

MICRO-PSYCHOKINESIS: EXCEPTIONAL OR UNIVERSAL?

By Mario Varvoglis and Peter A. Bancel

ABSTRACT: Most psychokinesis studies fall within either an elitist research tradition, involving exceptional participants and focusing on directly perceptible, macro-PK effects; or a universalist approach, exploring subtle micro-PK effects through massive data collection from unselected participants. However, Helmut Schmidt's highly significant body of micro-PK research was mostly elitist, involving intensive work with small numbers of selected individuals. We contrast his approach to that of the PEAR laboratory, with nearly 100 unselected participants and a highly standardized protocol. Although PEAR's 12-year benchmark study did produce significant cumulative results, a carefully designed replication, involving three laboratories, was nonsignificant. We argue that this apparent failure to replicate was due to the erroneous assumption that the original PEAR data were homogeneous across participants, when in fact they were dominated by two extreme outliers who contributed nearly a quarter of the total data. By ignoring this, the replication employed an overestimation of the effect size for the original study and underestimated the power needed to replicate. We conclude that research generally supports the view that micro-PK is not widely distributed, but exceptional, and that it is unproductive to attempt to tease extremely weak effects out of unselected volunteers. Like Schmidt, investigators should focus on optimizing testing conditions and work intensely with selected participants.

Keywords: micro-PK, RNG research, Schmidt, PEAR, experimenter effects

The term psychokinesis (PK) covers a wide spectrum of putative psi phenomena, ranging from metal bending, poltergeists, and table levitations, to distant influence on human physiology, animal activity, or plant growth. Psychokinesis also refers to situations where statistical deviations from chance in probabilistic systems, such as tumbling dice or coin-tosses, are observed to correlate with participants' wish or intention for a particular outcome. PK studies on probabilistic systems were carried out in various laboratories during the 1940s and 1950s, typically employing mechanical devices with tumbling dice or coin tosses. In the early 1970s hardware random number generators (RNGs) began to replace mechanical systems and quickly became widespread in parapsychology, first as self-standing units and later in computer controlled configurations. Compared to the earlier mechanical systems, RNGs greatly simplified the testing process and thus attracted a wide range of researchers who began exploring a spectrum of hypotheses and variables potentially associated with PK.

Effects reported in association with RNGs are generally qualified as micro-psychokinetic, suggesting a distinction between large-scale or directly perceptible PK and extremely subtle effects that can be inferred only through statistical methods. It is unclear whether the term should apply to all probabilistic systems, even when clearly macroscopic; nor is it certain whether the distinction between macro- and micro-PK points to fundamental differences or merely operational ones. Nevertheless, the fact is that this distinction tends to introduce a split in research strategies—in particular, with respect to the presumed agent or source of the effect. Historically, researchers interested in macro-PK have largely employed an elitist approach, seeking out either exceptional individuals who might produce PK in the laboratory, or exceptional circumstances, as in the case of field research on poltergeists. By contrast, insofar as it seems plausible that weaker PK effects may be more widespread or smoothly distributed in the general population, micro-PK research seems far more amenable to a universalist approach, employing massive data collection from a large number of unselected participants.

The issue we would like to examine here is whether the experimental literature indeed points to the idea that a weak form of micro-PK skill is distributed throughout the population at a detectable level, thus justifying a universalist research strategy, or whether it suggests instead a highly acentric distribution,

with micro-PK being a strong but rare or exceptional skill. This latter conclusion would encourage a return to highly selective, elitist strategies, rather reminiscent of what has been practiced in macro-PK research.

One approach for evaluating this question would be to conduct a meta-analysis comparing effect sizes in studies leaning toward elitist strategies with those that seem to adopt more universalist methods. The problem is that the two main meta-analyses of the RNG micro-PK literature (Bösch, Steinkamp, & Boller, 2006; Radin & Nelson, 1989) arrive at strikingly divergent conclusions, with the most recent one asserting that the small meta-analytic effect size found across studies can be explained by publication bias. As we have argued elsewhere (Varvoglīs & Bancel, 2015), although we agree that there is evidence for substantial publication bias, a subset of the micro-PK literature, involving a large number of high-*z* studies, renders implausible arguments that attribute the effect solely to publication bias and other methodological problems. The literature in fact is highly heterogeneous and meta-analyses cannot currently provide reliable estimates of micro-PK effect sizes.

As an alternative to a broadly-based meta-analytic approach, we focus here upon two major contributors to the experimental micro-PK literature that adopted opposite approaches on the elitist-universalist spectrum. The universalist approach is expressed in the PEAR lab's 12-year "benchmark" micro-PK study involving a highly standardized protocol and nearly 100 unselected participants. The elitist approach manifests quite clearly in the equally impressive and long-term body of research by Helmut Schmidt, who adopted a far more selective and personalized approach in his work with participants. Together these two bodies of research constitute a substantial portion of the micro-PK literature and thus afford a good approximation to the issue examined: Does experimental evidence for micro-PK reflect the minute contributions of the participant population as a whole, or is it rather the result of a few gifted individuals?

Schmidt's Research

Helmut Schmidt is rightfully considered the "father" of micro-PK RNG research: He was the first to introduce a practical hardware RNG for psi studies, was a highly prolific investigator over the course of three decades, played a major role in conceptualizing and modeling the phenomena, and produced by far the strongest and most consistent results in the field. Though a number of other researchers have had considerable success with RNG-PK studies, Schmidt's contribution was clearly exceptional. In our review of his work we found 22 experimental publications containing 50 independent studies, three quarters of which were significant ($p < .05$) and nearly half of which had *z*s above 3 (Varvoglīs & Bancel, 2015). Even if we admit some ambiguity in determining the number of independent studies in experiments that used different devices or participant groups, any tally of the combined significance of Schmidt's work leads to astronomical odds against the null hypothesis.

Why was Schmidt so phenomenally successful with RNG studies? To begin with, a close reading of his reports reveals a highly intuitive approach to the psychological facets of micro-PK research and a keen sense of how best to work with individuals. Indeed, his stance with regard to testing for micro-PK was straightforward and practical: Psi is neither egalitarian nor available on demand, and experiments should be run to proactively track it down and encourage its emergence. Among the strategies employed, foremost was the selection 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 participant pools, selection was frequently based on systematic preliminary tests. Schmidt was capable of investing months of his time preparing 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 adopting this 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 people's homes and make himself available on short notice should volunteers find themselves 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 a preferred feedback mode. In some instances, a session would only be initiated after a preliminary "warm-up" test was successful, and volunteers were generally encouraged to set their own pace and take breaks or chat with an experimenter if they felt tired or bored. Schmidt indeed allowed for variable contributions from individual participants, for the interruption of sessions, and even for participants to be dropped from an experiment if performance lagged. This stopping was a methodologically sound procedure since Schmidt set the total number of trials (as opposed to the total number of participants) in advance (Schmidt, 1973).

Thus, besides his general tendency to select promising participants, a second explanation of his results is that he was simply a very good experimenter. Whether tacitly or explicitly, Schmidt understood the psychology of getting results, which he applied through 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). 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.

This brings us to a third way in which his research was elitist. Besides selecting for talented participants and seeking to create psi-conducive testing conditions for them, Schmidt was a highly gifted micro-PK subject himself. His basic interest was 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 he occasionally used himself as a subject, having discovered that he was often as reliable in obtaining positive results as his other subjects. But of course, if this was the case, there is little reason to suppose that his psi skills emerged only when he intentionally evoked them. Parapsychologists (including Schmidt himself) have suggested different channels through which experimenter psi may manifest in micro-PK experiments: direct (albeit unintentional) action on the RNGs during testing, retroactive effects during data analysis (Weiner & Zingrone, 1986, 1989), or numerous intuitive decisions that tacitly guide the experimenter's sampling of the RNG (May, Utts, & Spottiswoode, 1995). Whatever the potential channel used, it seems likely that Schmidt's striking success as experimenter was partly related to his talent as psi subject—a fact that may considerably challenge the generalizability of his results, as well as their replicability across laboratories.

PEAR and the Consortium

The Princeton Engineering Anomalies Research (PEAR) laboratory, founded in 1979 by Robert Jahn, Dean of the engineering school at Princeton University, assembled a staff of physicists, psychologists, and technicians who worked in a basement lab in the campus engineering building for nearly 30 years. Although PEAR explored numerous PK target systems, including macroscopic probabilistic systems such as the random mechanical cascade (Dunne, Nelson, & Jahn, 1988), their primary focus was on RNG research. In stark contrast to Schmidt's strongly personal approach, the laboratory was committed to a strict universalist approach, involving volunteers whose participation depended essentially on their own availability and willingness, and a patient, progressive accumulation of data using the same protocol over many years.

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 participant's intention: Bias the output to go high, to go low, or to remain even. The experimental hypothesis was 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 between the two directional runs. We focus here on the benchmark experiment, which was a 12-year study that collected over 2.5 million experimental trials from 91 participants, equally distributed across HI, LO, and BL conditions. At its termination, the experiment

had attained high significance, yielding a z of 3.8 (Jahn, Dunne, Nelson, Dobyns, & Bradish, 1997). The result is particularly noteworthy, insofar as PEAR had a firm policy of publishing all its experimental results in either refereed journals or publicly accessible internal reports, thus ensuring that the huge volume of research produced over the years can be considered free of publication bias and file-drawer problems.

In 1996 PEAR 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 undertook an extensive replication of the PEAR benchmark experiment, using the same protocol, uniform RNGs among the groups, and stipulation of an equal contribution of trials from each laboratory. The replication collected roughly the same amount of data as the original benchmark experiment, with a total of 750,000 trials per condition (HI, LO, and BL) from 227 volunteer participants. The primary hypothesis was retained, with a prediction for RNG deviations consistent with volunteers' intentions, and the difference between HI and LO scores as the test statistic. Given effect size estimations from the PEAR experiment, this gave the replication about an 85% chance of succeeding at $p < .01$.

Following three years of intensive data collection, and a period of careful analysis, the Consortium published its much-awaited report (Jahn et al., 2000), but the results were disappointing. All three research groups found positive deviations, but the effect size was nearly an order of magnitude smaller than expected and the overall z came in at a nonsignificant 0.6. The combined PEAR and consortium results were still significant, with a z of 3.2, yet the apparent failure to replicate a solid and well-founded prediction, despite a well-planned collaborative study that included PEAR staff, remained quite surprising.

Why would a well-powered, rigorous replication fail to reproduce the previous results? Was there an essential difference between the experiments that had been overlooked? The Consortium examined a number of possibilities, but found none that was compelling, other than a suggestion that psi may be intrinsically elusive (see Atmanspacher & Jahn, 2003). However, we believe that the explanation is far simpler: The replication attempt underestimated the power needed to obtain results comparable to the PEAR benchmark study.

The replication's validity depended on two assumptions. The first assumption is that the original participant pool was representative of the general population, and the second is that the benchmark effect size is a valid estimate of an average participant's PK effect size. Under these assumptions, any new group of participants should provide about the same effect size, and the appropriate size for a replication can be determined by power analysis.

However, a close look at the PEAR benchmark study shows that there were two extreme outliers in the participant pool, with highly significant personal databases yielding z s of 5.6 and 3.4. The outliers had large individual effect sizes, and they contributed nearly a quarter of the data, with individual contributions far exceeding those of any other participants. This resulted in their contributing over 80% of the total HI-LO deviation. It is easy to see that they are not representative of the 89 other participants, as the overall z of the remaining three quarters of the database is only 0.8, whereas the combined z for the two outliers is 6.5. This difference is significant at the 5-sigma level. Indeed, if we exclude these two outliers, and focus on the database of the 89 remaining participants, we obtain nearly the same effect size and z score as in the nonsignificant consortium replication (see Appendix).

In short, we suggest that the consortium's apparent failure to replicate was due to an overestimation of the true population effect size due an inclusion of the outlier participants, and a consequent underestimation of the power needed to replicate. Had the replication design been based on an effect size without the outliers, then the power needed to replicate would have had to be nearly quadrupled. This means that the apparent replication failure does not call into question the original evidence seen in the benchmark PEAR databases, but only the assumption of homogeneity of its effects across participants.

Optimizing Micro-PK Research

What, then, is there to make of the universalist claim that positive results with unselected participants should be a straightforward matter given sufficient data? Insofar as the PEAR and consortium

results, with outliers removed, did produce some indication of a small effect, and that they are clearly free from any file drawer problem, the cumulative PEAR/Consortium results might justify pursuing an approach based on unselected participants, massive data collection, and analytical tools to tease out effects in the data. However, our analysis shows that studies would need to be significantly larger than those of the benchmark and Consortium experiments merely to provide statistical evidence for an effect. Given the enormous resource investment this approach has represented, involving several laboratories over the course of several years in the case of the Consortium replication, the returns obtained seem meager indeed—and the universalist strategy far from optimal.

We emphasize that the original benchmark result was entirely dominated by a disproportionate contribution from just 2% of the participant population. This basic observation challenges the idea that “anybody can do it”—at least from a pragmatic viewpoint—and points to the benefits of participant selection. From this perspective, the PEAR/Consortium studies, in which we can trace an overall significant effect to the large contributions of two exceptional participants, essentially validate the wisdom of Schmidt’s approach, which was to work intensively with a few participants rather than teasing extremely weak effects out of unselected volunteers.

It should be emphasized that these conclusions are, for now, limited to micro-PK research; they do not necessarily carry over to other parapsychology paradigms. Unselected participants may well perceive ganzfeld, presentiment, or DMILS protocols as more relevant and motivating than micro-PK protocols, and therefore produce far better results. Also, even if the universalist approach is unsatisfactory for proof- or process-oriented hypothesis testing in micro-PK research, it can still be useful for participant selection. Following the lead of Tart (1976), who found that preselection for high scorers in ESP tests seemed to pay off in terms of subsequent ESP training, it may be worth undertaking large-scale (e.g., internet-based) testing to locate promising individuals, and then progressively focus on the few who show the highest potential.

Of course, participant selection alone is hardly sufficient. The testing conditions, the meaningfulness of the task for the participant, and most importantly the investigator-participant relationship, have repeatedly been acknowledged as critical, even with the most gifted macro-PK participants. Why should this be any different with micro-PK? Taking our cue from Schmidt, we suggest working with participants in a highly personalized manner, with a strong focus on motivational conditions and a readiness to adapt testing conditions to the participant (rather than rigidly imposing compliance to a predefined protocol). In this context, it is worth recognizing a rather substantial body of process-oriented research, by a broad spectrum of investigators, exploring factors that enhance micro-PK performance—somewhat in the way that “noise reduction” procedures seem to enhance ESP performance. This research has been amply documented elsewhere (see Gissurarsson, 1997; Varvoglis & Bancel, 2015); some of the more promising optimization factors include a passive-volition set, goal-oriented visualization techniques, and meditation practice.

Any discussion of laboratory psi research is incomplete without addressing the issue of experimenter psi, as has been discussed by a number of authors (Kennedy & Taddonio, 1976; Palmer & Millar, 2015; Parker, 2013). Some argue strongly that most psi researchers who obtain consistent results—whether for micro-PK or other experimental paradigms—are themselves good psi participants. Kennedy and Taddonio (1976) remark:

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)

This is not to suggest that experimenters are the only source of psi in the lab; it seems reasonable to assume that micro-PK effects are associated with strong performers—be they participants or experimenters—in conjunction with favorable testing conditions. But to the extent that experimental results potentially reflect the PK input of investigators as “hidden subjects,” we are confronted with an inherent ambiguity in how to interpret results. How do we distinguish participant effects, assumed to be representative of a

larger population and lawful phenomena, from effects that may be due to the experimenters themselves, and potentially dependent upon the very hypotheses they pose? If this issue cannot be resolved, parapsychology may need to reconsider the classical experimental paradigm altogether and turn to radically different epistemological approaches (Atmanspacher & Jahn, 2003; Lucadou, 2001).

The complex issue of experimenter psi notwithstanding, we should not lose sight of the role that experimenter skill may play in obtaining results. It may be that successful investigators such as Schmidt simply know how to facilitate participants' talents. If so, we need to understand in a far more detailed way just how they do it. The number of researchers who systematically succeed in psi research is limited and as Parker (2013) has pointed out, an important body of tacit knowledge risks being lost. Perhaps, in addition to mastery of all the analytical tools that go with the territory, upcoming parapsychologists should train or be mentored by psi-conducive experimenters, studying and modeling their state of mind, mental set, expectations, rituals, and so forth, so they can ensure the longevity of their subtle craft.

In summary, rather than assuming "anybody can do it," we recommend that micro-PK, like macro-PK, be approached as a rare event, one that emerges under exceptional circumstances or as a result of exceptional ability. From this perspective, its investigation demands that experimenters have a special skill set, a process for participant selection, flexible protocols that can adapt to participants' state, mood, or performance, and proactive optimization procedures that may enhance participant scoring.

References

- Atmanspacher, H., & Jahn, R. G. (2003). Problems of reproducibility in complex mind-matter systems. *Journal of Scientific Exploration*, *17*, 243–270.
- 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.
- Dunne, B., Nelson, R. D., & Jahn, R. G. (1988). Operator-related anomalies in a random mechanical cascade. *Journal of Scientific Exploration*, *2*, 155–179.
- Gissurarsson, L. R. (1997). Methods of enhancing PK task performance. In S. Krippner (Ed.), *Advances in parapsychological research 8* (pp. 88–125). Jefferson, NC: McFarland.
- 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., Dunne, B., Bradish, G., Dobyns, Y., Lettieri, A., Nelson, R., . . . Walter, B. (2000). Mind/machine interaction consortium: PortREG replication experiments. *Journal of Scientific Exploration*, *14*, 499–555.
- Kennedy, J. E., & Taddonio, J. L. (1976). Experimenter effects in parapsychological research. *Journal of Parapsychology*, *40*, 1–33.
- Lucadou, W. v. (2001). Hans in luck: The currency of evidence in parapsychology. *Journal of Parapsychology*, *65*, 3–16.
- May, E. C., Utts, J. M., & Spottiswoode, S. J. P. (1995). Decision augmentation theory: Applications to the random number generator database. *Journal of Scientific Exploration*, *9*, 453–488.
- Palmer, J., & Millar, B. (2015). Experimenter effects in parapsychology research. In E. Cardeña, J. Palmer, & D. Marcusson-Clavertz (Eds.), *Parapsychology: A handbook for the 21st century* (pp. 293–300). Jefferson, NC: McFarland.
- Parker, A. (2013). Is parapsychology's secret, best kept a secret? Responding to the Millar challenge. *Journal of Nonlocality*, *2*. Retrieved from <http://journals.sfu.ca/jnonlocality/index.php/jnonlocality/article/download/28/22>
- Radin, D., & Nelson, R. (1989). Consciousness-related effects in random physical systems. *Foundations in Physics*, *19*, 1499–1514.
- Schmidt, H. (1973). PK tests with a high-speed random number generator. *Journal of Parapsychology*, *37*, 105–118.
- Schmidt, H. (1987). The strange properties of psychokinesis. *Journal of Scientific Exploration*, *1*, 103–118.
- Tart, C. T. (1976). *Learning to use extrasensory perception*. Chicago, IL: University of Chicago Press.
- Varvoglis, M. P., & Bancel, P. (2015). Micro-psychokinesis. In E. Cardeña, J. Palmer, & D. Marcusson-Clavertz (Eds.), *Parapsychology: A handbook for the 21st century* (pp. 266–281). Jefferson, NC: McFarland.
- 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.

Institut Métapsychique International
 51 rue de l'Aqueduc
 75010 Paris, France
 contact@mario-varvoglis.com

Appendix

Table 1 lists the relevant data from the PEAR benchmark experiment, showing the N , effect sizes, and z statistics for the full database, the outlier participant-operators (operator IDs: Op10 and Op78) and the outlier-removed 89-operator subset. Data for operators 10 and 78 can be read from Figure 4 in Jahn et al., 1997.

Table 1
Relative Effect Sizes of the Outlier Participants of the PEAR Benchmark Experiment

	N	μ	σ	z	Δ
All data	837,000	0.042	0.011	3.81	34,850
Operator 10	120,000	0.162	0.029	5.60	19,400
Operator 78	67,000	0.132	0.039	3.42	8,850
Ops 10 & 78	187,000	0.151	0.023	6.54	28,250
89 Operators	650,000	0.010	0.012	0.82	6,600

Note. N = number of trials per intention; μ = effect size of the mean HI-LO deviation; σ = theoretical standard deviation of μ ; z = z statistic for μ ; Δ = absolute HI-LO mean shift; $\sigma = \sqrt{100/N}$ = standard deviation of the HI-LO mean shift.

Abstracts in Other Languages

Spanish

LA MICRO-PSICOKINESIS: EXCEPCIONAL O UNIVERSAL?

RESUMEN: La mayoría de los estudios sobre psicoquinesis caen ya sea dentro de una tradición de investigación elitista, con participantes excepcionales y centrada en efectos directamente perceptibles, o sea macro-PK, o un enfoque universalista que explora efectos de micro-PK sutiles a través de la recolección masiva de datos con participantes no selectos. Sin embargo, la mayoría del trabajo muy significativo de Helmut Schmidt en la investigación de micro-PK fue en su mayoría elitista, trabajando intensamente con un pequeño número de individuos selectos. Contrastamos su acercamiento con el del laboratorio PEAR, con cerca de 100 participantes no selectos y un protocolo muy estandarizado. Aunque el estudio básico de 12 años de PEAR produjo resultados acumulados significativos, un estudio de replicación cuidadosamente diseñado con la participación de tres laboratorios no fue significativo. Argumentamos que este fracaso evidente en replicar se debió a la suposición errónea de que los datos PEAR originales eran homogéneos para todos los participantes, cuando en realidad estaban dominados por dos participantes con valores extremos que contribuyeron casi una cuarta parte del total de datos. Al hacer caso omiso de esto, la replicación sobreestimó el tamaño del efecto del estudio original y subestimó el poder necesario para replicar. Concluimos que la investigación en general apoya la opinión de que la habilidad micro-PK no está distribuida global sino excepcionalmente, y que es improductivo tratar de obtener efectos extremadamente pequeños con voluntarios no selectos. Al igual que Schmidt, los investigadores deben centrarse en la optimización de las condiciones de prueba y en trabajar intensamente con participantes selectos.

French

MICRO-PSYCHOCINÈSE : EXCEPTIONNELLE OU UNIVERSELLE ?

RESUMÉ : La plupart des études de la psychocinèse entrent dans une tradition de recherche élitiste, impliquant des participants exceptionnels et se focalisant sur des effets de macro-PK directement perceptibles ; ou une approche universaliste explorant des effets subtils de micro-PK à travers une collecte massive de données en provenance de participants non-sélectionnés. Toutefois, le corpus de recherche sur la micro-PK par Helmut Schmidt était principalement élitiste, impliquant un travail intensif avec un petit nombre d'individus sélectionnés. Nous comparons son approche à celle du laboratoire du PEAR, avec près de 100 participants non-sélectionnés et un protocole hautement standardisé. Bien que l'étude phare du PEAR, conduite durant 12 ans, ait produit des résultats cumulés statistiquement significatifs, une réplication conçue de façon minutieuse, impliquant trois laboratoires, a obtenu des résultats non-significatifs. Nous affirmons que cet apparent échec de réplication était dû à l'hypothèse erronée que les données originales du PEAR étaient homogènes, alors qu'en réalité elles furent largement influencées par deux participants aux résultats extrêmes qui ont contribué à près d'un quart du total des données. En ignorant ce fait, la réplication a employé une surestimation de la taille d'effet pour l'étude originale et une sous-estimation de la puissance statistique nécessaire pour la réplication. Nous concluons que la recherche supporte généralement la perspective selon laquelle la micro-PK n'est pas largement distribuée, mais reste une exception. Dès lors, il semble improductif d'essayer de détecter des effets faibles à partir de volontaires non-sélectionnés. Tout comme Schmidt, les chercheurs devraient plutôt se focaliser sur l'optimisation des conditions de test et travailler intensivement avec des participants sélectionnés.

German

MIKRO-PSYCHOKINESE: AUSSERGEWOEHNLICH ODER UNIVERSAL?

ZUSAMMENFASSUNG: Die meisten Untersuchungen zur Psychokinese fallen entweder in eine elitistische Forschungstradition mit ausgewählten Teilnehmern und Beschränkung auf direkt beobachtbare Makro-PK-Effekte oder in einen universalistischen Zugang, der schwache Mikro-PK-Effekte mittels umfangreicher Datensammlungen mit unausgewählten Teilnehmern erforscht. Helmut Schmidts hochsignifikantes Korpus von Mikro-PK-Forschungen war hauptsächlich elitistisch und resultierte aus einer intensiven Zusammenarbeit mit nur wenigen, zuvor ausgewählten Personen. Wir kontrastieren seinen Zugang mit jenem des PEAR-Labors, das aus beinahe 100 unausgewählten Teilnehmern und einem hochstandardisierten Protokoll bestand. Obwohl die 12 Jahre umfassende Benchmark-Studie von PEAR kumulativ ein signifikantes Ergebnis erbrachte, fiel eine sorgfältig geplante Replikationsstudie unter Einschluss von drei Laboratorien nichtsignifikant aus. Wir vertreten den Standpunkt, dass dieser offensichtliche Replikationsfehlschlag auf der irrtümlichen Annahme beruhte, die PEAR-Daten seien über alle Teilnehmer hinweg homogen verteilt, während sie in Wirklichkeit von zwei extremen Ausreißern bestimmt wurden, die nahezu ein Viertel der Gesamtdaten ausmachten. Durch fehlende Berücksichtigung dieses Aspekts wurde die Effektstärke der Originalstudie überschätzt und die Teststärke, die zur Replikation benötigt wurde, unterschätzt. Unsere Schlussfolgerung lautet, dass die bisherige Forschung allgemein die Anschauung unterstützt, dass Mikro-PK nicht weit verbreitet, sondern eher die Ausnahme ist, und dass es unproduktiv ist, extrem schwache Effekte aus unausgewählten Freiwilligen herauszukitzeln. Dem Vorbild Schmidts folgend, sollten sich Untersucher darauf fokussieren, die Testbedingungen zu optimieren, und mit ausgewählten Teilnehmern intensiv zusammenarbeiten.