedings of the IEEE*, **64**, 329-354.

Radin, D.I. and Nelson, R.D. (1989). Evidence for Consciousness-Related Anomalies in Random Physical Systems. *Foundations of Physics*, **19**, 1499-1514.

Radin, D. I. & Ferrari, D. C. (1991). Effects of consciousness on the fall of dice: A meta-analysis. *Journal of Scientific Exploration*, 5, 61-84.

Randall, J. L., and Davis, C.P. (1982). Paranormal deformation of nitinol wire: A confirmatory experiment. *Journal of the Society for Psychical Research*, 51(792), 368-373

Rosenthal, R. (1991). *Meta-analytic procedures for social research* (Rev. ed.). Newbury Park, CA: Sage.

Sasaki, K. S., and Ochi, Y. (1982). Observation of the deformation of pure aluminum plates by psychokinesis. *Journal of the Psi Science Institute of Japan*, 7(1), 7-13.

Scargle, J. D. (2000). Publication Bias: The “File-Drawer Problem” in Scientific Inference. *The Journal of Scientific Exploration*, **14**(1), 91-106.

Schmeidler, G. R. (1973). PK effects upon continuously recorded temperature. *Journal of the American Society for Psychical Research*, 67(4), 325-340

Schmeidler, G. R. (1984). Further analyses of PK with continuous temperature recordings. *Journal of the American Society for Psychical Research*, 78(4), 355-362

Thouless, R. H. (1945). Some experiments on PK effects in coin spinning. *Proceedings of the Society for Psychical Research*, 47(170), 277-281

Appendix

by Roger Nelson

The RNG literature was first examined in 1987 for a meta-analysis ultimately published in 1989 (Radin and Nelson, 1989). Though our search at the time was reasonably thorough, we did not locate all the relevant papers, and in the meantime many more RNG studies have been conducted. Dr. Fiona Steinkamp, at the Institut für Grenzgebiete der Psychologie und Psychohygiene (IGPP) in Freiburg, Germany, has undertaken a comprehensive update and extension of the original RNG meta-analysis, and she provided us with a complete set of papers which were not included in the 1987 survey, or were published later.

A preliminary meta-analysis of these additional papers was prepared as part of a presentation of the state of the art of mind-matter interaction research at the Science and Spirituality in Healing conference in Winston-Salem, NC, October 26-29, 2000. This update of the RNG meta-analysis is less formal than the 1987 effort and should be regarded as an approximation that can later be compared to the more exacting analysis intended by the IGPP team. In particular, the quality assessment for this effort used a five-point subjective rating scale, rather than the quantitative count of specific methodological features we had originally employed. The ratings were based on the same methodological criteria as in the 1987 analysis although the features were not explicitly counted. Judgements were made independently for each paper without consideration of possible implications for the intended use in various correlational studies. The relationship of study quality to effect size in the new body of experiments will eventually be treated more rigorously by the IGPP group, but we felt it was important to provide a preliminary estimate of quality to address this important meta-analytic question.

The new database includes a total of 65 additional articles describing RNG studies and 5 papers discussing a new experimental variant called "Fieldreg." The latter studies examine correlations between RNG outputs and implicit variations in attention within groups of people rather than explicit assignment of intention in individuals.

The new experiments were reported by 36 principal authors and include published articles, technical reports, theses, reports in proceedings, and abstracts in journals or books. The total of 70 papers includes 183 RNG studies and 22 Fieldreg studies, for a total of 205 reported outcomes. Some 27 of these papers, comprising 67 studies, were found with publication dates prior to 1987, which had not been discovered in the search at that time. Most of those were from conference proceedings or non-English-language journals.

It is important to note that the great majority of the post-1987 database were "process-oriented" experiments, in which two or more conditions (identified as separate studies in our meta-analytic database) were compared. This means that some of the conditions were not expected to produce anomalous deviations because the intention of the experiment was to compare differences among conditions rather than the simple effects of intention on RNG behavior. The present meta-analysis was not designed to address questions about underlying processes. Rather, we wished to assess the cumulative evidence for consciousness-related anomalies in the RNG literature. Therefore, a decision was made at the outset to disregard the implicit intentions embedded in complex, process-oriented experimental designs, and to regard each study or data subset in the same way, as if the participants always intended to achieve an anomalous effect in the direction of nominal intention. This is a conservative approach because it ignores predicted differential effects, and it penalizes any successful effort to reduce deviation as may be required in certain experimental conditions.