RESEARCH

A Faulty PK Meta-Analysis

Wilfried Kugel
Berlin

Submitted August 26, 2009; Accepted November 3, 2010

Abstract—This article starts with an introduction to the concepts and experimental methodology used in the investigation of micro-psychokinesis (micro-PK). After a summary of three PK meta-analyses that seem to show a genuine PK effect, I will comment on a paper by Holger Bösch, Fiona Steinkamp, and Emil Boller (BSB), entitled "Examining Psychokinesis: The Interaction of Human Intention With Random Number Generators—A Meta-Analysis" (BSB-MA). The paper was published in the July 2006 issue of the *Psychological Bulletin* and suggests that all evidence of micro-PK may be due to publication bias. I will then show that the BSB-MA contains a large number of serious errors, which include data selection bias, faulty data coding, a lack of correspondence between experimental and control datasets, faulty statistical analyses, and erroneous interpretation of results. In addition, the entire negative z-score in the meta-analysis results from only one study. This meta-analysis, therefore, produced spurious results.

Terms and Methods

In accordance with the definitions published by the Rhine Research Center (Durham, NC, USA), the term *parapsychology* describes the scientific study of certain paranormal or ostensibly paranormal phenomena, in particular ESP and PK (or Psi, as a general term used either as a noun or adjective to identify ESP or PK). Extrasensory perception (ESP) denotes paranormal cognition; the acquisition of information about an external event, object, or influence (mental or physical; past, present, or future) in some way other than through any of the known sensory channels. ESP includes telepathy and clairvoyance. Precognition denotes a form of ESP involving awareness of some future event that cannot be deduced from normally known data in the present. Psychokinesis (PK) denotes paranormal action: the influence of mind on a physical system that cannot be entirely accounted for by the mediation of any known physical energy.

In a test of ESP, a target is defined as the object or event that the percipient attempts to identify through information paranormally acquired. In a test of PK, a target is defined as the physical system, or its effect, that the subject attempts to influence.

There are so-called macro-Psi and micro-Psi experiments. Typically, "macro-experiments" are closely related to spontaneous experiences and are performed under more or less informal conditions, whereas micro-experiments are performed under strictly controlled laboratory conditions. This paper will only look at micro-experiments.

Target sequences are typically generated by a random number generator (RNG), also called a random event generator (REG), an apparatus (typically electronic) incorporating an element capable of generating a random sequence of outputs. In tests of PK, the RNG may itself be the target system that the subject attempts to influence. Typical physical processes underlying a true RNG are electronic "white noise" and radioactive decay. "White noise" diodes are very susceptible to electromagnetic disturbances. RNGs based on radioactive decay are the best choice because this process and its statistical character cannot be influenced by any known human means (including electromagnetic disturbances).

A pseudo random number generator (PRNG) is an algorithm for generating a sequence of numbers that approximate the properties of a true random number series. The sequence is not truly random in that it is completely determined by an initial value ("entry point" or "seed"). In practice, the output of many common PRNGs (including the random functions of PCs and Macintoshes) exhibit artifacts. A common way to generate pseudo random numbers is to generate a "seed value" for an entry in the decimal digits of an irrational number such as π or e. (The decimal representation of an irrational number never ends or repeats.) PRNG-generated number sequences are not at all proper targets for either PK or precognition experiments.

Let us look at how a typical PK experiment is designed: In a fixed group of successive trials, called a run, the subject attempts to influence the outcome of an RNG. At least one independent control run (better if there are more) must be performed with the same RNG but without any attempt by the subject to exert intentional influence on the outcome. In a successful experiment, the control runs show no significant deviation from the theoretically expected random series, where the run intentionally influenced by the subject shows such a significant deviation from chance expectation, as measured by statistical methods such as the Binomial Test, the Chi^2-Test, and others.

In 1976, Helmut Schmidt introduced the hypothesis of a "PK Effect on Pre-Recorded Targets" (Schmidt, 1976), i.e. whether direct mental influences might occur in a time-displaced or "backward-acting" manner.[1]

The "Decision Augmentation Theory" (DAT) by May et al. (1985, 1995a, 1995b, 1995c)[2] reconceptualized PK as a precognition-based selection process rather than one of actual influence. This means that the subject of a PK experiment or even the experimenter may "foresee" when and where a natural statistical fluctuation in the generation of target bits appears and choose the right moment and place to start the recording of the target sequence. This basic problem—that the result of a PK experiment may depend only on when and where the target bits are generated—cannot be avoided in any experimental design.

The design of a precognition experiment differs considerably from the design of a PK experiment: The subject's role in a precognition experiment is not to influence, but to foresee the outcome of an RNG. To exclude artifacts produced by malfunctions of the RNG, the random series produced by the RNG must always fit the a priori theoretical expectation under all circumstances. In principle, no control runs are necessary as long as the RNG produced random numbers, as statistically expected. (This is true for all other ESP experiments as well.) Although it is conceivable that in precognition experiments a PK influence may transform one random series into another one, there is no way such an effect can be measured.

A meta-analysis (MA) sums up several individual studies and attempts to give an overall result. Because of the different paradigms, no ESP experiments may be included in a PK meta-analysis. ESP studies need target sequences that correspond to chance expectation, whereas PK studies measure the deviation of target sequences from chance expectation. Therefore, if ESP studies are merged with PK studies in a PK-MA, the overall result should be closer to random expectation if more ESP studies are included.

On the other hand, PK studies could be included in precognition meta-analyses, since it is possible that in PK experiments the subject may "foresee" upcoming deviations of the RNG output from chance expectation (as conceptualized by the DAT).

Previous PK Meta-Analyses

Prior to 2006, three meta-analyses (MAs) of PK experiments were made.[3] I shall refer to these meta-analyses as RN1, RN2, and RN3, respectively. (For a comparison of these analyses, see Table 1.)

RN1 Meta-Analysis

RN1 (Radin & Nelson, 1989) examined the following hypothesis: "The statistical output of an electronic RNG is correlated with observer intention with prespecified instructions, as indicated by the directional shift of distribution parameters (usually the mean) from expected values." The random event output of the RNGs, investigated in the RN1, originated in electronic noise, radioactive decay, or randomly seeded pseudo random sequences.

The RN1 examined all available 152 references from 1959 to 1987, covering 597 experimental PK studies, including 258 studies from the Princeton

TABLE 1
Comparison of the Results of 4 PK Meta-Analyses

META-ANALYSES Time Period	RN1 (1989) 1959–1987	RN2 (1997) 1959–1996	RN3 (2003) 1959–2000	BSB (2006) 1969–2004
Total studies	597	339	515	380
Non-PEAR studies included	339	339	number n.r. > 339	
PEAR studies included	258	0	258 (collapsed to 1) + number n.r.	32 (partly collapsed to 1)
Additional non-PEAR studies (update)	--	0	176 (including PEAR studies)	yes, number n.r.
Additional PEAR studies (update)	--	1,004	number n.r.	26 (partly collapsed to 1)
PEAR studies excluded	--	258	0	yes, number n.r.
PK studies excluded	--	0	0	yes, number n.r
ESP studies included	no (not verified)	no (not verified)	yes (supplied by BSB), number n.r.	yes, > 30
z	>+12 (with PEAR, according to RN3) +6.53 (with PEAR, according to BSB)	+6.41 (without PEAR)	+16.1 (with PEAR) +3.81 (according to BSB)	-3.67 (with PEAR) +3.59 (without fast PEAR "Mega")
PEAR studies separated	0	258	0	0
PEAR studies total	258	1,262	258 (combined to 1) + number n.r.	32
PEAR z	significantly positive, value n.r.	significantly positive, value n.r.	significantly positive, value n.r.	negative, value n.r.

n.r., not reported

Engineering Anomalies Research Laboratory (PEAR). A highly significant effect (z = 6.53) was found.

Bösch, Steinkamp, and Boller (BSB) criticized RN1 as follows:

> The authors did not [...] specify definite and conclusive inclusion and exclusion criteria. [...] Participants in the included studies varied from humans to cockroaches [...] The meta-analysis included not only studies using true RNGs, which are RNGs based on true random sources such as electronic noise or radioactive decay, but also those using pseudo-RNGs [...], which are based on deterministic algorithms. (Bösch et al., 2006:501)

RN2 Meta-Analysis

In RN2, (Radin, 1997), Dean Radin revisited the RN1 and calculated an experimental effect of ≈51% ($p < 10^{-12}$). For a replication analysis, Radin separately examined all PEAR experiments and updated the database to 1996, which included a total of 1,262 PEAR studies. He stated:

> Princeton University mathematician York Dobyns found that the seven years of new PEAR RNG results closely replicated the preceding three decades of RNG studies reviewed in the meta-analysis [RN1]. That is, our 1989 prediction had been validated. [...] Roger Nelson [...] found that the main RNG effect for the full PEAR database of 1,262 independent experiments [...] was associated with odds against chance of four thousand to one (Nelson et al., 1991) [$p \approx 0.00025$]. (Radin, 1997:142f)

But BSB derogated the method used in RN2:

> Radin (1996) [sic] recalculated the effect size of the first RNG meta-analysis, claiming that the "overall experimental effect calculated per study, was about 51%" (p. 141). However, this newly calculated effect is two orders of magnitude larger than the effect of the first RNG meta-analysis (50.018%). The increase has two sources. First, Radin removed the 258 PEAR laboratory studies included in the first meta-analysis (without discussing why) and second, he presented simple mean values instead of weighted means as presented 10 years earlier. (Bösch et al., 2006:501)

RN3 Meta-Analysis

In 2003, an update (RN3) of the previous meta-analyses by Dean Radin and Roger Nelson was published. The paper states:

> A literature review found 64 new publications describing 176 RNG experiments that were not retrieved in the earlier meta-analysis [...]. 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 data point for the purposes of the present [...] 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. [...] 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 ($p \ll 10^{-50}$). (Radin et al., 2003)

But in contrast to $z = 16.1$, BSB reported a z-score of only 3.81.

Let us look at how Radin and Nelson found the "new publications" they mentioned in RN3 (as quoted above). In August 2000, Roger Nelson sent a request to the IGPP ("Institut für Grenzgebiete der Psychologie und Psychohygiene" in Freiburg, Germany) for additional studies to be included in an updated version (RN3) of the RN1 PK meta-analysis. In an e-mail to me on November 4, 2007, Nelson stated that the new studies (including some that were done but not found during the period of the earlier RN1) were aggregated by Bösch and Boller, and assessed mainly by Steinkamp. (This was not the database of the later BSB-MA.) Roger Nelson appended a description and a list of the additional data. To my surprise, I found the results of two of my 1979 telepathy studies (TELBIN VOR, TELBIN S-SS) in the list that Roger Nelson sent me. Not only did these ESP studies not belong there, but none of the values given are correct. Moreover, arbitrarily selected results of my 1980–1999 precognition studies (A through D; see: Kugel, 1992, 1999), including series A and B, in which physical roulette wheels were used, were also included.

On November 5, 2007, Roger Nelson informed me that all the items in the list he had sent me were courtesy of Bösch et al. and that he had not been aware that they included inappropriate studies in the RN3.

Criticizing the RN3-MA in 2006, BSB unequivocally quoted that "no inclusion and exclusion criteria were specified" (Bösch et al., 2006:501). But the fact is that BSB themselves, by adding arbitrarily studies to the RN3-MA database, included selected ESP data. This was not consistent with the inclusion and exclusion criteria later specified by BSB themselves.

The 2006 PK Meta-Analysis by Bösch, Steinkamp, and Boller

In the July 2006 issue of the *Psychological Bulletin*, Holger Bösch, Fiona Steinkamp, and Emil Boller published a paper entitled "Examining Psychokinesis: The Interaction of Human Intention With Random Number Generators—A Meta-Analysis" (Bösch et al., 2006) (BSB-MA).

The 2006 BSB-MA was part of a five-year consortium project on RNG experiments. The consortium was established in 1996, lasted through 2000, and was funded by the IGPP.

> The consortium comprised research groups from the PEAR laboratory (Princeton Engineering Anomalies Research Laboratory, Princeton University, School of Engineering/Applied Science, Princeton, New Jersey, USA, founded in 1979 and closed in 2007); the Justus Liebig University of Giessen, Giessen, Germany (GARP); and the Institut für Grenzgebiete der Psychologie und Psychohygiene (Institute for Border Areas of Psychology and Mental Hygiene) in Freiburg, Germany (FAMMI). (For the results, see Jahn et al., 2000)

BSB summarized the results of their MA as follows:

> The meta-analysis combined 380 studies that assessed whether RNG output correlated with human intention and found a significant but very small overall effect size. The study effect sizes were strongly and inversely related to sample size and were extremely heterogeneous. A Monte Carlo simulation revealed that the small effect size, the relation between sample size and effect size, and the extreme effect size heterogeneity found could in principle be a result of publication bias. (Bösch et al., 2006:497)

BSB described the inclusion and exclusion criteria of the studies they used for the MA as follows:

> After the comprehensive literature search was conducted, we excluded experiments that
>
> (a) involved, implicitly or explicitly, only an indirect intention toward the RNG. For example, telepathy experiments, in which a receiver attempts to gain impressions about the sender's viewing of a target that is randomly selected by a true RNG, were excluded (e.g., Tart, 1976). Here, the receiver's intention is presumably directed to gaining knowledge about what the sender is viewing rather than to influencing the RNG. We also excluded those that
>
> (b) used animals or plants as participants (e.g., Schmidt, 1970b);
>
> (c) assessed the possibility of a nonintentional or only ambiguously intentional effect, for instance, experiments evaluating whether hidden RNGs could be influenced when the participant's intention was directed to another task or another RNG (e.g., Varvoglis & McCarthy, 1986), or experiments with babies as participants (e.g., Bierman, 1985);
>
> (d) looked for an effect backward in time [retro-PK] or, similarly, in which participants observed the same bits a number of times (e.g., Morris, 1982; Schmidt, 1985), and
>
> (e) evaluated whether there was an effect of human intention on a pseudo-RNG (e.g., Radin, 1982).
>
> In addition, experiments were excluded if their outcome could not be transformed into the effect size that was prespecified for this meta-analysis. This excluded studies for which the data are not expected to be binomially distributed. As a result, for example, experiments that compared the rate of radioactive decay in the presence of attempted human influence with that of the same element in the absence of human intention (e.g., Beloff & Evans, 1961) were excluded. [...] From the 372 experimental reports retrieved, 255 were excluded after applying the inclusion and exclusion criteria.

Confirming this, in an email of November 7, 2007, Fiona Steinkamp informed me that studies using pseudo-RNGs [e] as well as retro-PK studies [d] were excluded from the BSB-MA, as well as studies that assessed random decay of a radioactive source only, since an output was needed that was a clear 1 or 0 for a study to be included.

Published Responses to the BSB-MA

Two responses to the BSB-MA were published in the same July 2006 issue of the *Psychological Bulletin*, following the original paper.

Wilson et al. summarized only the intention of the BSB paper:

> The authors argue that, for both methodological and philosophical reasons, it is nearly impossible to draw any conclusions from this body of research. [...] If we had to take a stand on the existence of an RNG psychokinesis effect on the basis of the evidence in Bösch et al., we would probably vote no. (Wilson & Shadish, 2006:524, 527)

Radin et al. stated:

> Bösch et al. postulated the heterogeneity is attributable to selective reporting and thus that psychokinesis is "not proven". [...] The authors maintain that selective reporting is an implausible explanation for the observed data and hence that these studies provide evidence for a genuine psychokinetic effect. [...] Bösch et al. excluded two thirds of the experimental reports they found.

Furthermore, Radin et al. mentioned errors in the statistical treatment of the MA by BSB (Radin et al., 2006:529, 531).

The *Psychological Bulletin* gave BSB the opportunity to reply to the two comments. The reply was also published in the July 2006 issue. BSB stated that their

> meta-analysis [...] demonstrated (a) a small but highly significant overall effect, (b) a small-study effect, and (c) extreme heterogeneity. [...] The authors reaffirm their view that publication bias is the most parsimonious model to account for all 3 findings. (Bösch et al., 2006a)

Timm stated at the November 2006 workshop of the "Wissenschaftliche Gesellschaft zur Förderung der Parapsychologie" (WGFP) that Bösch et al. based their work on unrealistic assumptions about the structure of parapsychological experiments. After correcting the statistical analysis, he arrives at a highly significant value for the existence of PK and states that the attempt by Bösch et al. to attribute the PK results to publication bias is untenable (Timm, 2006). At the same workshop, Ertel also pointed out that the BSB-MA contained statistical errors, and that effect of publication bias was negligible. He, too, attributed the results to a genuine, overall PK effect (Ertel, 2006).

Boller rejected the criticism of Timm and Ertel as spurious. It only addressed the general problems of PK research, he said, but not specifics of the BSB-MA (Boller, 2007).

Unfortunately, none of the authors who commented on the BSB-MA mentioned that the entire database of this PK-analysis was assembled incorrectly by including arbitrarily selected ESP data, presumably because they were not aware of this important fact.

My Criticism of the BSB Meta-Analysis

The Subject Matter

BSB conceded that they faced a difficulty:

> Deciding which experiments to include and which to exclude, even if the criteria are clearly defined, can be as delicate as are decisions concerning how to perform the literature search and decisions made during the coding procedure." (Bösch et al., 2006:503)

But were the authors capable of handling that difficulty? Defining the subject matter of their MA clearly seems to have posed a problem for them. Although references to the DAT appear in the BSB paper as sources (May et al., 1985, 1995a), there is no reference to it in the text. They did not take into consideration that PK experiments can be included in a precognition MA, but not vice versa.

PK Data Mixed with Arbitrarily Selected ESP Data

The title and the text of the paper states clearly that the subject matter of the MA is psychokinesis (PK). BSB state (Bösch et al., 2006:502): "The final database included only experimental reports that examined the correlation between direct human intention and the concurrent output of true RNGs." It follows that only genuine PK experiments should have been included in the BSB-MA.

The BSB paper fails to mention that ESP (Extrasensory Perception) studies, presumably mainly precognition studies, were also included in the MA.

Holger Bösch provided me with some original SPSS data files, in particular, "Experimental Data Description," "Experimental Data," "Control Data Description," and "Control Data."

According to a table in the file "Experimental Data Description," the "380 studies fulfilling our inclusion and exclusion criteria" covered 302 PK studies (79.5%), 40 precognition studies (10.5%), 4 mixed studies (1.1%), and 34 "other" studies (8.9%), whatever "other" may mean.

Confronted with my criticism regarding the inclusion of ESP studies in a PK-MA, Emil Boller wrote to me on September 27, 2007, that, to his knowledge, many authors in parapsychology agree that PK and precognition cannot be distinguished unambiguously. His co-authors, he stated, shared this opinion. Therefore, Boller argued, precognition experiments using genuine random number generators had to be included in the BSB-MA. Fiona Steinkamp made a similar argument when she wrote to me on November 7, 2007, that the BSB-MA, in keeping with the inclusion criteria, could include precognition studies as long as there was an intention to obtain a result in a specified direction and a true RNG was used. She added that the BSB-MA was not defined as a PK meta-analysis. But this is clearly not the case and contradicts even the title of the BSB paper, "Examining Psychokinesis: ...". BSB knew that

> PK [Psychokinesis] refers to the apparent ability of humans to affect objects solely by the power of the mind, and ESP relates to the apparent ability of humans to acquire information without the mediation of the recognized senses or inference. (Bösch et al., 2006:497)

But what do the authors mean by "human intention"? They write: "The participants' intention is generally directed (by the instructions given to them) . . . ".

And they explain:

> Telepathy experiments, in which a receiver attempts to gain impressions about the sender's viewing of a target that is randomly selected by a true RNG, were excluded [...]. Here, the receiver's intention is presumably directed to gaining knowledge about what the sender is viewing rather than to influencing the RNG. (Bösch et al., 2006:510, 502)

But the same is true for precognition experiments, where the receiver's intention is presumably directed to gaining knowledge about the RNG's outcome, but not directed by any instructions to influence the outcome of the RNG.

I was surprised that two of my experimental reports from 1979 and 1999 were marked with an asterisk as included in the BSB-MA (Bösch et al., 2006: 517, 520), which covered one telepathy study (Kugel et al., 1979) and four precognition studies (Kugel, 1999), i.e. ESP, but not PK. I will examine this in more detail below and will show that the BSB-MA is, in large part, highly questionable.

Here is a striking example of what I would call "selection bias" in the BSB-MA: In my experiments, I used the following random sources: In 1971, random number tables, freshly generated by a computer; in 1972, a high-frequency electronic ring-counter driven by "White Noise"; in 1973, a high-frequency electronic ring-counter driven by "White Noise" and additionally distorted by radioactive decay; and from 1979 on, RNGs driven exclusively by radioactive decay. All of my experiments are described in the research reports (No. 1 to No. 6) I submitted to the Psychological Institute and later (No. 7 and No. 8) to the Institute for Applied Informatics of the Technical University Berlin. All these reports were available for BSB in the archive of the IGPP. But only arbitrarily selected parts of report No. 7 were included in their MA, namely parts of TELBIN and MM, but the latter was not listed in the BSB references.

I will now turn to the results of my own precognition experiments, published in 1999 and inappropriately used by BSB. From the 4 studies (A through D), only two, studies C and D, were included in the BSB-MA, because those were the only two studies with electronic RNGs.[4] In my 1999 paper I explicitly stated "that in series D, there was no PK influence" (Kugel, 1999:142). BSB's file "Experimental Data" includes z-scores (−0.24; −0.36) only provided as side information. In my paper, I made it clear that "there was no hypothesis with respect to these scores." The z-scores under the (one-tailed) hypothesis of high scoring (+1.06 and +2.19) do not appear in the BSB-MA. This selection by BSB appears to be biased. The output of the RNGs was analyzed extensively before, during, and after the precognition experiments, and was found to be completely random. No control runs of series C and D were included in the BSB-MA, although a copy of my research report (Kugel, 2000) on these series has been available to the authors in the archive of the "Institut für Grenzgebiete der Psychologie und Psychohygiene e.V., Freiburg i.Br.” (IGPP) since the year 2000.

Inclusion and Exclusion Criteria

BSB had stated: “We have a list of 225 [correct number is 255] reports that did not meet our criteria, and it is available to anyone who asks” (Bösch et al., 2006a:536). Ignoring my request, BSB did not send me the list. This lack of information prevented me from looking at some of the issues in more detail.

Apparently, PK studies investigating the possible influence on radioactive decay were excluded because BSB were unable to handle the data statistically. Moreover, other studies were also excluded or only summarized, despite the stated criteria.

Radin, Nelson, Dobyns, and Houtkooper stated in 2006:

> Bösch et al. excluded two thirds of the experimental reports they found. [BSB found 372 reports on relevant experiments. But only 117 were used in their MA.] That selection may have introduced important factors that the reader cannot evaluate. In any case, the exclusion of data with a nonbinomial distribution, such as studies based on radioactive decay, is questionable. (Radin et al., 2006:531)

According to BSB‘s file “Experimental Data” only six data points of the pre-2000 PEAR studies were included, two of them covering several studies, collapsed to one data point each. These six data points cover PEAR studies until 1994 (according to Nelson, 1994) and include 483.69 million trials. But according to a summary published in 1997 by Jahn et al. (Jahn et al., 1997), the number of trials of PEAR experiments in a 12-year program (Jahn et al. did not report the exact time period) was about 499.44 million trials. This shows that from 1994 on, data of the pre-2000 PEAR studies were excluded from the BSB-MA. Furthermore, only some of the data from a 2004 PEAR study were included in the BSB-MA (Dobyns et al., 2004). Radin, Nelson, Dobyns, and Houtkooper criticized this in 2006:

> The reference in question reports two experiments, only one of which Bösch et al. considered. Of the two experiments, one is subdivided into three phases, each generating two data sets per phase, producing a total of seven data sets that can be distinguished as separate studies. (Radin et al., 2006:530)

This corresponds to my findings, as reported here, and clearly shows that BSB chose their database arbitrarily.

In their reply to Radin et al., BSB wrote:

> Meta-analytic results can be distorted [...] by the selection of publications to insert in the meta-analytic database. Even the most well-intentioned, comprehensive search strategy aimed at including published as well as unpublished

> manuscripts can be fallible. We do not deny that we inadvertently missed some relevant reports, despite having done our best to contact all researchers in the field and to search through all relevant journals and other publications. (Bösch et al., 2006a:536)

But BSB did not even try to contact me.

Inadequate "Control Data"

372 experimental reports were retrieved for the BSB-MA, but only 137 corresponding control studies. How did BSB define control studies?

> Many experimenters performed randomness checks of the RNG to ensure that the apparatus was functioning properly. These control runs were coded in a separate "control" database. [...] The purpose of control studies is to demonstrate that, "without intention," the apparatus produces results (binomially distributed) as expected theoretically.

Moreover, they wrote: "The control studies in this meta-analysis were simply used to demonstrate that the RNG output fits the theoretical premise (binominal distribution)" (Bösch et al., 2006:503,514). But despite their use of the term "corresponding control studies," they state, on the same page: "We have coded and analyzed unattended randomness checks as 'control' studies." This is an arbitrary decision, as I was able to show when I examined the SPSS file "Control Data": From my 1979 telepathy study (TELBIN), BSB arbitrarily took two of six pre-experimental hardware tests of the RNGs at 4,000 trials each. But why only two? Furthermore, the two hardware tests taken had nothing to do with the experiments, and the corresponding experiments were not included in the experimental database. BSB could just as well have used arbitrary series from random number tables. For the RNG tests, no target was set and the randomness check was done with the Chi^2-test. The Chi^2-value always has a positive sign. But from these Chi^2-values, BSB calculated two z-scores. With one of the values they were lucky, because the Chi^2-value was 0, giving $z = 0$. But the other z-value was given a negative sign! Everyone familiar with statistics knows this is strictly prohibited if the direction of the deviation is not known.

Faulty Data Coding

An equally striking example of questionable procedure is the data coding in the BSB-MA. In the case of my data, according to BSB's SPSS files "Experimental Data Description," "Experimental Data," "Control Data Description," and "Control Data," this was done by Emil Boller as "first coder No. 2". BSB included 20 items in their control database, taken from one of my studies ("MM"). MM is not listed as a reference by BSB, but published in the same research report as TELBIN in 1979. This exploratory study in a fivealternatives design included four telepathy experiments and two PK experiments of 250 trials each. Twenty random series of 250 trials each were generated as tests of the RNG. Only p-values of these 20 Chi^2-tests were reported by me. Notwithstanding the fact that the data of these 20 tests were not binomially distributed, BSB's files show 20 negative z-scores, which according to the file "Control Data" "had to be estimated from p values supplied". The corresponding (and statistically not significant) results of the two PK experiments of my study MM were reported only as two p-values, but combined into one completely fictional positive z-score by BSB. (A comment in the file "Experimental Data" says: "did stouffer z on both studies and then estimated hits—z was 0.38891"). These are serious errors in data coding.

Summary

For a comparison of all PK-MAs, including the BSB-MA, see Table 1.

The 2006 BSB-MA contained 302 PK studies, only 71.4% of the 423 studies, originally published in 1989 in the RN1. More than 40 non-PK (mainly ESP) studies were added to the database. It stands to reason, therefore, that the end result strongly differs from the previous meta-analyses.

As mentioned in the beginning of this paper, ESP studies need target sequences that correspond to chance expectation, whereas PK studies measure the deviation of target sequences from chance expectation. Therefore, if ESP studies are merged with PK studies in a PK-MA, which measures the deviation from chance, the overall result should more closely approximate random expectation when more ESP studies are added. In the BSB-MA, 40 of the 302 studies were ESP studies. This significantly reduces the overall result of the PK studies.

Let's look at the overall result of the BSB-MA. In their summary, BSB claim to have found two results, "a significant but very small overall effect size" ($z = -3.67$) and "that the small effect size, the relation between sample size and effect size, and the extreme effect size heterogeneity found could in principle be a result of publication bias" (Bösch et al., 2006:497). BSB's allegation of publication bias warrants no elaborate comments. Every scientist knows that it is difficult to publish nonsignificant results. This is true not only for parapsychological research, which addresses the problem candidly, as BSB well know (Bösch et al., 2006:515). Furthermore, from a statistical viewpoint, it is not a priori permissible to call a negative z-value "significant," as BSB did. To make a decision about the significance of a probable effect, BSB would have had to discuss all studies from the viewpoint of whether the studies were performed under a one-tailed or two-tailed hypothesis. Under a one-tailed hypothesis, used for most studies that were done, a negative z-value is completely meaningless. But all positive results of previous meta-analyses in

Table 2 of the BSB-MA are marked as one-tailed! To describe the overall result of the BSB-MA as “significant” is yet another error of its authors.

The overall result of the BSB-MA raises the question why, contrary to the three former meta-analyses, a total negative score of $z = -3.67$ was reported. The answer is obvious. The 2004 PEAR report by Dobyns et al., included in the BSB-MA, contains three studies with a total of more than 3×10^{11} trials, a number that is about 100 times higher than the approximately 10^9 trials of all other studies analyzed in the BSB-MA. These three studies showed negative deviations from chance expectation (“MegaREG fast REG” $z = -2.98$, “fast REG” $z = -2.43$, “Mega-Mega-REG, fast modus” $z = -2.08$). Dobyns et al. commented on these studies:

> In the initial phase of MegaREG, the 200-bit trials produced outcomes comparable with our standard experiments, while the 2-million-bit trials produced an effect somewhat larger in absolute scale, but inverted with regard to intention. [...] A related experiment called “MegaMega” [...] produced a reversed intentional effect of the same scale. (Dobyns et al., 2004:369)

BSB knew about that problem, as they wrote themselves: “Without these three studies, both models showed a statistically highly significant effect in the intended direction.” (Bösch et al., 2006:506) It is not known which changes had been made to the hardware and/or software of the PEAR RNG to achieve a 10^4 times higher RNG output rate. In all former PEAR experiments, the RNG output was 200 bits per trial. In the new “high speed experiments”, there were 2,000,000 bits per trial. No control data for these exceptional studies were included in the BSB-MA. That is regrettable, because Dobyns et al. stated in their 2004 paper (Dobyns et al., 2004:393) that “The noise source used has since suffered electronics failure.”

We can conclude, therefore, that the overall negative result of the BSB-MA ($z = -3.67$) could be due to a large amount of data from an RNG that may have been malfunctioning. Since BSB decided to use this material and to value it higher than all other studies together, they could just as well have decided to leave out all other studies. They would have come up with the same overall result. Excluding the 2004 Dobyns et al. study, the z-score of the BSB-MA would be +3.59.

Trying to find an explanation for their results, BSB claimed:

> However, another difference between the current and the previous meta-analyses lies in the application of inclusion and exclusion criteria. We focused exclusively on studies examining the alleged concurrent interaction between direct human intention and RNGs. All previous meta-analyses also included nonintentional and nonhuman studies. [...] This difference might explain the reduction in effect size and significance level. (Bösch et al., 2006:513f)

But this statement is not correct. BSB arbitrarily excluded genuine "intentional" PK studies and included "nonintentional" ESP studies (such as mine). This might explain a z-score of +3.59 (excluding the 2004 Dobyns et al. data), which is much lower than in the previous three PK meta-analyses.

This article will have served its purpose if it facilitates an assessment of the factors that contributed to the publication of faulty data and conclusions in an important field of parapsychological research.

Notes

1 The results of all retro PK experiments were excluded from the BSB-MA.

2 This problem was addressed earlier by the author (Kugel et al., 1978), including a new definition of Psi.

3 Despite the fact that I contacted the authors Dean Radin and Roger Nelson several times, I was not given access to the complete original databases of the meta-analyses they published in 1989, 1997, and 2003.

4 The studies A and B were performed with a toy roulette (A) and at real roulette tables at the Casino in the Berlin "Europa Center" (B). BSB write (Bösch et al., 2006:498): "Although there has been some variety in methods to address PK, such as coin tossing and influencing the outcome of a roulette wheel, these methods have been used only occasionally." This statement is not correct. I am not aware of one single experiment testing possible influences of intent on the outcome of a real roulette wheel. All experiments with real roulette wheels were precognition experiments.

References

Beloff, J., & Evans, L. (1961). A radioactivity test of psycho-kinesis. *Journal of the Society for Psychical Research, 41,* 41–46.

Bierman, D. (1985). A retro and direct PK test for babies with the manipulation of feedback: A first trial of independent replication using software exchange. *European Journal of Parapsychology, 5,* 373–390.

Boller, E. (2007). Erwiderung auf die kommentare von Suitbert Ertel und Ulrich Timm zur RNG-metaanalyse von Bösch, Steinkamp & Boller (2006). http://www.anomalistik.de/sdm_pdfs/erwiderung_boller_auf_timm_ertel.pdf

Bösch, H., Steinkamp, F., & Boller, E. (2006). Examining psychokinesis: The interaction of human intention with random number generators—A meta-analysis. *Psychological Bulletin, 132*(4), 497–523.

Bösch, H., Steinkamp, F., & Boller, E. (2006a). In the eye of the beholder: Reply to Wilson and Shadish (2006) and Radin, Nelson, Dobyns, and Hooutkooper (2006). *Psychological Bulletin, 132*(4), 533–537.

Dobyns, Y., Dunne, B., Jahn, R., & Nelson, R. (2004). The MegaREG Experiment: Replication and interpretation. *Journal of Scientific Exploration, 18*(3), 369–397.

Ertel, S. (2006). Kritischer kommentar zu einer meta-analyse von Bösch, Steinkamp & Boller. *Workshop der WGFP, Offenburg, 5*. November. (http://www.anomalistik.de/sdm_pdfs/ertel-boesch-kritik.pdf)

Jahn, R., Dunne, B., Nelson, R., Dobyns, Y., & Bradish, G. J. (1997). Correlations of random binary sequences with pre-stated operator intention: A review of a 12-year program. *Journal of Scientific Exploration, 11*(3), 345–367.

Jahn, R., Dunne, B., Bradish, G., Dobyns, Y., Lettieri, A., Nelson, R., Mischo, J., Boller, E., Bösch, H., Vaitl, D., Houtkooper, J., & Walter, B. (2000). Mind/Machine Interaction Consortium:

PortREG replication experiments. *Journal of Scientific Exploration, 14*(4), 499–555.

Kugel, W. (1992). Amplifying precognition: Two experiments with roulette. *European Journal of Parapsychology, 8*(1990–1991), Edinburgh 1992, 85–97.

Kugel, W. (1999). Amplifying precognition: Four experiments with roulette. *The Parapsychological Association 42nd Annual Convention: Proceedings of Presented Papers*; August 4–8, 1999, Stanford University, Palo Alto, California, pp. 136–146. Durham, NC: Parapsychological Association.

Kugel, W. (2000). Abschlußbericht. *Projektnummer 670110*, Juli 2000. Freie Universität Berlin, Fachbereich Physik, Gast-Forschungsprojekt "Prognostizierbarkeit stochastischer Ereignisse (Präkognition / Psychokinese)".

Kugel, W., Bauer, B., & Bock, W. (1979). Versuchsreihe TELBIN [Experimental series TELBIN]. *Arbeitsbericht 7* [*Research Report No. 7*], Forschungsprojekt-Schwerpunkt Biokommunikation, Fachbereich Informatik, Technische Universität Berlin (Germany).

Kugel, W., Bock, W., Bauer, B., & Pörner, B. (1978). *Materialsammlung Psychotronik/ Parapsychologie.* Berlin: Forschungsinstitut für Psychotronik.

May, E. C., Radin, D. I., Hubbard, G. S., Humphrey, B. S., & Utts, J. M. (1985). Psi experiments with random number generators: An informational model. *The Parapsychological Association 28th Annual Convention: Proceedings of Presented Papers*, pp. 235–266. Durham, NC: Parapsychological Association.

May, E. C., Utts, J., & Spottiswoode, J. P. (1995a). Decision augmentation theory: Toward a model of anomalous mental phenomena. *Journal of Parapsychology, 59,* 195–220.

May, E. C., Utts, J., & Spottiswoode, J. P. (1995b). Decision augmentation theory: Applications to the random number generator database. *Journal of Scientific Exploration, 9*(4), 453–488.

May, E. C., Utts, J., Spottiswoode, J. P., & James, C. L. (1995c). Applications of decision augmentation Theory. *Journal of Parapsychology, 59,* 221–250.

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

Nelson, R., Dobyns, Y., Dunne, B., Jahn, R. (1991). Analysis of variance of REG experiments: Operator intention, secondary parameters, database structure. *Technical Note PEAR 91004*. Princeton, NJ: Princeton Engineering Anomalies Research Laboratory.

Nelson, R. (1994). Effect size per hour: A natural unit for interpreting anomalies experiments. *Technical Note 94003.* Princeton, NJ: Princeton Engineering Anomalies Research Laboratory.

Radin, D. (1982). Experimental attempts to influence pseudorandom number sequences. *Journal of the American Society for Psychical Research, 76,* 359–374.

Radin, D. (1997). *The Conscious Universe.* San Francisco: Harper Edge, pp. 138–144.

Radin, D., & Nelson, R. (1989). Evidence for consciousness-related anomalies in random physical systems. *Foundations of Physics, 19*(12) (December 1989), 1499–1514.

Radin, D., & Nelson, R. (2003). Research on mind–matter interactions (MMI): Individual intention. In: W. B. Jonas & C. C. Crawford, (Eds.), *Healing, Intention and Energy Medicine: Research and Clinical Implications*, Edinburgh: Churchill Livingstone, pp. 39–48. Nearly identical Internet version *Meta-Analysis of Mind–Matter Interaction Experiments: 1959 to 2000* at http://www.boundarylab.org/bi/articles/rngma.pdf

Radin, D., Nelson, R., Dobyns, Y., & Houtkooper, J. (2006). Reexamining psychokinesis: Comment on Bösch, Steinkamp, and Boller. *Psychological Bulletin, 132*(4), 529–532.

Schmidt, H. (1970b). PK experiments with animals as subjects. *Journal of Parapsychology, 34,* 255–261.

Schmidt, H. (1976). PK effect on pre-recorded targets. *Journal for the American Society for Psychical Research, 70*(3), 267–291.

Schmidt, H. (1985). Addition effect for PK on pre-recorded targets. *Journal of Parapsychology, 49,* 229–244.

Tart, C. (1976). Effects of immediate feedback on ESP performance. In: J. D. Morris, W. G. Roll, & R. L. Morris (Eds.), *Research in Parapsychology 1975*, pp. 80–82. Metuchen, NJ: Scarecrow Press.

Timm, U. (2006). Kommentar zu Bösch, H., Steinkamp, F., & Boller, E.. *Workshop der WGFP, Offenburg, 4*. Novembre 2006. (http://www.anomalistik.de/sdm_pdfs/timm-boesch-kritik.pdf)

Varvoglis, M., & McCarthy, D. (1986). Conscious-purposive focus and PK: RNG activity in relation to awareness, task-orientation, and feedback. *Journal of the American Society for Psychical Research, 80,* 1–29.

Wilson, D., & Shadish, W. (2006). On blowing trumpets to the tulips: To prove or not to prove the null hypothesis—Comment on Bösch, Steinkamp, and Boller. *Psychological Bulletin, 132*(4), 524–528.