Micro-psychokinesis

Chapter · January 2015

1 author:

Peter Amalric Bancel
Institut Métapsychique International

MICRO-PSYCHOKINESIS

Mario Varvoglis and Peter Bancel

Institut Métapsychique International

Introduction

Case studies, fieldwork and laboratory research have associated psychokinesis (PK) with an extremely wide variety of phenomena: metal-bending, poltergeists, table levitations, materializations, induced physiological changes in living organisms, and statistical deviations from chance in probabilistic systems.

Thus while it is common to describe psychokinesis as '*the apparent ability of humans to affect objects solely by the power of the mind*'[1] this definition does not adequately cover *all* purported psychokinetic phenomena - PK does not only involve inanimate objects, and its sources may not be limited to human intentions. A more inclusive and neutral definition might be worded as follows: *PK is the putative ability of organisms to affect other systems - both animate and inanimate - without mediation of any known physical forces or energies.* Within this broad scope, a convenient classification distinguishes between three types of PK:

- *macro-PK* - generally including phenomena that can be directly perceived with the senses (such as the displacement or distortion of objects)
- *bio-PK* - or, as more commonly labeled, DMILS : the Direct Mental Influence on Living Systems, including aspects of human physiology and health, plant growth, etc.
- *micro-PK* - minute influences on inanimate, probabilistic systems, producing effects that can only be detected through statistical means. The target systems may include tumbling dice, coin tossing systems, or hardware random number generators (RNGs).

Though widely used in parapsychology, these heuristic distinctions are not mutually exclusive. For instance, DMILS could be subsumed under the micro-PK label, since such effects are generally too small to be directly observed, and must be inferred through repeated trials and statistical analyses. On the other hand, the basis for a distinction between micro-PK and macro-PK is itself questionable (Braude, 1986) and can be considered a useful, but superficial taxonomic distinction, based on the methods used to observe the phenomena (direct observation vs. statistical inference). In this case, we would be tempted to subsume macro-PK effects under the micro-PK explanatory umbrella, even if the methods used to observe the effects differ. Alternatively, it may be that there is no common basis to phenomena as diverse as a levitating table and the tiny statistical deviation of a RNG. In that case, micro-PK might not be about underlying physical processes at all, but indicative of other psi phenomena, such as precognition.

Notwithstanding these important interrogations, in this chapter we use the term micro-PK as a convenient shorthand for *statistically inferred evidence for the apparent influence of living systems on inanimate probabilistic systems, without mediation of any known physical forces or energies*. The qualifier apparent is intended to leave open the interpretational questions cited above, including the possibility that micro-PK effects are not direct influences upon physical systems, but merely correlations between guesses and outcomes.

In this chapter we deal with a subset of micro-PK research involving laboratory studies with a well-defined set of participants. Thus the studies we will be examining involve three key elements:

- a probabilistic *target system*, such as RNGs or tumbling coins or dice, with a finite set of possible outcomes and known theoretical probabilities, such as the equal chance of heads or tails in a coin toss);
- a *subject* (sometimes referred to as the experimental participant or operator) who, for a given number of trials, is asked to intend for or favor one particular outcome;
- an *experimenter*, who formulates hypotheses; defines experimental protocols; recruits the subjects; conducts the experimental sessions; and analyzes the data with statistical tools.

In many ways, then, the scope of this chapter is similar to that of Rex Stanford's (1977) excellent *Experimental Psychokinesis,* which reviews "*statistically evaluated studies of PK by humans upon nonliving systems*". Nevertheless, there are significant differences. Stanford's article constitutes a rather complete literature review which surveys three historical periods of research, each defined by a predominant target or device: dice tossing, placement studies of rolling dice and RNGs. By contrast, in this chapter we focus almost exclusively upon research with RNGs. This choice essentially reflects the historical fact that, soon after their introduction in parapsychology, RNGs became the focus of micro-PK research, while the earlier target systems were largely abandoned. Even within the RNG micro-PK literature, we are not seeking to review, even summarily, every study conducted since 1977. Instead, we survey the landscape of micro-PK research, highlighting certain studies that are representative of a particular approach[2].

A more substantive difference concerns our objectives. Stanford aimed to render the rather heterogeneous field of micro-PK research comprehensible within a larger conceptual context. He was not so much interested in the evidence for micro-PK (which he more or less took for granted) as tracking down its psychological and physical correlates in view of constructing a general model for psi functioning. Our primary objective is more proof-oriented: in our estimation, the jury is still out regarding whether the results confirm micro-PK as an effect which truly alters the behavior of physical systems. As we shall see, several researchers argue that RNG studies do *not* point to a influence on the target system, but rather to a psi-mediated selection process that tacitly guides the individual to sample the RNG in a way that produces favorable outcomes.

More importantly, the latest meta-analysis of RNG micro-PK research (Bosch, Steinkamp and Boller, 2003) questions whether there is any real effect to discuss at all. The authors of the meta-analysis assert that the small cumulative effect found across studies can be explained by publication bias. They conclude that, after nearly four decades of research micro-PK has not been convincingly demonstrated.

We take issue with both positions, but neither can be easily dismissed: from a meta-analytic perspective, the case for micro-PK is not as self-evident as it may have appeared almost 40 years ago. Thus, our objective is to examine the basic evidential issues, including the meta-analytic perspective, while attempting to elucidate its nature and highlight potentially relevant factors. Specifically, our chapter will pursue four basic lines of inquiry:

1. Can it be stated today that the RNG-PK literature provides evidence for a real effect - one that cannot be attributed to methodological weaknesses, inappropriate statistical techniques, filedrawer problems, and so forth?

2. If we *can* establish the probable reality of an effect, does the evidence favor an influence model, or is it in fact pointing to a receptive form of psi, such as precognition?

3. Are certain methodological approaches more successful than others? Are there physical or psychological moderator variables that systematically relate to micro-PK scores and point to the most appropriate directions for future research?

4. To what degree may we consider the effect replicable, and thereby amenable to study with conventional experimental methods? This issue is intimately related to questions of experimenter effects and subject variability[3].

The birth of a field

A short history of micro-PK research

The underlying approach to micro-PK was anticipated four centuries ago by Francis Bacon, one of the pioneers of the scientific method. Published posthumously in his opus *Sylva Sylvarum,* Bacon states:

> It is good to consider upon what things imagination hath most force: and the rule is, that it hath most force upon things that have the lightest and easiest motions. As for inanimate things, it is true that the motions of shuffling of cards, or casting of dice, are very light motions. (Bell, 1964, p.203).

Bacon's ghost must have had a good chuckle when J.B. Rhine launched micro-PK research with dice in the late 1930s, following a gambler's claim that he could influence the outcome of die-tosses. Initial studies were clearly exploratory, but controls improved progressively, as manual hand-tosses were replaced with dice-tossing machines, independent records of results were introduced, and so forth. Nevertheless Rhine's publication of his results drew highly critical reactions (namely by Girden, 1962) claiming that dice-PK studies were fraught with methodological problems, such as structural biases in the dice and lack of automatic recording of the results. While a number of these critiques overstated the case against the dice-PK experiments, especially in view of the better studies, most researchers were discouraged from investing in this research research. It was clear that improved methods for generating, controlling and recording experimental trials were needed.

The field changed quickly when researchers turned towards an entirely novel generation of PK target systems based on the intrinsically indeterminate processes associated with quantum phenomena, such as the decay of radioactive isotopes. Following several researchers' inconclusive attempts in this direction, the physicist Helmut Schmidt created a relatively compact and flexible device using a strontium-90 source. Schmidt's device used the detection of radioactive decay at a Geiger counter to stop and register the state of a high frequency two-state counter. Since the decay occurs at random intervals, the counter is equally likely to stop in either state, and the output is thus equivalent to the toss of a fair coin. The quantum indeterminacy of the decay, guaranteeing the randomicity and independence of the successive outputs, can thus be readily used as the basis of a precognition task (guessing the outcome of the next 'coin toss'). It could also be used as a PK experiment - the focus of this chapter - by asking the subject to intend or wish for a particular outcome, then initiate a trial with a button-push and receive immediate feedback.

The introduction of RNGs was a major advance for the field. It brought PK research into the era of modern laboratory automatization and resolved a number of outstanding methodological problems by ensuring truly random target sequences, allowing for automated data collection and simplifying collection of control data to verify the RNG's randomness and stability. The RNG could be modified relatively easily to explore different modalities of feedback, trial generation rate, and so forth, thus encouraging experimenters to explore protocols that would otherwise have been excluded as overly labor-intensive. The evolution of RNGs from self-standing, custom-made systems, to commercially available integrated circuits based on electronic noise, permitted their integration with personal computers, further facilitating data-collection and analysis, and opening possibilities for process-oriented research. All in all, the RNG's methodological advantages, simplicity of use and flexibility led to its widespread adoption, and quickly transformed micro-PK research from a near-moribund activity to a major trend in the field - as amply documented in Stanford's (1977) chapter and more recent literature reviews (Schmeidler, 1984; 1987; 1990; 1994; Gissuarson, 1997) and meta-analyses (Radin & Nelson, 1989; Bosch et al, 2006).

Helmut Schmidt

Helmut Schmidt was a key player in micro-PK research for several decades. His work is particularly interesting, not only because he was a highly successful and prolific PK experimenter, producing by far the strongest and most consistent results in this area of research, but also because of the role he played in conceptualizing and modeling the phenomena. It is thus instructive to take a close look at his experimental and theoretical work.

Broadly speaking, there are two main thrusts to Schmidt's research. Early experiments focused on establishing that PK is independent of the internal details of the RNG device and depends instead on the sensory and psychological experience of the subject. His later research, which built naturally on the formalisms derived from his first experiments, focused on the possibility of retroactive PK. In this brief overview we can only give a general feel for the evolution of Schmidt's experimental work and his thinking[4].

Early experiments. The first published report (Schmidt, 1969a) used a four-state RNG. The device was enclosed in a box featuring a set of four small lamps with adjacent buttons that corresponded to the four equiprobable RNG outputs. Subjects would select and press a button to initiate each trial. Upon trial initiation, the device would generate a random output and light the appropriate lamp as feedback. The subject's task was to indicate by the button push which lamp would light. The protocol was thus initially conceived as a test for precognition since, strictly speaking, the random event was generated only after each button push. In two separate studies, tens of thousands of trials were collected and highly significant results obtained (Z-scores of 4.6 and 6.5)[5]. Schmidt, however, immediately recognized the ambiguity here: the results could be interpreted as a precognitive intuition, or as being due to an active PK mechanism tacitly exerted by the subjects and influencing the RNG so that its output conforms to their guesses.

In experiments more explicitly turned to micro-PK, he was particularly interested in determining whether the internal structure or complexity of the RNG affects the individual's ability to influence it (as would be expected in a mechanistic or cybernetic model). In one study Schmidt (1974) explored the impact that two different RNGs would have on scoring: one was a simple binary RNG that output a single bit (0 or 1) while the other rapidly produced 100 bits returning a hit if the bit sum was greater than 50. For each trial, one of the RNGs was selected randomly as the feedback device and neither subject nor experimenter knew which one was momentarily in use. From the subject's viewpoint, all trials were experientially identical: they would initiate a trial by selecting one of two buttons, and immediately receive feedback through the lighting of one of two lamps adjacent to the buttons. An overall Z of 5.2 was obtained, with no significant difference between the two RNGs. These results were strictly due to the RNG responsible for feedback; the one that was not the target of the subject's intention, on each trial, was at chance.

Experiments such as this one (Schmidt, 1969b; 1970a; 1972; 1973) convinced Schmidt that the internal structure of the RNG device is of little consequence, and the key factor for results is the subject's state and experience. He formulated this as an 'equivalence principle':

> If we have two [structurally different, but] truly random generators, operating such that the generators are from the outside physically indistinguishable, then a PK effort affects [each] system to the same degree, i.e., the systems are also indistinguishable in their response to a PK effort. (Schmidt, 1987, p.108).

He concluded that psi effects are goal-oriented, primarily based on the subject's intentions and mental state:

PK may not be properly understood in terms of some mechanism by which the mind interferes with the machine in some cleverly calculated way ... it may be more appropriate to see PK as a goal-oriented principle, one that aims successfully at a final event, no matter how intricate the intermediate steps (Schmidt, 1974b).

Retroactive-PK. The equivalence principle was meant not so much as a theoretical model, but as a guide for devising new experiments (Schmidt, 1978). It led Schmidt to the following thought experiment. Imagine devices with internal RNGs that produce statistically indistinguishable outputs. One device outputs its signals immediately, as soon as they are generated. The other records the generated signals into a large buffer and only begins to output them 24-hours later. Under the equivalence principle, insofar as the device outputs are indistinguishable to the subject, the response to PK effort should be the same. In other words, the equivalence principle provided a conceptual formalism for understanding experiments with prerecorded targets and encouraged Schmidt to study retroactive PK (or, retro-PK) more closely.

One early experiment (Schmidt, 1976) involved both prerecorded and real-time RNG data. Schmidt arranged for a protocol in which the RNG's random output would generate a barely audible click as feedback. Participants were asked to listen carefully for the clicks, and Schmidt's hypothesis was that this mental state of intense attention and anticipation would alter the RNG output so that the frequency of clicks would increase. This hypothesis was confirmed for real-time sessions, in both pilot and formal experiments, with Z's of 3.1 and 3.3. It was also confirmed in the prerecorded condition (Z of 2.9) that was experientially identical for the participants, where the clicks were played back from a prerecorded tape. In another clever experiment (Schmidt, 1976) real-time and prerecorded feedback sounds were randomly mixed. Additionally, the prerecorded sounds were actually presented 4 times, interspersed among the other clicks so that the subjects did not know they were hearing the same sounds repeatedly. Schmidt's hypothesis here, in accordance with his mathematical model, was that the 4-fold increase of focus on a given event would increase the impact on the earlier generation of that event. Both real-time and prerecorded trials were significant, but the scoring rate obtained for the prerecorded (and 4-times heard) trials was indeed significantly higher than that obtained for the real-time generated targets (52.9% vs. 50.8%).

Experiments with prerecorded RNG sequences continued in the following years (Schmidt, 1979a; Terry & Schmidt, 1978; Schmidt & Dalton, 1999). While not all yielded significant results, the cumulative[6] Stouffer Z of 10 studies was over 4, and Schmidt considered that this work made a strong case for his surprising idea that PK could affect data generated at an earlier time.

In his last experimental contributions, Schmidt sought to use protocols with prerecorded RNG targets as a means to include independent investigators, including skeptical ones, in psi experiments. The idea was to design a protocol which would not be susceptible to accusations of investigator fraud, insofar as controls would be shared between several independent researchers. The basic procedure involved several steps. The first was to use an RNG to generate a preset number of binary sequences, that were to remain unobserved until they would be used in a series of experimental PK runs. A sealed copy of these sequences was sent to an independent collaborator, whose role was to ensure the experiment's validity. The collaborator used his or her own RNG to determine the target direction for each PK run, and sent these target assignments to the experimenter, who then ran the session with the subject. When all runs were completed, the collaborator unsealed the copy of the data and performed pre-specified statistical tests to obtain the result of the experiment. Finally, the two datasets were compared to ensure that they matched.

Five experiments with this general protocol were completed (Schmidt, Morris and Rudolph, 1986; Schmidt and Schlitz, 1989; Schmidt, Morris and Hardin, 1991; Schmidt and Braud, 1993; Schmidt and Stapp, 1993). The experiments employed a variety of feedback methods and psychological approaches, as well as different subject populations, so they cannot be considered as exact replications. Still, they all examined whether the protocols can produce convincing evidence of PK for independent, outside collaborators, and potentially be used for subjects working on their own. The cumulative score of the five studies gives a Stouffer Z of 3.67 which gives considerable support for a PK effect under the conditions of a tight protocol involving independent collaborators.

An assessment of Schmidt's work. Taken as a whole, Schmidt's experiments present extremely strong evidence for micro-PK. We reviewed 22 experimental publications containing 50 independent studies, of which 3/4 reported significance above the 0.05 threshold and nearly half had Z's above 3. There is some ambiguity in counting studies and a dozen or so were reported as pilots, but any tally one can make leads to astronomical odds against the null hypothesis. Non-ideal behavior of the RNGs could give systematic biases in the data, but Schmidt was careful to calibrate his devices often and reported control data runs in most publications. At any rate, excluding from consideration studies with inadequate reporting of control data does not change the picture. Some critics were not fully satisfied that the RNGs were reliable (Hyman, 1981), but decades of testing by Schmidt and the many independent researchers who used his devices adequately allays such concerns.

Replication is of course a necessary step to establish any effect firmly. As we'll see in the different meta-analyses we will be reviewing, Schmidt's work inspired hundreds of studies. The results have been quite variable, some being nearly on par with Schmidt's while others being null. In this context, the retro-PK research is particularly noteworthy. It produced much excitement in the parapsychological community for a good 15 years, partly because of the emergence of the observational theories (OTs) in the 1970s and 1980s (see Schmidt, 1982,1984; Walker, 1984; Millar, 2012; Houtkooper, 1993; Stapp, 1994) which gave a conceptual grounding to the work with prerecorded targets[7]. Numerous replications with prerecorded targets were attempted and some reported suggestive evidence for retroactive PK (Davis & Morrison, 1978; Houtkooper, 1977; Talbert & Debes, 1982; Weiner & Zingrone 1986; Weiner & Zingrone 1989), while other studies were null or inconclusive (Bierman 1985; Bierman, 1988; Bierman & Weiner, 1980; Bierman & Houtkooper, 1981; Houtkooper et al, 1980; Millar & Broughton, 1977; Morrison & Davis, 1979; Schouten, 1977).

It is pretty clear, given the consistency and strength of his results, that Schmidt was an outlier among micro-PK researchers; one cannot resist wondering why he was so phenomenally successful with RNG studies. He was certainly a good PK subject himself, and if we accept his suggestion that psi is goal-oriented, independent of the complexity and knowledge of underlying conditions, then it is plausible that he tended to obtain the results he was looking for by exerting his own psi. This, at least, would account for the difficulty some researchers have had in repeating his results, and limit the generalizability of his findings.

We will return to the issue of experimenter psi presently. But an alternative, or at least complementary hypothesis is that Schmidt was simply a very good investigator. His personal investment in RNG research, his creativity in hypothesis testing and his sheer perseverance over the course of three decades, may have honed his ability to tease out effects that are subtle and difficult to reproduce, but quite real. Furthermore, although the most prominent aspect of his work was the search for simple principles relevant to physics, a closer reading reveals a highly intuitive approach to the psychological facets of PK research and a keen sense of how best to work with subjects.

Schmidt's stance with regard to eliciting psi from subjects was straightforward and practical: psi is neither egalitarian nor available on demand, and experiments should be run with this in mind. Among the strategies employed, foremost was the use of people with established success in micro-PK tests. He sought out and then tested mediums, psychics and people who reported extraordinary experiences. When working with larger subject pools, selection was frequently based on systematic preliminary tests. Schmidt was capable of investing months of his time for a single experiment testing many dozens of people before settling on a handful for the experiment:

For my own experiments, I found it inefficient to gather data from a very large number of people, because poor scores of the majority tend to dilute the effect of the successful performers. Therefore I pre-selected promising subjects, and then used these subjects immediately in a subsequent formal experiment with a specified number of trials. Unfortunately, the process of locating and pre-selecting promising subjects is time consuming and often frustrating. (Schmidt, 1987, p. 105).

Besides this subject selection strategy, Schmidt was particularly careful to provide an inviting and friendly environment for participants. In some cases he would arrange to do experiments in peoples' homes, and make himself available on short notice should a subject find herself well-disposed for a session. Participants could also postpone a session if they did not feel ready and were also given latitude in deciding on preferred feedback. In some instances, a session would only be initiated if a preliminary test was successful, and subjects were generally encouraged to set their own pace and take breaks or chat with an experimenter if they felt tired or bored.

[T]he subjects were tested only under conditions that they liked. In particular, in order to avoid unpleasant and frustrating associations with the tests, they were not forced to operate under all of the different test conditions under study. In order to obtain statistically meaningful results concerning the existence of PK in the data, the total number of trials under each test condition was pre-specified, but it was left open how much each subject should contribute to the test. This arrangement would correspond to what might appear the most efficient procedure in future applications of psi: one would select a number of promising subjects in pretests and then use each subject when and as long as he seemed to be in a particularly favorable state for high scoring (Schmidt, 1973, p. 107).

Schmidt used this strategy extensively: he allowed for variable contributions from individual subjects, for the interruption of sessions, and even for participants to be dropped from an experiment if performance lagged. Session stops and subject culling probably increased the portion of data collected under psychologically favorable conditions, and permitted Schmidt to use his intuition (or psi) to decide when to intervene in sessions[8]. That Schmidt was personally successful at PK tasks undoubtedly contributed to this intuitive sense of how to design and execute his experiments.

We may summarize Schmidt's strategy rather simply: he aimed to study the underlying principles of micro-PK and address questions of temporality, causality and the goal-oriented nature of psi; to do so, he needed strong effects and did whatever it took to obtain them - including, using himself as subject. Whether tacitly or explicitly, Schmidt understood the psychology of getting results - through subject selection, skillful creation of good psychological conditions, and flexibility in hypothesis testing (e.g., favoring psi-missing rather than psi-hitting, when circumstances seemed to call for this). These considerations must nuance our assessment of subsequent replication rates from other investigators: the tacit actions and judgments that are clearly important determinants of experimenters' success and subjects' performance are difficult to integrate into fixed protocols.

On the other hand, this 'psi at all costs' strategy, does have drawbacks. It is poorly adapted for systematically studying the range of psychological or cognitive factors that may impact psi and begs the question as to whether micro-PK is universal or just for the gifted few. It raises the distinct possibility that many of the effects discovered are basically expressions of the experimenter's psi. And it complicates independent replications insofar as it is unlikely one can identify and reproduce all of the elements that went into his experiments: motivation, experimenter skills, intuition, and perhaps, tacit psi ability.

The PEAR lab, in stark contrast to Schmidt's strongly personal approach, was committed to unselected subject populations, stable protocols, and a rather patient, cumulative philosophy for exploring micro-PK.

Micro-PK as universal

The PEAR research

The Princeton Engineering Anomalies Research (PEAR) laboratory was founded by Robert Jahn, Dean of the engineering school at Princeton. He assembled a staff of physicists, psychologists and technicians who worked in a basement lab of the campus engineering building for nearly 30 years (the PEAR lab closed in 2007). PEAR's primary focus was micro-PK RNG research, although extensive effort also included work on macroscopic random systems (some reminiscent of the Rhine era), and remote viewing studies.

Early on, the laboratory adopted a number of orientations and ground rules for its operations. A strict universalist approach was adhered to, using only unselected experimental participants ('operators', in the lab lingo) whose participation depended essentially on their own availability and willingness. In order to have a full appreciation of experimental conditions, lab staff were required to contribute to experiments as participants. Like Schmidt, PEAR built its own RNGs in-house[9]. Extensive testing and calibrations were published and there is little doubt that the devices were reliable and stable enough to detect the small effects sizes expected in the universalist approach. Also, the PEAR lab had a firm policy of publishing all its experimental results in either refereed journals or in publicly accessible internal reports. This lends a significant added value to the lab's contributions since the huge volume of research produced can be considered free of publication bias and filedrawer problems.

The RNG used was a binary device which produced feedback as either a counter readout or a graphical display. Data were generated at the rate of one trial per second and consisted of the sum of 200 consecutive random 0 or 1 bits (the expected mean trial output thus being 100)[10]. All the RNGs in the PEAR lab used exclusive-OR (XOR) processing of the raw bits before summed to make a trial, to eliminate potential first order biases in the bit streams.

PEAR used a 'tripolar' protocol for each experimental run, which consisted of three separate PK efforts of equal length. These were termed HI, LO and BL (baseline) and indicated the direction of the participants intention: bias the output to go high, to go low, or to remain even. The experimental hypothesis was then that the HI runs would give a positive deviation from the mean and the LO runs a negative deviation. The statistical test was based on the difference of the two directional runs. The BL run was intended as a control and the participants instructions were to hold no intention while the baseline data accumulated.

A particularity of the lab was to conduct long-term studies of a single protocol over many years. Rather than designing short studies with a preset number of trials, major experiments were cumulative, and occasional updates of progress would be communicated in reports or publications. While not all work followed this open-ended model, the most important and consequential micro-PK experiment did (Jahn et al., 1997). This was the so-called 'benchmark' experiment, a twelve year study that collected over 2.5 million experimental trials from 91 participants. At its termination, the experiment had attained a high significance, yielding a Z-score of 3.8.

Aside from the benchmark study, over the course of the years the PEAR lab explored many other of issues concerning psi functioning, including the kind of fundamental questions that so interested Schmidt. For example, they undertook an extensive study of whether micro-PK depends on the distance by performing experiments with subjects at remote locations. Remote-subject data was collected over the course of a 6 year period (Dunne & Jahn, 1992), and involved participants from all over the globe, with remote locations ranging from 1 to 9000 miles from the Princeton laboratory. The protocol specified that, at an agreed upon time, a staff member would start a run of, typically, 1000 trials and then leave the room where the laboratory RNG was housed. The participants wished-for a specific sequence of high, low and baseline aims (that they chose themselves, but had no feedback from the RNG. Once data for the full series of three (HI, LO, BL) target conditions were collected (a process that took one hour) the remote subject would inform the experimenter of the order of his intention - thus ensuring that the experimenter had remained blind to the intended direction of results until after these had been recorded.

Thirty participants contributed a total of 265 series. Results were significant for the high aim, and null for the low and baseline aim, but the HI-LO difference produced a significant Z-score of 2.2. No dependence of effect size on distance was found, and results were essentially similar to those obtained with the usual benchmark protocol, where the participant was seated in front of the RNG[11]. A subset of these series also included *temporal* displacements - a theme familiar from Schmidt's work - in which the time of the subject's efforts was offset from the time of data collection. Time-displaced data comprised about 20% of the remote-subject sessions (a total of 60 series, representing 100,000 trials), and included time offsets from a few minutes to many days before or after the RNG sessions were triggered. The off-time sessions were also consistent with the benchmark experiment. The time-displaced data, like the distance analysis, found that the time of effort and the significance of the effect were uncorrelated: long time intervals between subjects' efforts and RNG sampling were as likely to produce effects as shorter ones.

The Consortium replication

In 1996 the PEAR lab formed a research consortium with two German groups, the Institut für Grenzgebiete der Psychologie und Psychohygiene in Freiburg, and the Center for Psychobiology and Behavioral Medicine at Justus-Liebig Universität in Giessen. The consortium planned a broad program of collaborative micro-PK research, and undertook an extensive replication of the PEAR lab's benchmark local-subject experiment as its first project. The original experiment had collected nearly a million data trials in each condition of the HI-LO-BL protocol (aim high, aim low and baseline control). The replication used the same protocol, uniform RNGs among the groups and stipulated an equal contribution of trials from each laboratory. The primary hypothesis was maintained: trial values would deviate according to the subjects' intentions, with the test statistic being the difference of HI and LO scores. The goal was to accumulate a pre-determined amount of data of roughly the same size as the PEAR experiment. This gave the replication about a 95% chance of succeeding at a P-value level above 0.01.

Taking a cue from PEAR's previous analyses, pre-planned secondary hypotheses would look at how the data varied with experimental parameters such as the feedback mode or the length of data runs. All told, six secondary hypotheses were planned to test for anomalous, non-random structure in the database.

During three years of intensive data collection, 750,000 trials per condition were collected from 227 participants. After a period of careful analysis, the consortium published its much-awaited report (Jahn, et al., 2000), but the results were disappointing. Although all three groups found positive deviations, the effect size was nearly an order of magnitude smaller than expected and the overall Z-score came in at an insignificant 0.6. The evident failure to replicate a solid and well-founded hypothesis seemed paradoxical. Why would a well-powered, exact replication fail to reproduce the previous results? And what, then, to make of the universalist claim that, with enough data and a broad sampling of people, obtaining positive results should be a straightforward matter?

Despite the failure to confirm the main hypothesis, a statistical combination of the six preplanned secondary analyses resulted in a P-value of 0.02, suggesting that anomalous deviations *were* present in the data. Two of the planned hypotheses were primarily responsible for this result. One was a 'series-position' effect in which effect sizes decrease as participants complete successive data series. This decline effect, presumably due to subjects' initial enthusiasm giving way to a loss of interest, had been seen in the earlier PEAR data and is of course well-known in psi research. The second planned analysis, inspired by the PEAR data, looked at variability of the effect size under different conditions of feedback, run-length, gender and target assignment. It suggested that results varied with experimental parameters in a way that could not be explained by chance.

No replication can exactly repeat all experimental conditions, especially when human subjects are involved. Was there an essential difference between the experiments that had been overlooked? The consortium looked at a number of possibilities, but found argued that none was compelling. Minor changes were implemented in RNG design and feedback selection, but from past experience these alterations should be inconsequential. The participant pool was of course different, but the same approach to unselected, volunteer participants was used, as was the approach to establishing an ambiance in the laboratory conducive to PK. Lacking an obvious explanation for the replication 'paradox', the consortium turned to the suggestion that micro-PK, and psi in general, might belong to a class of phenomena that cannot be replicated in the conventional understanding of 'objective' scientific investigation. They conclude:

> The change from the systematic, intention-correlated mean shifts found in the prior studies, to this polyglot pattern of structural distortions, testifies to inadequate understanding of the basic phenomena involved and suggests a need for more sophisticated experiments and theoretical models for their future elucidation (Jahn et al., 2000).

The statement is less anodyne that it appears: it suggests that psi may be not only slippery and elusive as a practical matter, but even *inherently* so. A considerable amount of literature on this idea (Kennedy, 2003; Atmanspacher and Jahn, 2003) discusses how replication problems relate to the subjective or self-referential nature of mind-matter interactions, and some suggest that psi is fundamentally elusive (von Lucadou, 2006). However, we raise here an important point that, in this case, seems to provide a simpler explanation.

The replication's validity depends on two assumptions. The first assumption is that the original participant pool was representative of the general population, and the second was that the benchmark effect size was a good estimate of an average participant's PK effort. Under these assumptions, any new group of participants should provide about the same effect size, and the appropriate size of a replication can be determined by power analysis. A problem, however, is that there were two extreme outliers in the PEAR benchmark participant pool. Together they contributed nearly a quarter of the data and more than half of the total HI-LO deviation. Their personal databases are highly significant with Z's of 5.6 and 3.4. It is easy to see that they are not representative of the 89 other participants since the overall Z of the remaining database is only 0.8. Ironically, the 89-participant database has nearly the same size and Z-score as the replication. But the point is that, if the outliers had been removed before designing the study, an appropriate effect size would have been estimated, and the replication size would have nearly quadrupled. Based on this observation, the replication failure can be interpreted not as the result of elusive micro-PK, but simply as due to an under-powered study.

The consortium replication was clearly important, as it attempted what is generally considered a gold standard for proof-oriented research: the well-powered replication. Its failure was disappointing, but did provoke a healthy discussion around the issues of replicability. Considering the outlier subjects gives at least a plausible explanation for why the replication failed, without undermining confidence in the earlier data Furthermore, "failure" in this instance might be too strong a term. First, as we have seen, several internal structure findings were replicated. Second, it should be stressed that, while it called into question the reliability of replication, it did *not* call into question the original evidence seen in the benchmark PEAR databases. Indeed, when the results of the two experiments (PEAR and consortium) are combined, they produce Z of 3.2 which remains a highly significant effect.

The Big Picture

Meta-Analyses of micro-PK research

We now turn to the analytical evidence for PK as provided by meta-analysis (MA). Three major micro-PK meta-analyses in micro-PK have been published (May et al., 1985; Radin & Nelson, 1989; Bosch et al., 2006). We will focus on the latter two (henceforth, RN and BSB) since the mainstream journals where they appeared published peer reviewed critiques that generated interesting debates.

Generally speaking, meta-analysis (MA) permits an estimation of effect sizes from a large number of independent studies; when these studies are small or low-powered, it can combine them to approximate the power of a larger study. Underlying assumptions are that studies measure the same effect, are of good and comparable quality, and are representative of the research that has actually been done. In practice, it can be difficult to ensure that assumptions are satisfied and this is often a major challenge for MA. An example is the well-known publication bias problem, whereby the literature contains too many significant reports because insignificant studies have difficulty getting through the publication process. Such artifactual 'selection' biases can greatly inflate effect size estimates and lead to erroneous conclusions about effects.

RN included 597 micro-PK studies published up through 1987. A quality assessment found that effect sizes were independent of study quality, indicating that the MA results would not be adversely impacted by quality problems. The failsafe filedrawer estimated that for each published study 90 unpublished studies would be required to cross the failsafe threshold. This is a huge failsafe number that, at face value, seems reassuring. Their estimate of the effect size had a Z of over 6. RN conclude, a bit over-optimistically, that the effect is both robust and replicable:

> The overall effect size obtained in experimental conditions cannot be adequately explained by methodological flaws or selective reporting practices [...] This meta-analysis shows that effects are not a function of experimental quality, and that the replication rate is as good as that found in exemplary experiments in psychology and physics. (Radin & Nelson, 1989, p.1510)

The paper received a good deal of attention and in 2003 the authors published an update introducing an additional 176 studies (Radin & Nelson, 2003). They confirmed their earlier conclusion regarding the quality assessment but, using a more current method, estimated a failsafe filedrawer per publication of only 5

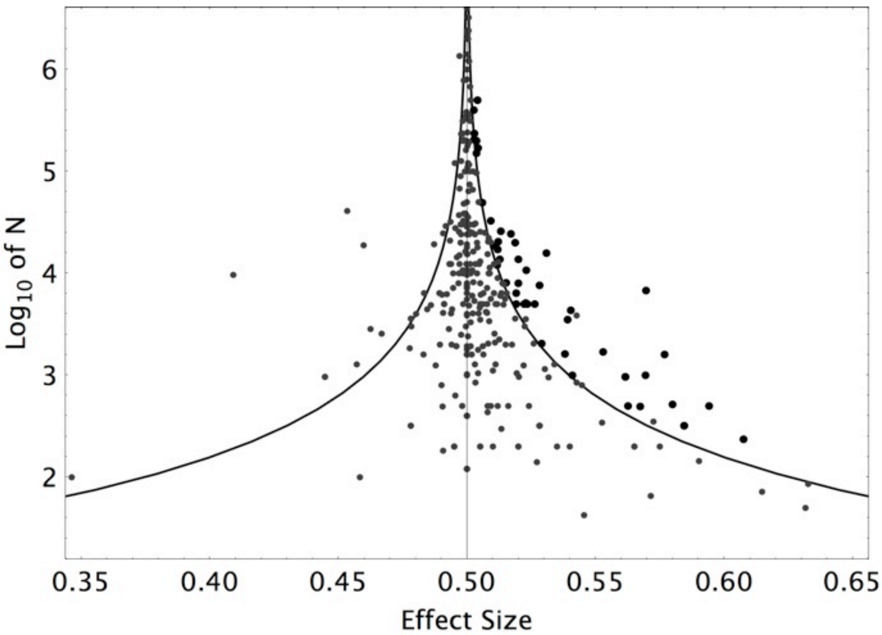

Figure 1. Funnel plot of study effect sizes for the BSB database. The solid curves are lines of constant Z=2.6 (P-value 0.005) and separate high and low Z regions. An indication that effect sizes are dependent on study size is evidenced by the rightward skew of the points. The logarithmic vertical scale accomadates studies that vary of several orders of magnitude.

(rather than the previous 90). They claim this value is nevertheless large enough to satisfy worries of publication bias and maintain their confidence in the 'repeatability' of PK.

A long critique of the Radin & Nelson meta-analyses was published by Schub (2006). One of his main critiques is a lack of consistency between different meta-analytical methods. He shows that three different methods[12] for calculating the significance of the effect give wildly different Z's of 16, 6 and 1.3. For Schub, this extreme inconsistency between methods indicates that there are serious problems with the data that violate the assumptions of MA's.

At about the same time an extensive, independent MA was published in *Psychological Bulletin* (Bösch, Steinkamp & Boller, 2006; referred to as BSB). Two of the authors were members of the consortium replication and, in light of the replication's failure, their goal was to determine whether MA could provide evidence for PK. Following the replication, they restricted themselves to local-time, local-subject studies. The analysis for fixed and random effects models on 377 studies yielded effect size Z-scores of 4.1 and 3.6, respectively.

BSB's conclusion, similar to Schub's, was that publication bias, indicated by the extreme, skewed heterogeneity in the funnel plot (Figure 1), was responsible for the effect. They modeled this artifact with a Monte Carlo simulation by allowing for bias jumps at P-values of 0.05 and 0.01. The simulation found that 1544 unpublished papers could explain the effect size (a failsafe ration of 4) and produce a skewed heterogeneity. To quote their conclusions:

> Publication bias appears to be the easiest and most encompassing explanation for the primary findings of the meta-analysis. If the time comes when the funnel indicates a systematic effect, a model to explain the effect will be more than crucial. ... [For now,] Girden's (1962) verdict of "not proven" (p. 530), which he mooted more than 40 years ago in the same journal with respect to dice experiments, also holds for human intentionality on RNGs.

In rebuttals to Schub and BSB, Radin et al (2006a; 2006b) argue that funnel asymmetry can have other origins than publication bias, and explain that some models of PK can in fact give this effect (such as some versions of the goal-oriented models mentioned above). At the heart of the rebuttal is the contention that a simple bitwise deviation, assumed by both BSB and Schub, might not be the way PK actually behaves. For MA, this possibility changes the interpretation of heterogeneity, and raises questions about how to define the effect size. To add weight to their view, Radin et al (2006a) make the valid point that BSB's publication bias model is unsatisfactory since it leaves nearly half of the heterogeneity was unaccounted for. After the exchanges, both sides remained unconvinced by the opposing arguments.

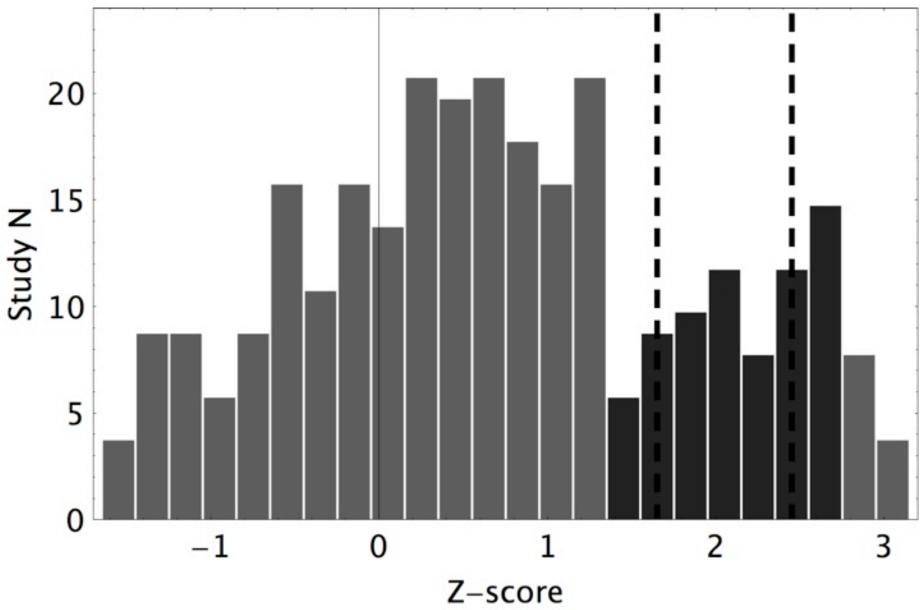

Figure 2. The Z-score distribution of studies for the BSB database illustrating evidence for publication bias. The dashed vertical lines are at Z's with P-values of 0.05 and 0.01. In the absence of bias, the distribution tail above a Z of one should decrease smoothly. Publication bias is signaled by an increase in the distribtion near P-values associated with publication acceptance (dark bars). This artifact is clearly visible around the standard P-values of significance threshold.

To shed light on this standoff, we provide our own re-assessment of the MAs. It is clear that correctly treating the asymmetric heterogeneity is key to resolving the debate, and that the heterogeneity limits MA's effectiveness in providing evidential support for micro-PK. Subsets of the database might be amenable to effect size estimation if they are sufficiently homogeneous. But for the full database, the best use of MA is to understand the sources of heterogeneity, rather than to estimate effects. This was ultimately the approach taken by BSB and, despite a tendency of RN to downplay the issue, they are right in asserting that publication bias and other selection mechanisms constitute a serious problem. Figure 2 shows that there is fairly good evidence for publication bias in the database; this needs to be taken fully into account when arguing for micro-PK. On the other hand, RN are right to insist that BSB fail to successfully model the data heterogeneity. A substantial source of heterogeneity, we suggest, lies in those studies that cannot be reasonably accounted for by publication bias - namely the studies with high Z-scores. About 40 of the total dataset have Z's between 2.6 and 5; these contribute nearly half of the total heterogeneity. To generate these studies by artifactual selection would require an extremely large filedrawer. The BSB simulation ignores them, which is why the residual heterogeneity of their model is so large[13].

Given the relatively large proportion of these studies in the database, and the difficulty of explaining them by artifacts, we feel that the *meta-analytic* evidence for micro-PK is in fact quite strong[14]. However, to quantify the strength of this evidence requires modeling the tail of the Z-score distribution, rather than the full distribution, where publication bias and other artifacts are at issue.

A model for micro-PK?

While we conclude that there is strong evidence for a real effect in the micro-PK database, we have not addressed the *nature* of the effect. Among the competing models, two are of particular interest for their simplicity and contrasting predictions: influence models, that propose perturbations in the RNGs behavior, and selection models which claim that the RNGs remain unperturbed. According to the latter models, (often called Decision Augmentation Theory, or DAT) subjects use precognition to time their actions (such as a button press to start a trial), in a manner such that they 'select' a data segment with a deviation that happens to accord with the target direction.

Statistical tests that can distinguish between the models have been devised by May et al. (1995a; 1995b). In both these simple models, the respective effect sizes are predicted to be constant, but the analysis units are different. In the Bitwise Influence (BI) model, the bit is the natural data unit: a PK effect shifts the mean RNG output, and the shift is modeled as a small, uniform bitwise change in the binary probabilities. Under the DAT model, the subject decides to start her experimental run at a time when the data will fluctuate favorably. The natural data unit here is the run (or other data segment corresponding to an initiatory decision on the part of the subject).

To test the models against each other, the data are first divided into blocks of DAT units, and the Z-score for each block is calculated. A linear regression is performed of Z against $\sqrt{N_{bits}}$ (where N_{bits} is the number of bits in each block). The models predict a regression slope of zero if the data follow the DAT model, and a positive slope for the BI model.

May et al. ran the regression for a dataset of 128 micro-PK experiments and found that their DAT model gave a better fit. A problem, however, is that the fit is in fact quite poor due to a large scatter in the data which effectively gives a low power to the test. This is reminiscent of the limitations imposed by heterogeneity in the meta-analyses. The test can be improved by using homogeneous databases that use a uniform protocol and are free of artifactual selection problems. Fortunately two exist: The PEAR benchmark database and that of the GCP.

The DAT test has been done for both of these and each favors the BI model fairly decisively. The PEAR analysis (Dobyns, 2000) rejected DAT with a Z-score for the significance of the regression slope of 4.4. The same analysis on the GCP data favored[15] the BI, with a Z of 2.6 (Bancel, 2014a). While the PEAR and GCP data give some evidence for bitwise changes in RNG output distributions, it should be stressed that this only applies to these two homogeneous datasets. It is also possible that micro-PK effects include contributions from both BI and DAT mechanisms.

Also, it should be stressed that the evidence does not necessarily argue for a forward-in-time causal mechanism. Etter (Shoup & Etter, 2000) has proposed an experiment to test for retro-causal PK and a series of these experiments were reported by Radin (2006). Radin constructed a binary RNG that had several internal RNGs operating serially (a Markov chain) so that the final output probability was p=1/2. The experiment recorded the state of all the internal RNGs in the chain for each trial. Shoup and Etter had shown that data from the internal RNGs would be different for forward- and retro-causal effects. Radin's analysis favored retro-causality, a result with potentially profound consequences. However, since not all of the complex experiments in the series were successful, we should be cautious in drawing firm conclusions. It would be interesting to see more work along these lines.

A very different theoretical approach has been proposed and tested by von Lucadou (2007). His Model of Pragmatic Information (PMI) combines ideas from systems theory and a generalized version of quantum mechanics that is applicable to situations where humans have meaningful interactions with their environment. The theory proposes that these interactions are akin to quantum entanglement and explains micro-PK effects as non-causal, non-local correlations that are measurable, but inherently elusive; they defy attempts of direct replication. In a series of experiments, von Lucadou (1995; 2006) gives evidence in support of his theory. These experiments, which are on-going, are rather involved and it would be good to see a thorough accounting and independent verification of the analyses.

Process-oriented research: enhancing micro-PK

Historically, two research approaches have dominated PK investigations. The earliest was elitist, involving intensive work with a few, highly selected subjects. This was, of course the dominant approach in investigations of physical mediums and macro-PK and one of Helmut Schmidt's preferred strategies, as we have seen[16]. One would think that Schmidt's success would encourage researchers to systematically seek out gifted subjects, but admittedly this is easier said than done[17]. Other than employing widespread PK screening tests (Schmidt's approach), the usefulness of simpler means, such as inventories for personality traits or cognitive style, seems quite limited[18].

At the other extreme of the elitist strategy lies the strategy of working with unselected subjects, very small effects, massive data collection, and analytical tools to tease out patterns in the data. It could be argued that the PEAR / Consortium studies justify pursuing this universalist approach, insofar as they did produce evidence for a small effect in the general population. However, given the weakness of the effects observed, and the limited resources of the field, a purely universalist approach will likely be ineffective. A more promising path is to explore methods and protocols that enhance subject scoring - perhaps in a manner analogous to the Ganzfeld in ESP research. This 'third way' between the elitist and universalist strategies has essentially been addressed through process-oriented studies, which attempt to specify moderator variables associated with micro-PK scoring.

The first and perhaps most obvious of these is the subject's motivation. Rhine and his collaborators noted from dice studies that results were typically high in the first few trials of a run, and then declined, a finding considered indicative of a loss of motivation over the course of a run or series (Rhine & Humphrey, 1944). Such effects have also turned up in micro-PK research, whether within a single experimental run or across series (André, 1970; Honorton & May, 1976; Broughton & Perlstrom, 1985; Dunne et al., 1994; Jahn et al, 2000, Bancel, 2014b). These position effects suggest that the strong scores in early runs come while individuals are fresh and excited - and before a chain of unsuccessful trials discourages them. It would seem useful to integrate this rather obvious point by using very short testing periods or closely monitoring the subject's state and interrupting sessions with the first sign of boredom or fatigue.

Another psychological factor that has been found to relate to micro-PK is stress. Stanford & Kottoor (1985) ran 60 participants under one of three conditions: a classic RNG task; the RNG task with an abrupt auditory disruption that participants had not been warned about; and the same task again but with a prior warning. They found significant PK missing with the unwarned disruption (the most stressful condition). Broughton & Perlstrom (1986; 1992) compared micro-PK scoring of people who were susceptible to stressful states (as measured by two anxiety inventories) and those who were not. It was found that participants prone to anxiety were more likely to show psi missing in a competitive PK tasks than those who were not as anxiety-prone.

Realizing early on the impact of stress, in some of his earliest experiments Schmidt explicitly *provoked* stress in his subjects in order to induce strong negative scoring (Schmidt & Pantas, 1972). Of course, there is an alternative, too: if stressful or anxious mental states lead to null results or even psi-missing, perhaps states that are more relaxed and composed might lead to positive micro-PK results. In his 1977 article, Stanford reported several studies suggesting that positive scoring is related to states of 'effortless intention' (Camstra, 1973; Andrew, 1975; Steilberg, 1975; Braud, et al, 1976) or meditative practice (Matas & Pantas, 1971; Schmidt & Pantas, 1972; Honorton & May, 1976). Subsequent research has generally supported this association. Braud & Braud (1979) reported significant PK hitting in association with a 'right brain' mental set, induced through instructions for relaxation or a 'passive volition' mental set, whereas an effortful, 'left-brain' set led to chance scores or to psi-missing. Debes & Morris (1982) also obtained significant PK hitting when subjects were given relaxation instructions, and significant missing following effortful instructions. Also, most studies implicating meditators or meditative practices support earlier findings (Schmidt & Schlitz, 1991; Bancel, 2014b; Radin, 2012; Tressoldi, 2014). Although a few exceptions do exist (Braud & Hartgrove, 1976; Winnett & Honorton, 1977), PK studies with meditators generally suggest that a focused but relaxed mental state of 'effortless intention' is more productive than an effortful state where one is preoccupied with performance. Perhaps, also, meditative practices help counter decline effects, as they train the mind to sustain attention and focus on a single thought or task.

A related factor is mental strategy. Schmidt repeatedly suggested that micro-PK is goal-oriented:

> The finding that the success rate is rather independent of the physical structure of the generator... suggests that goal orientation may be a feature of the underlying mechanism rather than a matter of mere psychological attitude. It appears as if the subject, by concentrating on the final outcome, could induce nature to let the previous random events properly fall into place such as to lead to the desired outcome (Schmidt, 1987, p.109)

It is thus noteworthy that several studies have indeed explored goal-oriented strategies, that emphasize a focus on the desired end result and 'letting nature take its course' (Levi, 1979; Morris, Nanko & Phillips, 1978; Nanko, 1981; Morris, Nanko & Phillips, 1982). These have been sometimes contrasted with strategies emphasizing an ongoing process (imagining a flow of energy emitting from the person and affecting the RNG outputs). The results for the goal-oriented strategies have generally been significant and superior to those emphasizing an energy flow; 7 out of 8 studies exploring these strategies have produced results in the positive direction (Gissuarson, 1997).

In sum, these three related factors - a passive volition set, meditative practice, and a goal oriented mental strategy - are quite consistently related to positive PK scoring. Coupled with the obvious motivational factors, they point to several fruitful directions for enhancing individuals' results in micro-PK.

The omnipresent issue of experimenter psi

In our discussion concerning psi-favorable moderator variables, we have sidestepped a major contributor to scoring: the experimenter. From our presentation of Schmidt's work, it seems obvious that experimenter skills can have a major impact on subjects' performance. But it also suggests the very real possibility that results are related to the experimenter's own psi abilities.

While the issue of unintentional experimenter psi has been amply discussed (Kennedy & Taddonio, 1976; Parker, 2012), it is particularly conspicuous in micro-PK research which typically involves extremely weak effects and in which the most successful investigators often show very good results when testing *themselves* for micro-PK[19]. As Kennedy (1976) puts it:

> The case for experimenter PK seems clearly drawn when one considers that experimenters are typically more motivated than their subjects to achieve good results, that PK need not involve a conscious intent, and that most successful PK experimenters are themselves successful PK subjects. (p. 17)

A case in point is Schmidt's animal-psi work (1970b, 1973, 1979b), associating RNG outputs with the needs of different organisms (a cat, cockroaches, brine shrimp, yeast cultures or fruit flies). It seems far simpler to imagine Schmidt as the single source of the effects observed rather than the organisms tested, especially since results seem to have varied as a function of Schmidt's own mindset[20]. Similar interrogations arise with other investigations of animal-psi, as in the innovative work of René Peoc'h (1988, 2001)[21]. The point is not that we consider animal psi as an implausible hypothesis: if we accept that micro-PK is goal oriented, but not limited to conscious intentions and complex cognitive processes, it is conceivable that animals' needs bias the RNG outputs. Rather, the issue is that it is difficult to decide on the true source of an effect, once we recognize that experimenters probably do not turn off their PK skills when it comes to testing their ideas.

Similar interrogations arise in investigations with human subjects. The studies by PEAR and other researchers are quite consistent in suggesting that micro-PK is independent of distance. Yet in several of these studies, it is conceivable that, rather than the distant subject, the effects obtained are due to the experimenters who trigger the RNG *locally*, and tacitly bias its outputs in accordance with their hypotheses[22]. Or, they may affect the RNGs at the moment of data analysis: if we accept retro-PK, we can imagine how an investigator, hopeful to find a significant effect in a particular 'cell' of the data, will unintentionally influence those data during analysis.[23]

Even when micro-PK seems to correlate with a specific subject variable, the possibility of experimenter psi remains. One frequently cited case is a study by Honorton & Barksdale (1982). Honorton hypothesized that PK would be associated with muscular tension. He tested six subjects alternating between muscle tension and relaxation while attempting to influence an RNG. Results were significant specifically in the tension condition. Subsequently, however, Barksdale ran ten subjects in the same experiment, and obtained no evidence for micro-PK in any condition. Honorton was then tested as the sole subject and obtained significant positive results in the tensing condition and significantly *negative* ones in the relaxation condition. Together, these studies suggest that Honorton was the real source of the effect all along, unintentionally confirming his expectation whether as subject *or* as experimenter.

The issue of experimenter psi seems particularly relevant in studies sampling 'silent' or 'hidden' RNGs, while experimental participants are engaged in other tasks. Significant results have been reported in a number of laboratory studies (Honorton & Tremmel, 1978; Varvoglis & McCarthy, 1986; Berger, 1988; Varvoglis, 1988), but also in studies with hidden RNGs placed in non-laboratory settings ('power spots', personal development workshops, sport-event stadiums). The Global Consciousness Project is by far the largest and most sustained example of these field investigations of micro-PK[24]. The question is, do these studies point to a subtle "PK field", that is that spreads outwards from individuals under certain attentional or emotional circumstances and indiscriminately affects physical systems[25]? Or do they simply point to experimenter psi, (whether real-time or retroactive)? Clearly, in *all* these studies RNGs are *not* hidden to the investigators; the latter are well aware of them, and quite invested in their outputs.

To the extent to which experimental results potentially reflect the PK input of experimenters as 'hidden subjects', we are confronted with an inherent ambiguity in the interpretations of experimental results. How do we distinguish subject-based effects, hopefully representative of a larger population and lawful phenomena, from effects that may be limited to the experimenters themselves, and dependent upon the very hypotheses they pose? Clearly, the possibility of experimenter-psi challenges traditional research methodologies and raises the question whether radically different epistemological approaches are needed (Atmanspacher & Jahn, 2003; von Lucadou, 2001). In this regard, psi research may be pointing to the need for reconsidering the experimental paradigm of parapsychology altogether.

Conclusions

In our introduction we posed several key questions for micro-PK research. The most basic of these is the status of the evidence. We have seen two lines of support for a micro-PK effect.

The first is in the PEAR data, which is based on a sound, tested methodology and is free from publication bias. The benchmark experiment together with the consortium replication had a significant combined Z-score, and both databases showed secondary structures that deviated significantly from chance. Additionally, other PEAR experiments, such as the remote subject database, also produced strong results.

The second line of support comes from a revised perspective on the micro-PK meta-analyses. We saw that an understanding of the heterogeneity points us to a reliable subset of the literature which is relatively free from counter-arguments based on publication bias. A chance explanation for the high-Z studies, taken collectively, is highly implausible. The remainder of the studies do, however, exhibit clear evidence for publication biases and this limits their usefulness for meta-analysis. For this reason and the evident difficulty in defining effect sizes, we feel that meta-analyses cannot currently provide reliable estimates of micro-PK effect sizes or guidance for large replications based on power analysis. We have seen a simple but striking example of this in the failure of the consortium replication of the PEAR data: from our perspective, this may well have been due to an inadequate power analysis that failed to identify outlier contributions from two high performance subjects.

The difficulty of replication is evident throughout the micro-PK literature, but the consortium experiment is unique in pointing to a plausible reason for non-replication - and in suggesting that we should not underestimate the role of subject selection. We have seen that selection of strong subjects clearly played a key role in the experimental success of Helmut Schmidt and in one instance at least, his test-retest approach to subject selection was integrated into a formal protocol which yielded results on a par with his (Haraldsson, 1970). Thus systematizing subject selection is clearly a direction that would be interesting to pursue.

Schmidt was also skillful in creating conditions for experimental success, through his flexible approach to protocols and his manner in interacting with subjects. We sought to highlight these experimenter skills, as these are not always conveyed in reports and represent a body of tacit knowledge that is not fully integrated in attempted replications. However, the probable contribution of experimenter psi is equally important to consider, not only in the case of Schmidt but, one suspects, in the case of other successful researchers as well. Although it is difficult to give a clear demonstration, it seems quite plausible that the micro-PK effect is globally dominated by strong PK performers, be they subjects or experimenters.

As for the alternative, universalist approach, we do have some indications for its potential. These are best seen by the analyses showing additional anomalies in the PEAR benchmark and the consortium data, such as series position effects. Since subjects in these two cases were unselected, this gives support for a research strategy which assumes the presence of a weak psi effect, broadly distributed in the population. In this context, however, it seems essential to create testing conditions that enhance performance for relatively unselected subjects. Some directions are worth exploring include: motivation enhancement, mental strategies such as 'passive volition', goal-oriented visualization techniques, meditation, and flexible conditions that adapt to subjects' psychological states[26].

The most interesting question is the one for which we have the least to offer by way of an answer: how do we explain micro-PK? We saw how two opposing perspectives, influence and precognition, can be tested under the specific models. It is clear that clean datasets with adequate power are necessary for these kinds of tests, and in the two available, the tests favor the influence models. Nevertheless, we are very far from being able to claim to understand microPK, and several competing (or complementary) models need to be systematically evaluated, such as retro-causal models, observational theories and generalized quantum theory. We can only hope that more theory-driven research will emerge in the near future and elucidate the nature of microPK.

References

André, E. (1972). Confirmation of PK action on electronic equipment. *Journal of Parapsychology*, 4, 283-293.

Andrew, K. (1975). Psychokinetic influences on an electromechanical random number generator during evocation of "left-hemispheric" vs. "right-hemispheric" functioning. In J.D. Morris, W.G. Roll, & R.L. Morris (Eds.), *Research in Parapsychology 1974* (pp. 58-61). Metuchen, N.J. : Scarecrow Press.

Atmanspacher H. & Jahn R.G. (2003). Problems of reproducibility in complex mind-matter systems. *Journal of Scientific Exploration,* 17, 243–270.

Bancel, P.A. (2011). Reply to May and Spottiswode's 'The Global Consciousness Project: Identifying the Source of Psi'. *Journal of Scientific Exploration,* 25, 690-694.

Bancel, P.A. (2014a). An Analysis of the Global Consciousness Project. In D. Broderick & B. Goertzel, (Eds.), *Evidence for Psi: Thirteen Empirical Research Reports* (pp. 255-277). McFarland.

Bancel, P.A. (2014b). The PEAR protocol: an experiment with meditators. Preprint.

Bell, M. (1964). Francis Bacon, pioneer in parapsychology. *International Journal of Parapsychology*, 6, 199-208.

Bem, D. (2011). Feeling the future: Experimental evidence for anomalous retroactive influences on cognition and affect. *Journal of Personality and Social Psychology*, 100, 407-425.

Berger, R.E. (1988). Psi effects without real-time feedback. *Journal of Parapsychology*, 52, 1-27.

Bierman, D.J. (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.

Bierman, D.J. (1988). A test on possible implications of the OT's for ganzfeld research. *European Journal of Parapsychology*, 7, 1–12.

Bierman, D.J. & Houtkooper, J.M. (1981). The potential observer effect or the mystery of irreproducibility, *European Journal of Parapsychology,* 3, 345-372.

Bierman, D.J. & Weiner, D.H. (1980). A preliminary study of the effect of data destruction on the influence of future observers. *Journal of Parapsychology,* 44, 233-243.

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.

Braud, L.W. & Braud, W.G. (1979). Psychokinetic effects upon a random event generator under conditions of limited feedback to volunteers and experimenter. *Journal of the Society for Psychical Research*, 50, 21-32.

Braud, W.G. & Hartgrove, J.L. (1976). Clairvoyance and psychokinesis in transcendental meditators and matched control subjects: A preliminary study. *European Journal of Parapsychology*, 1, 6-16.

Braud, W., Smith, G., Andrew, K., & Willis, S. (1976). PK influences on random generators during evocation of "analytic" versus "non-analytic" modes of information processing. *Research In Parapsychology, 1975*, 85-88. Metuchen, NJ: Scarecrow Press.

Braude, S. (1986). *The limits of influence.* New York: Routledge and Kegan Paul. (Revised edition,

University Press of America, 1997).

Broughton, R.S. & Perlstrom, J.R. (1986). PK experiments with a competitive computer game. *Journal of Parapsychology*, 50, 193-211.

Broughton, R.S. & Perlstrom, J.R. (1992). PK in a Competitive Computer Game: A Replication. *Journal of Parapsychology* 56, 292-305.

Button, K.S., Ioannidis, J.P.A., Mokrysz, C, Nosek, BA, Flint, J, Robinson, ESJ & Munafò, MR (2013). Power failure: why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, 14, 365-376.

Camstra, B. (1973). P.K. Conditioning. In W. Roll, R. Morris, J. Morris, (Eds.), *Research in Parapsychology. 1972* (pp. 25-27). Metuchen, NJ: Scarecrow press.

Davis, J.W., & Morrison, M.D. (1978). A test of the Schmidt model's prediction concerning multiple feedback in a PK test. In W.G. Roll (Ed.), *Research in Parapsychology 1977* (pp. 163-168). Metuchen, NJ: Scarecrow Press.

Debes, J., & Morris, R.L. (1982). Comparison of Striving and Non-striving Instructional Sets in a PK Study. *Journal of Parapsychology,* 46, 297-312.

Dobyns, Y. (2000). Overview of Several Theoretical Models on PEAR Data. *Journal of Scientific Exploration,* 14, 163-194.

Dunne, B.H., Dobyns, Y.H., Jahn, R.G. & Nelson, R.D. (1994). Series Position Effects in Random Event Generator Experiments. *Journal of Scientific Exploration,* 2, 197-215.

Dunne, B.J. & Jahn, R.G. (1992). Experiments in Remote Human Machine Interaction. *Journal of Scientific Exploration,* 6, 311-332.

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

Gissurarson, L.R. (1992). Studies of methods of enhancing and potentially training psychokinesis: A review. *Journal of the American Society for Psychical Research*, 86, 303-346.

Gissurarson, L.R. (1997). Methods of enhancing PK task performance. In S. Krippner (Ed.), *Advances in Parapsychological Research 8* (pp. 88-125). Jefferson, NC: Mc Farland Company.

Haraldsson, E.T. (1970). Subject Selection in a Machine Precognition Test. *Journal of Parapsychology,* 34, 182-196.

Honorton, C. & Barksdale, W. (1972). PK Performance with Waking Suggestions for Muscle Tension Versus Relaxation. *Journal of the American Society for Psychical Research,* 66, 208-214.

Honorton, C. & May, E.C. (1976). Volitional control in a psychokinetic task with auditory and visual feedback. In J.D. Morris, W.G. Roll, & R.L. Morris (Eds.), *Research in parapsychology 1975* (pp. 90-91). Metuchen, NJ: Scarecrow Press.

Honorton, C. & Tremmel, L. (1979). Psi correlates of volition: A preliminary test of Eccles' "neurophysiological hypothesis" of mind-brain interaction. *Research in Parapsychology 1976*. 36-38. Metuchen, NJ: Scarecrow Press.

Houtkooper, J.M. (1977). A study of repeated retroactive psychokinesis in relation to direct and random PK effects. *European Journal of Parapsychology*, 1, 1-20.

Houtkooper, J.M. (1993). Observational theory: A research program for paranormal phenomena. *Lisse*. The Netherlands: Swets and Zeitlinger.

Houtkooper, J.M., Andrews, R. Ganzevles, P.C.J. & van der Sijde, P.C. (1980). A hierarchical model for the observaional theories: a study of subject identity and feedback in repeaed retroactive psychokinesis. *European Journal of Parapsychology*, 3, 221-246.

Hubbard, R & Lindsay, R.M. (2008).Why P values are not a useful measure of evidence in statistical significance testing. *Theory and Psychology*, 18, 1, 69-88.

Hyman, R. (1981). Further comments on Schmidt's PK experiments. *Skeptical Inquirer*, 5, 34-40.

Jahn, R.G., Dunne, B. J.,Nelson, R.G., Dobyns, Y.H., & 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, 345-367.

Jahn, R.G., Mischo, J., Vaitl, D., Dunne, B.J., Bradish, G.J., Dobyns, Y.H., et al. (2000). Mind/Machine interaction consortium: PortREG replication experiments. *Journal of Scientific Exploration*, 14, 499-555.

Kennedy, J.E. (2003). The capricious, actively evasive, unsustainable nature of psi. A summary and hypotheses. *Journal of Parapsychology*, 67, 53-74.

Kennedy, J.E. & Taddonio, J.L. (1976). Experimenter effects in parapsychological research. *Journal of Parapsychology*, 40, 1-33.

Levi, A. (1979). The influence of imagery and feedback on PK effects. *Journal of Parapsychology,* 43, 275-289.

Lucadou, W.v. (1995). The model of pragmatic information (MPI). *European Journal of Parapsychology*, 11, 58-75.

Lucadou, W.v.(2001). Hans in Luck: The Currency of evidence in parapsychology. *Journal of Parapsychology*, 65, 3-16.

Lucadou, W.v. (2006). Self-Organization of temporal structures-A possible solution for the intervention problem. In D.P. Sheehan (Ed.), *Frontiers of Time. Retrocausation - Experiment and Theory* (pp. 293-315). Melville, NY: American Institute of Parapsychology.

Lucadou, W.v., Römer, H., & Walach, H. (2007). Synchronistic Phenomena as Entanglement Correlations in Generalized Quantum Theory. *Journal of Consciousness Studies*, 14, 50-74.

Matas, F., & Pantas, L. (1971). A PK experiment comparing meditating vs. nonmeditating subjects. In T*he Parapsychological Association 14th Annual Convention*: *Proceedings of presented papers*, 8, 12-13.

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

May, E.C., Utts, J.M., & Spottiswoode, S.J.P. (1995b). Decision Augmentation Theory: Applications to the Random Number Generator Database. *Journal of Scientific Exploration*, 9, 453-488.

Millar, B. & Broughton, R. S. (1977). An investigation of the psi enhancement paradigm of Schmidt. . In J.D.Morris, W.G. Roll, & R.L. Morris (Eds.), *Research in parapsychology 1976* (pp. 23-25). Metuchen, NJ: Scarecrow Press.

Millar, B (2012). Towards a Forensic Parapsychology in the OT Paradigm. *Journal of Non-locality*, 1, 1.

Morris, R., Nanko, M., & Phillips, D. (1978). Intentional observer influence upon measurements of a quantum mechanical system: A comparison of two imagery strategies. In T*he Parapsychological Association 21st Annual Convention: Proceedings of Presented Papers,* 266-275. Durham, NC: Parapsychological Association.

Morris, R.L., Nanko, M., & Phillips, D. (1982). A Comparison of Two Popularly Advocated Visual Imagery Strategies in a Psychokinesis Task. *Journal of Parapsychology,* 46, 1-16.

Morrison, M.D., & Davis, J. W. (1979). PK with immediate, delayed, and multiple feedback: A test of the Schmidt model's predictions. In *The Parapsychological Association 21st Annual Convention: Proceedings of Presented Papers*, 97-117. Durham, NC: Parapsychological Association.

Nanko, M. (1981). Use of goal-oriented imagery strategy on a psychokinetic task with "selected" subjects. *Journal of the Southern California Society for Psychical Research*, 2, 1-5.

Palmer, J (1986). An evaluation of the current status of parapsychology. *Army Research Institute Research note* 86-63.

Parker, A. (2013). Is Parapsychology's Secret, Best Kept a Secret? Responding to the Millar Challenge. *Journal of Nonlocality*, 2.

Peoc'h R. (1988). Chicken imprinting and the Tychoscope : ANPSI Experiment, *Journal of the Society for Psychical Research*, 55, 1-9.

Peoc'h R. (2001). Chicks' distant psychokinesis (23 kilometers). *Revue Francaise de Psychotronique*, 2.

Radin, D. (2006) Experiments Testing Models of Mind-Matter Interaction *Journal of Scientific Exploration,* 20, 375-401.

Radin, D.I., May, E.C., & Thomson, M.J. (1985). Psi experiments with random number generators: Meta-analysis Part 1. Unpublished manuscript.

Radin, D., Michel, L., Galdamez, K., Wendland, P., Rickenbach, R., & Delorme, A. (2012). Consciousness and the double-slit interference pattern: Six experiments. *Physics Essays*, 25, 157-171.

Radin, D., Michel, L., Johnston, J. and Delorme, A., (2013). Psychophysical interactions with a double-slit interference pattern. *Physics Essays*, 26, 553-566.

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

Radin, D.I., & Nelson, R.D. (2003). A meta-analysis of mind-matter interaction experiments from 1959 to 2000. In W.B. Jonas, & C.C. Crawford (Eds.), *Healing, Intention and Energy Medicine: Research Methods and Clinical Applications*, 39- 48. Edinburgh, UK: Churchill Livingstone.

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

Radin, D., Nelson, R., Dobyns, Y., Houtkooper, J. (2006b). Assessing the evidence for mind-matter interaction effects. *Journal of Scientific Exploration,* 20, 361–374.

Rhine, J.B., & Humphrey, B.M. (1944). The PK effect: Special evidence from hit patterns. I. Quarter distribution of the page. *Journal of Parapsychology*, 8, 18-60.

Rush, J.H. (1986). Findings from experimental PK research. In H. Edge, R. Morris, J. Palmer, & J. Rush (Eds.), *Foundations of parapsychology* (pp. 237-275). Boston: Routledge and Kegan Paul.

Schmeidler, G. (1984). The basic problem, research methods, and findings. In S.Krippner (Ed.), *Advances in Parapsychological Research 4* (pp. 12-30). Jefferson, NC: McFarland Publishing.

Schmeidler, G. (1987). Psychokinesis: Recent studies and a possible paradigm shift. In S.Krippner (Ed.) ,*Advances in Parapsychological Research 5* (pp. 9-38). Jefferson, NC: McFarland Publishing.

Schmeidler, G. (1990). PK: Recent publications and an evaluation of the quantitative research. In S.Krippner (Ed.), *Advances in Parapsychological Research 6*. (pp. 13-53) Jefferson, NC: McFarland Publishing.

Schmeidler, G. (1994). PK: Recent research reports and a comparison with ESP. In S.Krippner (Ed.), *Advances in Parapsychological Research 7* (pp. 198-237). Jefferson, NC: McFarland Publishing.

Schmidt, H. (1969a). Precognition of a Quantum Process. *Journal of Parapsychology*, 33, 99-108.

Schmidt, H. (1969b). Clairvoyance Tests with a Machine. *Journal of Parapsychology*, 33, 300-306.

Schmidt, H. (1970a). A PK Test with Electronic Equipment. *Journal of Parapsychology*, 34, 175-181.

Schmidt, H. (1970b). PK Experiments with Animals as Subjects. *Journal of Parapsychology*, 34, 225-261.

Schmidt, H. (1971). Mental Influence on Random Events. *New Scientist and Science Journal*, June 24, 757-768.

Schmidt, H. (1973). PK Tests with a High-Speed Random Number Generator. *Journal of Parapsychology*, 37, 105-118.

Schmidt, H. (1974). Comparison of PK Action on Two Different Random Number Generators. *Journal of Parapsychology*, 38, 47-55.

Schmidt, H. (1974b). Psychokinesis. In Mitchell, E. D. (Ed.), Psychic exploration: A challenge for science. New York: G. P. Putnam's Sons.

Schmidt, H. (1976). PK Effect on Pre-Recorded Targets. *Journal of the American Society for Psychical Research,* 70, 267-291.

Schmidt, H. (1978). Can an Effect Precede Its Cause? A Model of a Non-causal World. *Foundations of Physics* 8, 463-480.

Schmidt, H. (1979a). Use of Stroboscopic light as rewarding feedback in a PK test with pre-recorded and momentarily generated random events. *Research in Parapsychology* 1978, 115-117. Metuchen, N.J.: Scarecrow Press.

Schmidt, H. (1979b). Search for Psi Fluctuations in a PK Test with Cockroaches. *Research in Parapsychology* 1978, 77-78. Metuchen, N.J.: Scarecrow Press.

Schmidt, H. (1982). Collapse of the State Vector and PK Effect. *Foundation of Physics* 12, 565-581.

Schmidt, H. (1984). Comparison of a Teleological Model with a Quantum Collapse Model of Psi. *Journal of Parapsychology,* 48, 261-276.

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

Schmidt, H. & Braud, W. (1993). New PK Tests with an Independent Observer. *Journal of Parapsychology* 57, 227-239.

Schmidt, H. & Dalton, K. (1999). PK tests with repeated efforts on prerecorded targets. *Journal of Parapsychology* 63, 275-283.

Schmidt, H., Morris, R. & Hardin, C.L. (1986). Channeling Evidence for a Psychokinetic Effect to Independent Observers. *Mind Science Foundation report*, San Antonio, TX: Mind Science Foundation.

Schmidt, H., Morris, R. & Rudolph, L. (1986). Channeling Evidence for a Psychokinetic Effect to Independent Observers. *Journal of Parapsychology,* 50, 1-15.

Schmidt, H. & Pantas, L. (1972). Psi Tests with Internally Different Machines. *Journal of Parapsychology*, 36, 222-232.

Schmidt, H. & Schlitz, M. (1989). A large-scale pilot PK experiment with pre-recorded random events. In L. Henkel & R. Berger (Eds.), *Research in Parapsychology 1988* (pp. 6-10). Metuchen, NJ: Scarecrow Press.

Schmidt, H. & Stapp, H. (1993). Study of psychokinesis with pre-recorded random events and the effect of pre-observation. *Journal of Parapsychology* 57, 331-348.

Schouten, S.A. (1977). Testing some implications of a PK observational theory. *European Journal of Parapsychology*, 1, 21-31.

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

Shoup, R. & Etter, T. (2002). Can Causal Influence Propagate Backwards in Time? - A Simple Experiment in Markov Chains and Causality. Boudary Institute report. retrieved from http://www.gotpsi.org .

Stanford, R.G. (1977). Experimental psychokinesis: A review from diverse perspectives. In B. Wolman (Ed.), *Handbook of parapsychology* (pp. 324-381). New York: Van Nostrand Reinhold.

Stanford, R.G., & Kottoor, T.M. (1985). Disruption of attention and PK-task performance. In *The Parapsychological Association 28th Annual Convention: Proceedings of Presented Papers* (pp. 117-132). Durham, NC.: Parapsychological Association.

Stapp, H.P. (1994). Theoretical model for a purported empirical violation of quantum theory. *Physical Review A*, 50, 18-22.

Steilberg, B.J. (1975). Conscious concntration versus visualization in PK tests. *Journal of Parapsychology*, 39, 12-20.

Talbert, R. & Debes, J. (1981). Time-displacement psychokinetic effects on a random number generator using varying amounts of feedback. In *The Parapsychological Association 24th Annual Convention: Proceedings of Presented Papers,* 58-61. Durham, NC: Parapsychological Association.

Terry, J. & Schmidt, H. (1978). Conscious and subconscious PK tests with pre-recorded targets. In Roll, W. (Ed.), *Research in Parapsychology 1997* (pp. 36-40). Metuchen NJ: Scarecrow Press.

Tressoldi, P., Pederzoli, L. Caini, P., Ferrini, A., Melloni, S., Richeldi, D., Richeldi, F. Duma, G.M. (2014). Mind-Matter Interaction at a Distance of 190 km: Effects on a Random Event Generator Using a Cutoff Method. *NeuroQuantology*, 12.

Van de Castle, R. (1958). An exploratory study of some personality correlates associated with PK performance. *Journal of the American Society for Psychical Research*, 52, 134-150.

Varvoglis, M.P. (1988). A "psychic contest" using a computer-RNG task in a non-laboratory setting. In *The Parapsychological Association 31st Annual Convention: Proceedings of Presented Papers,* (pp. 36-52). Durham, NC: Parapsychological Association.

Varvoglis, M.P., & 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.

Walker, E.H., (1984). A review of the Quantum Mechanical Theory of Psi Phenomena. *Journal of Parapsychology*, 48, 4, 277-332.

Weiner, D.H., & Zingrone, N.L. (1986). The checker effect revisited. *Journal of Parapsychology*, 50, 85-121.

Weiner, D.H., & Zingrone, N.L. (1989). In the eye of the beholder: Further research on the “checker effect.” *Journal of Parapsychology*, 53, 203–231.

Winnett, R. & Honorton, C. (1977). Effects of meditation and feedback on psychokinetic performances: Results with practitioners of Ajapa Yoga. In J.D. Morris, W.G. Roll, & R.L. Morris (Eds.), *Research in parapsychology 1976* (pp. 97-98). Metuchen, NJ: Scarecrow Press.

Endnotes

[1] Quoted from Bosch, Steinkamp & Boller (2003)

[2] See also Gissuarson (1997), Rush (1986) and Schmeidler (1984, 1987, 1990).

[3] The reproducibility is also problem in a other fields. Button et al., 2013; Begley and Ellis, 2012.

[4] More complete expositions are in Braude, 1979; Palmer, 1986.

[5] The Z-score is the number of standard deviations from chance. It provides a more compact notation than the familiar P-value. P-values of 0.05 and 0.01 correspond to Z's of 1.65 and 2.32. Z's above 3 are highly significant. In practice Z's as low as 1.65 are considered sufficient to reject a null hypothesis, although this is increasingly viewed as inadequate evidence for that purpose (for a readable entry to the P-value debates see Hubbard and Lindsay, 2008).

[6] The cumulative (Stouffer) Z-score is the sum of scores divided by the square root of the number of studies.

[7] Inspired by quantum mechanics, the OTs claimed to establish a direct connection between micro-PK and certain fundamental models of physics, while at the same time setting limits to the manner and the degree to which PK can affect physical processes.

[8] The stopping was a statistically valid procedure because the total number of trials was set in advance.

[9] PEAR, as well as some others, preferred random event generator (REG) to RNG. For consistency, we will be using the term RNG throughout. Similarly, we use 'subjects' or 'participants', as opposed to the PEAR-specific 'operators'.

[10] Bit generation rate was 1000/sec, but for each second only the sum of 200 random bits was recorded. The trials were binomial random variables with N=200, p=1/2 and a standard deviation of about 7.

[11] Other experiments have explored remote micro-PK (Stevens,1999; Tressoldi et al, 2014; Dickstein & Davis, 1979; Radin, 2013) and have generally yielded statistically significant results. None of them have found a relationship between distance and effect size.

[12] The techniques were the Stouffer Z, and fixed and random effects models.

[13] In fact, the first filedrawer calculation by RN included these studies, which is why the filedrawer they found exceeded 50,000.

[14] We would expect 2 studies to have a Z-score of 2 or more in the database; instead, we note that there were 40.

[15] An earlier calculation found agreement with DAT. This incorrect conclusion was due to analysis errors (Bancel, 2011).

[16] Even Rhine, considered the father of the universalist approach, tended to focus on those few individuals who showed real promise.

[17] It is worth recalling here the proportion of outliers in the 12 year PEAR benchmark series: 2 out of 91.

[18] Some promising research has been reported (Van de Castle, 1958; Radin et al, 2012; Schmidt & Schlitz,1989). However, a series of studies testing trait variables yielded little evidence for any systematic relationship with micro-PK (Gissuarson 1992).

[19] Helmut Schmidt, Dean Radin, and Charles Honorton are good examples.

[20] His early experiments gave positive results but replications in 1979, when Schmidt had become concerned about experimenter psi, were null.

[21] A RNG drove a robot's random displacements. Chicks had been conditioned to perceive the robot as a maternal substitute and when placed before the chick's transparent cage, the robot displacements toward the chicks were highly significant. Similar results obtained when the RNG was situated 23 kilometers away.

[22] See Rex Stanford's (1977) conformance behavior model

[23] Supportive evidence for analyzer or checker effects have been reported for decades; Weiner & Zingrone (1986, 1989) have conducted particularly interesting research in this area

[24] Roger Nelson's chapter reviews the field-REG studies and the GCP project in detail.

[25] The most recent analyses indicate that the GCP result may well be an experimenter effect (Bancel, 2014a).

[26] Preliminary evidence for an effective protocol comes from a recent experiment by Bem (2011), which is conceptually identical to binary push-button experiments (Schmidt, 1974) and was replicated by 7 independent laboratories (Bem et al., preprint, 2014). When these studies with unselected subjects are combined a Z-score of 4 obtains.