Journal of Scientific Exploration, Vol. 20, No. 3, pp. 361–374, 2006 0892-3310/06

Assessing the Evidence for Mind-Matter Interaction Effects

DEAN RADIN

Consciousness Research Laboratory,
Institute of Noetic Sciences, Petaluma, CA

ROGER NELSON AND YORK DOBYNS

Princeton Engineering Anomalies Research Laboratory,
Princeton University, Princeton, NJ

JOOP HOUTKOOPER

Center for Psychobiology and Behavioral Medicine,
Justus-Liebig-University of Giessen, Giessen, Germany

Abstract—Experiments suggesting the existence of mind-matter interaction (MMI) effects on the outputs of random number generators (RNG) have been criticized based on the questionable assumption that MMI effects operate uniformly on each random bit, independent of the number of bits used per sample, the rate at which bits are generated, or the psychological conditions of the task. This "influence-per-bit" assumption invariably leads to the conclusion that the significant cumulative results of these experiments, as demonstrated in meta-analyses, are due not to MMI effects but rather to publication biases. We discuss why this assumption is doubtful, and why publication bias and other common criticisms of MMI-RNG studies are implausible.

Keywords: meta-analysis—random number generator—mind-matter interaction—critiques

We appreciate Martin Schub's interest in the Radin and Nelson (1989, 2003) meta-analyses of mind-matter interaction (MMI) experiments involving random number generators (RNGs). His paper is one of several critiques of the RNG meta-analyses that have appeared in recent years. Three of those re-examined data used in previous RNG meta-analyses; they include Scargle (2000), Ehm (2005), and now Schub (this issue). A fourth article is a new meta-analysis by Bösch et al. (in press). All of these critiques and analyses are commendable. Fresh examinations of controversial data are useful in leading to more refined analyses and clearer interpretations of the results.

Schub's critique raises three major points and a number of minor ones. The first major point involves a debatable assumption about the nature of MMI and the consequences of that assumption. The other major points are two common critiques of meta-analyses, especially those involving controversial topics. They involve the possible biasing effects of publication bias and assertions about low experimental quality. Because these latter two points are notoriously difficult to

assess objectively, the resulting ambiguity allows the critic to dismiss meta-analytic outcomes as insufficiently persuasive. This predictably leads to calls for a highly repeatable, "perfect" experiment to settle the issue. The appeal of a proof-positive experiment is understandable, but it is also entirely unrealistic. No experiment can ever achieve that level of perfection. Why then do some scientists persist in exploring the frontier in spite of its uncertainties, while others remain stubbornly skeptical in the face of ever-accumulating evidence? We address this important question, which is of special interest to readers of this journal, in our concluding remarks.

What is MMI?

Schub assumes, as have others (Bösch et al., in press; Ehm, 2005) that MMI "operates" by uniformly influencing each generated random bit regardless of the number of bits used per sample, the rate at which bits are produced, or the psychological conditions of the task. He expresses this key assumption through his surprise that different weighting schemes lead to different statistical results, and with phrases such as this "is not what one would expect from a stationary, repeatable effect."

The problem with this assumption is that there is no valid reason to expect that MMI should behave in this way. Indeed, we are unaware of any sort of human performance that is unaffected by such parametric changes. For example, a factory worker who can identify defective parts with 100% accuracy as they roll by at one or two per second will do no better than chance if the conveyor belt is suddenly accelerated to 1,000 parts per second. And yet the bit rate in the various MMI-RNG experiments range over not just three orders of magnitude, but six orders, and they involve significant physical differences in the underlying mechanisms for generating random bits.

Then there are psychological considerations. Let's say we conduct two RNG experiments. The first involves 1,000 highly experienced meditators, each of whom is selected based on his or her performance on previous, similar RNG tasks. Each participant is asked by a cordial, enthusiastic investigator to engage in a daily intention-focusing practice for six months in preparation for the experiment. When the experiment occurs, the participant is asked to intentionally influence the generation of a single truly random bit. The outcome of that one random decision will be provided as feedback either as a moment of silence (counted as a miss) or as a breathtaking fireworks display (a hit). If the fireworks display is observed, then the participant will also win something especially meaningful, like a scholarship or other accolade. In the second RNG study, a bored student investigator indifferently recruits an apathetic college sophomore, who is asked to press a button and mentally influence 1,000 random bits generated in one second, with no feedback of the results, and with no consequences regardless of the outcome.

The physical context of these two studies may be identical, employing the same RNG and the same statistics to evaluate the resulting datasets, each of

which consists of 1,000 random bits. But it is clear that the psychological contexts differ radically. If we presumed that the only important factor in this type of experiment was the number of bits generated, then the two studies should provide about the same outcome. But if MMI effects are moderated by variables such as the amount of time or effort one can apply in focusing mental intention towards the random events, or one's skill, or motivation, or how the bits were generated, then the first study might well result in an effect size orders of magnitude larger than the second.

The point is that one's view of what is *meant* by MMI shapes the definition of effect size, and as such it is important to underscore that the hypothesis under consideration is not merely a proposal about pure physics. Rather, the two M's in MMI propose an interaction in which the psychological and the physical are related in a meaningful way. A simple physics-centric assumption may be appropriate for an experiment designed to measure, say, the gravitational constant, but it is inappropriate for an experiment involving MMI.

Incidentally, the effect size of interest in most behavioral, social, and medical meta-analyses is based on performance at the subject level, and not on individual trials, samples, or runs within subjects, or accumulated across subjects. The reason that individual bits have become the effect size of interest in MMI experiments is undoubtedly because these are the most readily available and commonly reported data, and not because they are the most appropriate data. Indeed, as Schub states, "since the original hypothesis behind the experiments is that subjects can enhance hit rates, methods not based on hit rate are not tests of the original hypothesis." We agree, but then what hit rate is most appropriate? Schub means hit rate calculated based on individual bits. We suggest that better measures may be hit rates based on subjects' performance, or derived from the statistical significance of an entire experiment.

Alternatives to Influence Models

To the neophyte, an influence-per-bit model may appear to be the only way to imagine the "mechanism" of MMI. But there are other interpretations. The continuing debate over influence vs. precognition models, as formalized for example in Decision Augmentation Theory (Dobyns & Nelson, 1998; May et al., 1995), shows that there is much room for creative thought in modeling how RNG experiments can produce genuine non-chance results without invoking an influence-per-bit explanation.

Another alternative model is that the outcomes of MMI studies reflect a goal-oriented or teleological process associated with the psychological characteristics of the overall task (Schmidt, 1987). Under this proposal, RNG outcomes are best measured by the statistical results of entire sessions, or by units associated with feedback per effort. A third alternative is that the key variable in these studies is the amount of *time* the participant spends in mental effort (Nelson, in press). These alternatives are testable, and indeed when Bösch et al. (in press) tested our

suggestion that one might find z scores to be independent of sample size (Radin & Nelson, 2003), they confirmed that this was indeed the case in their RNG meta-analysis (after excluding three large "outlier" studies, which they argued was the preferred way of analyzing the data).

Schub questioned our "bold" proposal of a constant-z score, writing that "even if the [MMI] effect were real, the correlation [between z score and sample size] would be very small, and possibly non-significant, if the hit rate were also very small." Assuming that by "hit rate" Schub meant effect size per bit, then his comment is incorrect. The number of bits used in individual MMI experiments has ranged from about 10^2 to 10^8 bits. Assuming a stationary bitwise effect size (e) of any non-zero magnitude, then the z scores resulting from the largest studies would be 1,000 times larger than those from the smallest studies due to the relationship $z = e\sqrt{N}$. This would guarantee a very significant correlation. There is some evidence in support of this relationship within the PEAR RNG data (Radin, in press), but it is not supported by analysis of the remainder of the literature (May et al., 1995), possibly because the dozens of design variations in those latter studies obscured systematic relationships between z and N.

Different Ways of Presenting the Data

Schub writes, "Statistical significance in these meta-analyses can clearly be arrived at by a variety of methods which yield a wide variety of results. The results vary so widely because of some serious problems with the data ...." Actually, the results vary not because of "serious problems," but for a more mundane reason: Different analytical assumptions naturally lead to dissimilar outcomes. It is true that this class of experiments has been examined in numerous ways over the years, and for much the same reason that Schub did. By exploring various ways of combining the data, one can test different assumptions about the postulated effects. Of course, as long as the different models produce outcomes ranging from very to astronomically significant, conclusions about the *existence* of MMI effects are not in question.

Data from individual experiments can also be partitioned in different ways, e.g. as Schub notes the entire PEAR Lab RNG database can be collapsed into a single datapoint or it can be regarded as hundreds of separate experiments. In addition, the ways results are discussed in various papers are often adjusted to match the intended audience. In articles intended for physicists, we have emphasized measurement concepts such as effect sizes and standard errors, and we assumed a certain degree of statistical sophistication (Radin & Nelson, 1989). For example, referring to our 1989 meta-analysis in *Foundations of Physics*, Schub writes "Oddly, it does not quote a statistical significance of its primary result, though a z-score of about 6.8 can be read from one of the graphs." Actually, in that paper we report that in $N = 597$ experiments the average $z = 0.645$. We anticipated that most readers of *Foundations of Physics*

would be sufficiently adept at elementary statistics to understand that this meant the Stouffer z score was $sz = \bar{z}\sqrt{N} = 15.76$. We also assumed that those readers knew that an effect associated with 6.8 (weighted) or 15.76 (unweighted) standard errors beyond chance is so unlikely that citing the associated p-values is unnecessary.

For other audiences, such as psychologists and medical researchers, we have highlighted probabilistic concepts familiar in those disciplines, such as z scores and p-values (Radin & Nelson, 2003). For the general public we have been obliged to simplify the presentation and use probabilistic concepts associated with lotteries and casino games, such as hit rates and odds ratios (Radin, 1997, in press).

Publication Bias

Based on an asymmetric funnel plot, Schub concludes the presence of publication bias. The asymmetry in the plot is driven by heterogeneity in the bitwise effect size, with smaller studies (fewer total bits) showing much larger per-bit hit rates than larger studies. Under a uniform influence-per-bit model, this heterogeneity could indeed suggest a publication bias. However, the above discussions about the modulating effects of physical and psychological parameters provide a number of reasons for *expecting* heterogeneity, even in the complete absence of publication bias.

To illustrate why, consider Figure 1. This illustrates the failure of the funnel plot technique when applied to simulated data with no selective reporting bias (by design) and with data that are genuinely non-homogeneous. These data were synthesized according to a model comparable to a constant-z hypothesis, and then plotted according to their associated bitwise measurement uncertainty, as Schub would have done. The asymmetry in the upper graph is clear; the lower graph shows the effect of applying the standard "trim and fill" algorithm (Duval & Tweedie, 2000) to estimate the presence and impact of publication bias. Figure 1 demonstrates that when the underlying data do not fit the assumptions used in the standard funnel plot, then the analyst will unavoidably arrive at an incorrect conclusion. With intrinsically non-homogeneous data, typified by those observed in RNG studies, the standard funnel plot approach virtually guarantees an incorrect conclusion of publication bias.

In justifying his conclusions based on the funnel plot, Schub suggests that publication bias is a viable explanation because "In some fields, the marginal effort and expense of conducting a study is so high that very few studies ever go unreported. The RNG studies are probably not in that category, since the time it takes to conduct another study can be fairly modest." The problem with this explanation is that it is again predicated on the assumption that the only meaningful factor in these experiments is the total number of bits generated per study. If that assumption is even partially incorrect, then the publication bias argument falls apart.

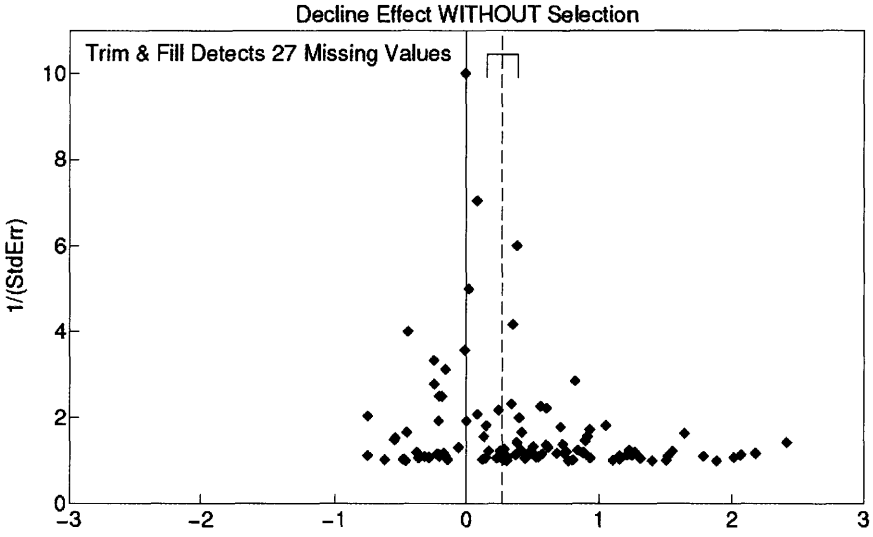

Fig. 1. Funnel plot (above) for simulated data generated according to a model comparable to a constant-z score hypothesis, without publication bias, and (below) after application of the standard trim-and-fill technique.

In a related criticism, referring to Scargle's (2000) method for estimating the number of studies required to nullify a meta-analytic result, Schub criticizes the Radin and Nelson (1989) paper because "Using Rosenthal's [filedrawer estimate] method, R&N estimate a 'failsafe' number of 54,000 unpublished

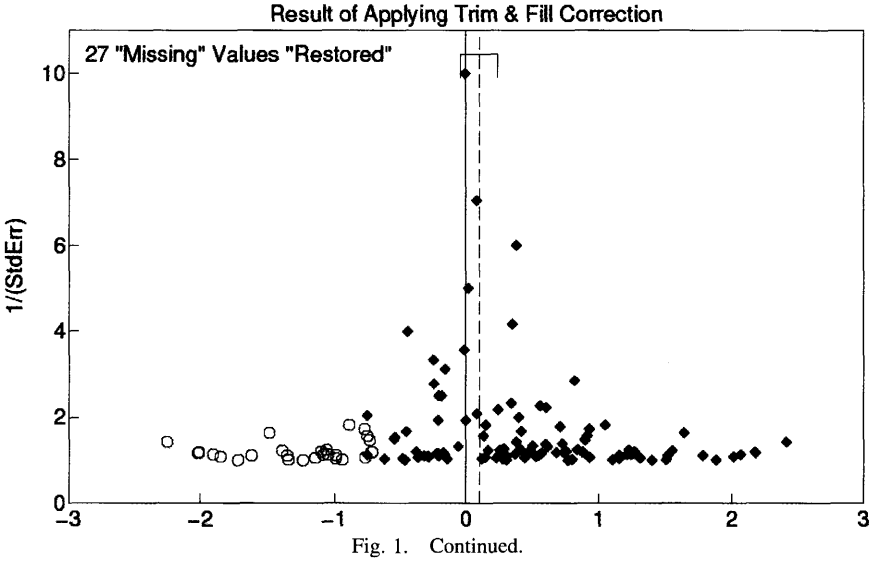

Fig. 1. Continued.

studies, but had they used [Scargle's] z = –0.1085 instead of z = 0 for unpublished studies, the estimate would have been only 2,681." This is a peculiar criticism because it suggests we should have employed a method which was developed eleven years later. Parapsychological studies have always attracted especially sharp criticisms, but this is the first instance we know of in which a critique was based on a failure to be precognitive.

In any case, imagine that we *had* used precognition and estimated that the filedrawer consisted of "only 2,681" non-significant, unpublished, irretrievable studies. Such a figure would have required every investigator who had ever reported an RNG experiment to also conduct but not report an additional thirty studies. Is that plausible? Not counting the PEAR Lab, which has no file-drawer problem due to a long-established laboratory policy, the most prolific investigator has been Helmut Schmidt. On average, Schmidt published about three studies per year for thirty years, many times the average rate of the other researchers. To achieve the estimated filedrawer of 2,681 studies, each of the other ninety investigators would have had to match Schmidt's prodigious output for a full ten years, and for not a single one of those studies to be reported or retrievable, even in surveys among colleagues who have known each other for decades. This is a highly implausible scenario.

Aside from the non-zero mean of the meta-analytic data, Schub also notes the increased variance of the distribution of results. This makes the claim of publication bias even more problematic, because the unexpected number of studies in the tails requires that a substantial amount of data must be *missing* from the center of the distribution. Of course, the peculiar void in the range $-0.4 < z < 0$, as noted by Schub, would be in accord with this thesis. But how realistic is this supposed missing data? Consider a center-suppressing biased publication process in which all studies with $z > 1.645$ are published, a fraction A of studies with $z < -1.96$ are published, and a fraction B of studies with $-1.96 < z < 1.645$ are published. This two-parameter model can be fit to the observed average $z = 0.645$ and variance $\sigma^2 = 2.903$ to yield $A = 0.635$ and $B = 0.037$. This leads to an overall publication rate of 0.1, from which it follows that a filedrawer of $597 \times 9 = 5{,}373$ studies is required to fit both the mean *and the variance* of the published results. The high frequency of studies in the tails, which Schub dismisses as unimportant, is actually an uncomfortable fit even based on this two-parameter model. Several published MMI-RNG studies have reported results with $z > 6$, and the probability of even one such study appearing in a total population of 5,970 experiments is itself very unlikely, with $p = 5.9 \times 10^{-6}$.

Schub counters with "direct evidence for publication bias in at least one original paper (Schmidt, 1976), where it is explicitly stated that complete records were not kept for some studies which did not produce positive results ...." Regarding the total of four studies cited, what Schmidt actually reports are experiments he describes "only as suggestive pilot studies." Schmidt was careful to maintain the distinction between "suggestive pilot" or screening

studies designed to locate talented participants vs. his pre-planned confirmatory tests. Schub is correct, however, that Schmidt's screening studies in that paper were included in our meta-analysis. So it is instructive to see what effect they had on the combined results. A Stouffer z of all fifteen experiments reported in Schmidt's paper was $z = 7.96$. When the four screening studies were removed, the composite z (non-significantly) declined to $z = 6.19$, still astronomically above chance.

More to the point are the results of the Bösch et al. (in press) RNG meta-analysis, as they partitioned pre-planned (N = 253 studies), post-hoc (N = 77), and not clearly specified (N = 50) experiments. For the pre-planned studies they found, based on a random effects model, evidence significantly in favor of the MMI hypothesis (after excluding three very large sample size studies, as they recommended).

To avoid future problems associated with selective reporting, Schub recommends a policy in which planned MMI experiments should be registered. This commendable policy has been implemented for decades by the *European Journal of Parapsychology*. In addition, an editorial policy encouraging authors to report all studies, positive or negative, was adopted by the *Journal of Parapsychology* in the mid-1970s, precisely to avoid problems associated with selective reporting.

Filedrawer estimates and models for publication bias are useful; however, to more directly investigate whether selective reporting could plausibly account for the observed results, in late 2005 we conducted a survey among many of the researchers who have conducted RNG studies. The survey revealed that the average number of unreported, non-retrievable experiments per investigator was about one, and some of those missing studies were reportedly statistically significant. Based on our survey, we believe that thousands of supposed filedrawer studies simply do not exist, and thus that selective reporting is an improbable explanation for the observed results.

Quality Assessment

Another reason that Schub was unimpressed by the RNG meta-analyses was due to a misinterpretation of the quality scoring method used in the Radin and Nelson (1989) analysis. Based on the raw numbers used to create those scores, he concluded that "The average quality score in this data ... is quite low—only 4.4 out of a possible 16." And that "Many quality features are surprisingly rare. Automatic recording ... is known present in only 57% of the non-PEAR studies."

For the 1989 meta-analysis we formed a conservative measure of methodological quality based on a combination of all known factors thought to be relevant to quality in this experimental domain. That analysis provided a range of *relative* scores useful for studying relationships among reported methods, effect sizes, and statistical significance. The scores were not intended to provide an absolute measure of experimental quality.

For example, automatic recording in RNG studies is so obvious a design factor that undoubtedly some investigators thought it unnecessary to report this feature. Nevertheless, for conservative reasons our scoring method relied exclusively on design features that were actually reported.

To gain an assessment of actual experimental quality, the best approach is to seek expert opinion based on a broad knowledge of the literature, both pro and con, and hands-on laboratory experience. Our opinion is that the overall quality of these studies is quite high, but as authors of positive RNG meta-analyses, we would not expect skeptics to uncritically accept our judgment. Ideally, we could refer to the opinion of meta-analysts who are not persuaded that RNG studies support the MMI hypothesis, and who have derived an independent assessment about quality based on their reading and coding of all of the relevant reports. Fortunately, such an opinion exists in the analysis of Bösch et al. (in press), who concluded that "The safeguard (quality) analyses indicated that the average study quality is very high." In addition, they found that the subset of highest quality studies (42% of all studies) resulted in a significant effect based on a random effects model (after excluding three outlier studies, as they recommend).

In referring to a quality vs. effect size correlation reported in *The Conscious Universe* (Radin, 1997), Schub writes, "A simple correlation coefficient is calculated between hit rate and quality. Radin concludes that hit rates are not related to quality, but the correlation coefficient he quotes is 0.165 ..., which is significantly positive ...." This is incorrect, and probably due to Schub misreading a footnote. The cited correlation refers not to quality vs. effect size, but rather to a significant improvement in quality vs. year of publication. This correlation was mentioned in the book to demonstrate that experiments have considerably improved over the years in response to criticisms and advancements in experimental protocols. The correct correlation for hit rate vs. quality is, as noted in the book, $r = 0.01$ (Radin, 1997: 312), which is not significant.

Minor Points

Statistical significance. Schub claims in his Table 1 that after applying a random effects model to the data that the null hypothesis begins to look more attractive. What do other analysts find when using the same model? Bösch et al. (in press) found that their random effects model resulted in an overall $z = 4.08$, after excluding three "outlier" studies. They cite this as evidence for a "statistically highly significant effect in the intended direction." It is worth noting that Bösch et al. (in press) *do not* interpret this as evidence for an MMI effect, but rather reflective of a selective reporting bias. We disagree with that interpretation for reasons already discussed, but the fact remains that an independent group using the same analytical model finds a highly significant

outcome where Schub does not. The difference may be due to Bösch et al.'s (in press) updated literature and their focus on a subset of MMI studies involving explicit human intention on truly random RNGs. In any case, one may argue that the "true" z score for these data, depending on which model one wishes to apply, probably resides in the range of $z = 4$ through 16, where the lower portion of this range involves weighting z scores by sample size.

Applicability of random effects modeling. Another problem with Schub's argument based on the non-significance of his random effects model is its inapplicability to a phenomenological existence test. That is, a non-significant random effects model does not imply that MMI does not exist. As Schub notes in his Appendix, "First a study effect size is chosen from a distribution with the between-studies variance, and then the observed hit rate is chosen from a distribution with the within-study variance." What he does not discuss is the fact that the standard random effects analysis assumes that any excess between-studies variance comes from random confounds that are irrelevant to the effect being measured, and statistical significance is calculated solely for the residual constant effect after the excess between-studies variance has been discarded. The problem with this approach for MMI-RNG experiments is that *any non-zero effect size* in any study is a demonstration of the existence of MMI. While spurious systematic effects may arise from methodological flaws, excess random between-studies variance in the random effects model *strengthens* rather than weakens the case for existence of the effect, a fact which does not appear in the standard significance report for the random effects main effect alone.

Optional stopping. Schub claims as an additional problem that "... in many of the original papers, the number of trials is not fixed ahead of time, so experiments might be run until a fixed statistical significance is obtained." In fact, it can readily be demonstrated by application of the well-known martingale theorem of probability theory that such optional stopping by individual experimenters cannot induce a spurious effect from the vantage point of a meta-analyst reviewing the literature. The only stopping point of importance is the final stopping point of the meta-analysis itself. Since this is established chronologically by the decision when to undertake the meta-analysis, it is presumably independent of the outcome (barring, perhaps, precognition by the analysts), and therefore it cannot induce a bias. While this is relevant to the publication bias problem since the martingale theorem can be defeated if some of the data are hidden, we may note that experiments which seek to attain significance by optional stopping—but have not yet done so—will also be among the largest and longest databases (since first-crossing probability for any statistical threshold declines with time). However, the largest and longest running databases are also those *least* likely to be either voluntarily hidden or irretrievable by meta-analysts. Thus, optional stopping is an implausible source of statistical inflation in the meta-analyses.

Repeatability. Is the MMI effect repeatable? Schub opines that "The subsequent failure by PEAR and others to replicate the original PEAR

findings ... argues that the meta-analysis is not valid." This statement is an overly simplistic view of replication. A few new experiments do not eradicate the results of a twelve-year experimental series from one laboratory, or a forty-year larger body of independently replicated data. But, Schub argues, "Half the authors have never published a significant effect." It is true that the majority of MMI studies do not reach statistical significance (two-tailed). But when all those "failed" studies are combined, their cumulative result is a highly significant Stouffer $z > 5$. This illustrates why significance counting is no longer used as a measure of replication—it fails to take into account the arbitrariness of what "statistically significant" means.

Nonsense hit rates. Schub mentions that "The hit rate of 0.509 quoted for the 1989 data in both *The Conscious Universe* and the 2003 paper is much larger than any hit rate shown in Table 1 ... it turns out that mostly it is because the calculation that gives 0.509 includes a number of nonsense hit rates." This is true, and we are grateful to Schub for spotting this mistake. There were twenty-four experiments coded with $N = 1$ because the actual sample size in those studies was unknown. Of those twenty-four studies, twenty-two were reported as "not significant," so for them we assigned the most conservative value of $z = 0$, equivalent to the null hypothesis hit rate of 0.5. When re-averaging all study hit rates after excluding two studies where $z > 0$ and $N = 1$, the average hit rate dropped slightly from the cited 0.509 to 0.507.

Reduced variance in control studies. Schub was concerned that "Even the control studies do not behave as expected, and show evidence of what are probably reporting biases .... The PEAR control data [too] is dramatically deficient in the tails." This is true, but is it really due to reporting bias? In the case of the PEAR Lab data, there is no reporting bias by design, precisely to avoid such arguments. The reduced variance of the remainder of the literature-reported control studies suggests that the PEAR variance effect replicates.

Ad hoc models. Schub faults our use of Rosenthal's "ad hoc" suggestion that a ratio of 5 to 1 failsafe to known studies might be considered a robust effect, but he doesn't find objectionable the equally ad hoc nature of Matthews' (1999) recommendation that people who are highly skeptical about a hypothesis should hold out for $z > 4.75$. The fact is that the literature does report individual high-quality RNG experiments that have surpassed Matthews' (1999) threshold (such as those of PEAR and Schmidt), and yet those studies have failed to sway the opinions of those who prefer to believe that such results must be due to flaws, fraud, or selective reporting.

Changing z scores. In comparing data used for our 2003 meta-analysis paper to those used in *The Conscious Universe*, Schub "found 15 z's had changed from negative to positive, and 15 from positive to negative." After analyzing the sign changes, it appears that the cause was probably confusion over how the intentional goal in each experiment was coded. We originally coded each study's aim as high or low, and then entered the associated z scores according to whether the results were in alignment with this goal. For example, if a study resulted in

a below chance outcome, and the intentional goal was low, then the associated z score would be recorded as positive. This double-negative bookkeeping was confusing, and it apparently resulted in shifting the signs of some of the figures when updating the database. We are grateful to Schub for bringing this to our attention, and it does raise a question about the impact of these sign shifts on the overall results. A conservative way to make this assessment is to evaluate the worst case scenario, in which all of the changed z scores are shifted towards the negative. Fortunately, the majority of studies (81%) were high aim, so even the worst case scenario would not be expected to affect the overall results much. In fact this scenario changed the original Stouffer z (for the analysis reported in *The Conscious Universe*) from 12.7 to 10.5, a non-significant decline.

On the Need for Certainty

At the beginning of this paper, we asked why some scientists persist in exploring anomalies while others persist in criticizing them. We believe the answer involves a meta-issue—a basic temperamental difference that creates two classes of investigators. Most scientists would probably agree that demonstrating the existence of a genuine MMI effect would be of profound importance, and thus careful consideration of this topic is warranted. But different predilections invariably lead to different assessments of the same evidence. Scientists who worry about Type I errors insist on proof-positive before taking the evidence seriously, while those who are more concerned about Type II errors prefer to take a more affirmative stance to counteract the prejudices invariably faced by anomalies research.

Type I preference leads to Bösch et al.'s (in press) comment that "this unique experimental approach will gain scientific recognition only when we know *with certainty* what an unbiased funnel plot ... looks like" (emphasis added). Or to Schub's statement that "perfect methodological quality is a logical necessity in parapsychology, since a paranormal effect can only be demonstrated if *all* possible non-paranormal causes are ruled out" (emphasis in the original). Demands for perfect, absolutely certain results are laudable in principle, but of course in practice such perfection does not exist. Of greater importance, when such statements are unpacked they are found to hide an irresolvable epistemological paradox.

Collins (1992) called this paradox the *Experimenters' Regress*, a Catch-22 that arises when the correct outcome of an experiment is unknown. To settle the existence question under normal circumstances, where results are predicted by a well-accepted theory, the outcome of an experiment is simply compared to the prediction. If they match, then we "know" that the experiment was conducted in a proper fashion and the outcome is taken as correct. If the outcome does not match, then the experiment was flawed. Unfortunately, when

it comes to a pre-theoretical hypothesis like MMI, to judge whether an experiment was performed "perfectly," or to know "with certainty" what an unbiased funnel plot looks like, we first need to know whether MMI exists. But to know that, we need to conduct the correct experiment. But to conduct the correct experiment, we need a well-accepted theory to inform us what "correct" means. But ... and so on.

For scientists with Type I temperament, this loop is destined to cycle indefinitely in spite of the application of the most rigorous scientific methods. The stalemate can be broken only by Type II scientists who are willing to entertain the possibility that Nature consists of many curious phenomena, some of which are not yet adequately accounted for by prevailing theories. In short, Type I scientists adopt the working hypothesis that claimed anomalies do not exist, and they seek evidence supporting that hypothesis; Type II scientists adopt the opposite approach. Given the biasing effects of one's adopted beliefs, it is not surprising that scientists tend to perceive their predilections. Of course, the goal in science is to get beyond personal beliefs and narrow in on the objective truth as best as possible. This is why critiques are so valuable, but also, when anomalies are in fact genuine, why debates tend to persist until both parties drop from exhaustion and new generations rise to the task with fresh energy and fewer prejudices.

Conclusion

From the earliest investigations of MMI with RNGs, researchers struggled to understand the physically perplexing but psychologically sensible goal-oriented nature of these phenomena (Schmidt, 1987). After a decade of replicated proof-oriented experiments suggested that something interesting was going on, most investigators began to concentrate on process-oriented research in an attempt to understand these complex psychophysical interactions. We sympathize with critics who assume that MMI implies a stationary, uniform, force-like effect on individual random bits, because that is also what many earlier researchers assumed. Unfortunately, that simple idea is not what Nature is revealing in these experiments.

Independent reviews of the relevant literature continue to indicate that the average methodological quality in these studies is high, and that on composite the results of hundreds of experiments are not due to chance. Some reviewers believe that selective reporting plausibly explains these results; we believe there are too few researchers and resources available to account for the filedrawer estimates. As a result, we find the cumulative data sufficiently persuasive to indicate that something interesting is going on. Like many of our colleagues, having reached that conclusion we find it more fruitful to focus on understanding the nature of these effects rather than spending time on a quest for the perfect experiment with absolutely certain results. That Holy Grail only exists in fantasy. This anomaly exists in the real world.

References

Bösch, H., Steinkamp, F. & Boller, E. (2006). Examining psychokinesis: The interaction of human intention with random number generators. A meta-analysis. *Psychological Bulletin, 132,* 497–523.

Collins, H. M. (1992). *Changing Order: Replication and Induction in Scientific Practice.* (2nd ed.). University of Chicago Press.

Dobyns, Y. H. & Nelson, R. D. (1998). Empirical evidence against decision augmentation theory. *Journal of Scientific Exploration, 12,* 231–257.

Duval, S. & Tweedie, R. (2000). A nonparametric "Trim and Fill" method of accounting for publication bias in meta-analysis. *Journal of the American Statistical Association, 95,* 89–98.

Ehm, W. (2005). Meta-analysis of mind-matter experiments: A statistical modeling perspective. *Mind and Matter, 3,* 84–132.

Matthews, R. A. J. (1999). Significance levels for the assessment of anomalous phenomena. *Journal of Scientific Exploration, 13,* 1–7.

May, E. C., Utts, J. M. & Spottiswoode, S. J. P. (1995). Decision Augmentation Theory: Applications to the random number generator database. *Journal of Scientific Exploration, 9,* 453–488.

Nelson, R. D. (2006). Time normalized yield: A natural unit for effect size in anomalies experiments. *Journal of Scientific Exploration, 20,* 177–200.

Radin, D. I. (1997). *The Conscious Universe.* HarperCollins.

Radin, D. I. (in press). *Entangled Minds.* Simon & Schuster.

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

Radin, D. I. & Nelson, R. D. (2003). Research on mind-matter interactions (MMI): Individual intention. In Jonas, W. B., & Crawford, C. C. (Eds.), *Healing, Intention and Energy Medicine: Research and Clinical Implications* (pp. 39–48). London: Harcourt Health Services.

Scargle, J. D. (2000). Publication bias: The "file-drawer" problem in scientific inference. *Journal of Scientific Exploration, 14,* 91–106.

Schmidt, H. (1976). PK effect on pre-recorded targets. *Journal of the American Society for Psychical Research, 70,* 267–291.

Schmidt, H. (1987). The strange properties of psychokinesis. *Journal of Scientific Exploration, 1,* 103–118.

Schub, M. H. (2006). A critique of the parapsychological random number generator meta-analyses of Radin and Nelson. *Journal of Scientific Exploration, 20,* 402–419.