

Kritischer Kommentar
zu einer Meta-Analyse von Bösch, Steinkamp & Boller:
„Examining psychokinesis: The interaction of human intention
with random number generators - A meta-analysis“

Psychological Bulletin 2006, 132, 497-523.

Suitbert Ertel
Georg-Elias-Müller-Institut für Psychologie
Göttingen

Ein Abriss hiervon auf einem Workshop der WGFP
in Offenburg am 5. 11. 2006 vorgetragen.

Zusammenfassung

Die Meta-Analyse der RNG-Untersuchungen, die Bösch, Steinkamp & Boller (2006) vorlegen, weist einen Grundfehler auf: Als zentrale Moderatorvariable, die den Umfang (*study size*) der 380 von den Autoren gesammelten Untersuchungen operationalisieren soll, verwenden sie die in diesen Untersuchungen jeweils abgerufenen Bitmengen. Diese sind abhängig vom jeweils gewählten Bittempo, einer nicht-standardisierten, extrem variablen technischen Eigenschaft der Versuchsgeräte (RNG). Nach der vom Bittempo abhängigen Bitmenge lässt sich der Umfang der Testerhebungen in den Untersuchungen nicht bemessen, hierfür ist eine invariable Messeinheit zu verwenden. Verwendet man als Umfangvariable richtigerweise den mit konstanten Zeiteinheiten gemessenen Testbeitrag der beteiligten Probanden, dann verschwindet der *Small-Study*-Effekt, mit dem die Autoren einen vermeintlichen Publikationsbias belegen wollen. Ein Publikationsbias, wenn überhaupt vorhanden, darf als vernachlässigbar gering erachtet werden. Der PK-Gesamteffekt, über alle Untersuchungen gemittelt, ist hochsignifikant. Der ebenfalls hochsignifikante Unterschied der Effektstärken, der zwischen selektierten und nicht-selektierten Testpersonen auftritt und der schon bei zwei früheren Meta-Analysen anderer Autoren aufgewiesen wurde und den Tatbestand unterschiedlicher Psi-Begabungen belegt, spricht für eine überwiegende Echtheit der in den 380 Untersuchungen erfassten psychokinetischen Phänomene.

1. Vorschau ¹

Der folgende Kommentar behält den Charakter eines Workshop-Beitrags bei, von dem er ausgeht. Das Verständnis des Textes setzt die Kenntnis des Artikels von Bösch et al. voraus.

Der Kommentar gilt folgenden Fragen:

- **Gibt es den Small-Study-Effekt wirklich?** Für die Autoren Bösch et al. ist der Small-Study-Effekt "evident". Der Nachweis seiner Existenz bildet den Eckstein ihrer gesamten Argumentation.
- **Wie ist die Asymmetrie der PK-Effektgrößen im unteren Bereich der Study-Size im Funnelplot zu erklären?** Für Bösch et al. handelt es sich um Anzeichen eines Publikationsbias. Eine andere Erklärung suchen die Autoren nicht.
- **Haben Bösch et al. den Versuchsumfang (*study size*) richtig operationalisiert?** Sie operationalisieren den Umfang einer Studie durch die Bitmenge, die von den verwendeten RNG-Geräten der jeweiligen Untersuchungen abgerufen wird. Die Entscheidung wird nicht begründet, eine andere Operationalisierung wird nicht erwogen.
- **Beeinflusst der Versuchsumfang die PK-Effektgröße tatsächlich?** Diese Frage sollte mit geeigneteren Mitteln untersucht werden.
- **Ist die PK-Effektgröße nicht auch von der Selektion psi-begabter Testpersonen abhängig?** Die Abhängigkeit der PK-Leistung von individuell unterschiedlichen paramentalen Sensitivitätsgraden würde man nicht durch einen Publikationsbias erklären können.
- **Haben Bösch et al. wirklich Grund, an der Existenz der PK-Effekte zu zweifeln?** Zweifel an vorliegenden empirischen Befunden hat sich empirisch und argumentativ zu legitimieren. Nach einer fehlerbereinigten Auswertung der Daten von Bösch et al.

¹ Der Kommentar stützt sich auf eine Reanalyse der von Bösch et al. in ihrer Meta-Analyse verwendeten Daten, von denen mir Holger Bösch-Hartmann dankenswerterweise eine Kopie überließ. Auch sei ihmgedankt für die Beantwortung diverser Fragen zu seiner Studie. Den Kommentatoren Radin, Nelson, Dobyms & Houtkooper (2006), deren Text im gleichen Heft des Psychological Bulletin publiziert wurde, standen die Daten nicht zur Verfügung, der vorliegende Kommentar akzentuiert und ergänzt einige dort vorfindbare Argumente. Erst nachdem ich den vorliegenden Kommentar beendet hatte, erfuhr ich von vier weiteren Beiträgen zum Thema der Mind-Matter Interaction im neuesten Heft des Journal of Scientific Exploration, die hier nicht berücksichtigt werden konnten (Nelson; Radin et al.; Radin; Schub). Für die Durchsicht des vorliegenden Textes und für Kommentare möchte ich mich bei Joop Houtkooper und Ulrich Timm bedanken.

und angesichts einschlägiger Ergebnisse anderer Autoren zur REG-PK könnte die Legitimation des von den Autoren wiederholt geäußerten Zweifels an der Echtheit der psychokinetischen Effekte zum Problem werden.

Bösch et al. sind der Meinung:

We believe that, ultimately, all of these findings [positive overall effect of REG studies; existence of a small-study effect; heterogeneity of the database] could be explained through publication bias and there is currently no other model available to clarify the data any better“ (2006b, 536).

Doch waren die Autoren selbstkritisch genug, um dem zu entsprechen, was Greenhouse & IYengar (1994) Meta-Analytikern ans Herz legt (p. 384).

“At every step in a research synthesis decisions are made that can affect the conclusions and inferences drawn from the analysis... In other words, it is important to check how sensitive the conclusions are to the method of analysis or to changes in the data”

Im Folgenden soll eingebracht werden, was den Autoren angeblich nicht verfügbar war: ein anderes und besseres Modell. Tatsächlich hätten die Autoren ein solches selbst aufgreifen und auf ihren Fall anwenden können.

2. Versuchsumfang: Trial und Run vs. Bitmenge

PK-Daten, gewonnen mit einem Random Number Generator (RNG), werden standardmäßig nach Einheiten quantifiziert, die den *Runs* und *Trials* der Testdurchführung entsprechen. Der Umfang einer Untersuchung U (*study size*) ergibt sich aus der Anzahl J der Runs und der Anzahl K der Trials, die von N Testpersonen absolviert werden ($U = f(J, K, N)$).

Bösch et al. legen als Auswertungseinheit die Summe der in den Untersuchungen produzierten bzw. abgerufenen Bits zugrunde, für deren Notation sie "N" wählen ($U = f("N")$). Ihr "N" ist der Gesamt-Output der jeweils benutzten RNG-Geräte, das Arbeitsergebnis der jeweiligen Technik, wenn man so will, nicht das der Testpersonen. Mit dieser Fehlentscheidung war ein Scheitern des Bösch et al.-Projekts vorprogrammiert, was im Folgenden näher begründet wird. (Um die für Personenzahlen allgemein verwendete Notation N vom "N" der Autoren, das etwas anderes bedeutet, besser unterscheiden zu können, wird zur Bezeichnung von "N" von Bösch et al. hinfort N_B verwendet).

Das *Bit* ist eine technische Eigenschaft des verwendeten RNG, es wird z.B. aus einer Zener-Diode gewonnen. Ein Bit ist etwas anderes als ein Spielwürfel des klassischen PK-Versuchs. In einem PK-Versuch mit RNG-Technik nimmt die Testperson z. B. einen Lichtpunkt wahr, dessen Auf- und Abbewegung auf einem Display vom Fluss der zufallsproduzierten bzw. abgerufenen RNG-Bits gesteuert wird, während sie die Bewegung des Lichtpunktes mental zu beeinflussen versucht. Man benötigt die im Gerät innerhalb der Zeitspanne eines jeweiligen Trials (oder eines Runs) produzierten bzw. abgerufenen Eins- und Null-Bitmengen lediglich zur Berechnung ihrer Abweichung von der zufallserwarteten Gleichverteilung. Tritt eine Abweichung auf und ist sie hinreichend stabil, wird dies als PK-Einfluss gewertet. Die Bits sind der Testperson erlebnismäßig nicht gegeben.

Tabelle 1

Gliederung der Einheiten eines Standard RNG-Experiments mit zugeordneten Bitmengen.

Series 1									Series
Run #1			Run #2			Run #3			Runs
low	high	base	low	high	base	low	high	base	Trials
2	2	2	2	2	2	2	2	2	Bitmenge #1
200	200	200	200	200	200	200	200	200	Bitmenge #2
1000	1000	1000	1000	1000	1000	1000	1000	1000	Bitmenge #3
10^6	10^6	10^6	10^6	10^6	10^6	10^6	10^6	10^6	Bitmenge #4

Tabelle 1 zeigt die Einteilung eines typischen RNG-Experiments in Abschnitte. Eine Testperson absolviert eine *Versuchsserie*, die in z. B. drei *Runs* aufgeteilt wird. Jeder Run wird noch einmal aufgeteilt, meist in drei *Trials*, die mit verschiedenen Beeinflussungsintentionen absolviert werden. Beim Trial "low" bemüht sich die Testperson, z.B. einen Lichtpunkt mental nach unten zu bewegen, beim Trial "high" will er ihn nach oben befördern, beim Trial "baseline" will er die Ausgangsstellung des Punktes möglichst beibehalten. Ein Trial kann z. B. zwei Minuten dauern, während dieser Zeit können beispielsweise 200 Bits abgerufen werden.

3. Die Bitmenge N_B ist abhängig von der Bitdichte und somit zur Operationalisierung des Versuchsumfangs ungeeignet.

Wieviel Bits für einen Trial produziert bzw. abgerufen werden, hängt von der *Bitdichte* des RNG ab, d. h. von der *Bitmenge pro Zeiteinheit*. Die Bitdichte ist je nach Wahl der Technik verschieden. Einem Trial von z. B. zwei Minuten können 2 Bits, oder auch 200, 1000, 1000000 Bits zugrunde liegen. Der Aufwand der Testpersonen, die Dauer ihrer Testtätigkeit und die Zahl der dabei geleisteten Trials sind von der Bitdichte unabhängig.

Frage: Warum überhaupt wollen Bösch et al. den Versuchsumfang operationalisieren, was hat er mit dem PK-Effekt zu tun? Antwort: Die Autoren benötigen den Versuchsumfang, um zwischen *big studies* und *small studies* differenzieren zu können. An dieser Differenzierung sind sie interessiert, weil sie einen Beweis brauchen für ihre Annahme, dass ihr Datensatz durch einen Publikationsbias beeinträchtigt wurde. Der wäre gegeben, wenn sich die PK-Effektgrößen in einem Funnelplot bei den *small studies* asymmetrisch verteilen, d. h. wenn sich vor allem bei den umfangkleineren Untersuchungen ein relativer Mangel an Studien mit niedrigen oder negativ gerichteten Effektstärken zeigt.

Tatsächlich wäre die Existenz eines Publikationsbias mit Nachweis eines *small study*-Effekts aus zwei sich ergänzenden diskutablen Gründen evident; erstens aus rein statistischen Gründen: Bei *small studies*, wenn also wenig Tests durchgeführt werden, liegt bei einem Signifikanzniveau von sagen wir $p = .05$ eine größere Effektstärke vor als bei den *big studies*. Ein Effekt muss bei nur wenig Beobachtungen stärker sein, um signifikant zu werden. Eliminiert man aus dem Studiensample alle nicht-signifikanten Fälle, dann bleiben im unteren, dem *small studies* Bereich des Funnelplots Studien mit höherer Effektstärken zurück, die im oberen Bereich nicht oder seltener vorkommen.

Zweitens, die Neigung, nicht-signifikante Ergebnisse oder Ergebnisse mit erwartungswidriger Richtung nicht zu publizieren (*file drawer effect*), soll sich nach Bösch et al. bei *small studies* leichter auswirken als bei Untersuchungen, die größeren personellen und finanziellen Aufwand erfordern. Dieser Punkt soll terminologisch verdeutlicht werden: Zwischen dem hypothetischen *File-Drawer-Effekt* und einem zusätzlich angenommenen Moderator dieses Effekts, dem *File-Drawer-Schwellen-Effekt* (so möchte ich ihn nennen) ist zu unterscheiden. Danach wird die Schwelle für eine File-Drawer-Entscheidung, der Verzicht auf Veröffentlichung nicht-signifikanter Versuchen mit geringerem Umfang früher erreicht als bei solchen mit größerem Umfang. M. a.W. man lässt nicht-signifikante Forschungsergebnisse eher in der Schublade liegen, wenn weniger Versuche gemacht wurden. Die umfangsgrößereren Studien, die nichtsignifikant ausfallen, bleiben seltener unpubliziert. Bösch et al. glauben, mit ihrer *Figure 2* (p. 508), die eine Asymmetrie der Effektgrößen im unteren Bereich der *study size* N_B aufweist (" *small study*"-Effekt) , die Existenz eines Publikationsbias bewiesen zu haben.

Die "*small study*"-Asymmetrie wird auch sichtbar anhand eines Teil-Datensatzes von nur 140 Studien, die ich dem Pool der 380 Bösch et al.-Studien entnommen habe (s. das Funnelplot in Abbildung 1). Auf die Teilmenge der 140 Studien wurde zurück gegriffen, weil für die weiteren Auswertungen nur Studien verwendbar waren, die bei den folgenden Variablen Einträge aufwiesen (die Variablen-Nomenklatur ist dem Bösch-Datensatz entnommen):

1. Anzahl der **Gesamtbits** einer Studie (N_B) = *Bitmenge*
2. Anzahl der **Bits pro Sekunde** (FREQ) = *Bittempo* (oder *zeitliche Bitdichte*)
3. Anzahl der **Testpersonen** (PARTS)

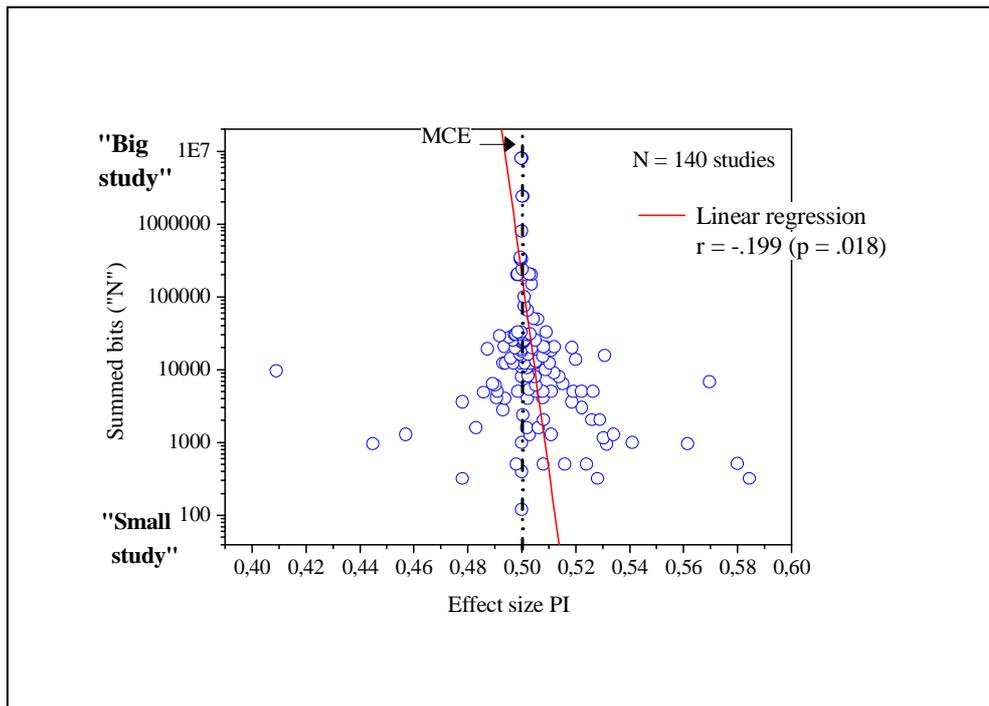


Abbildung 1

PK-Effektgrößen PI von 140 Studien in Abhängigkeit von N_B (logarithmiert), d. h. von der Variablen, die Bösch et al. als Maß für den Studienumfang (den Untersuchungsaufwand) zugrunde legen.

Abbildung 1 dient einerseits zur Erinnerung an *Figure 2* von Bösch et al., mit der sie weitgehend übereinstimmt, andererseits auch als Beleg dafür, dass der für die weiteren Berechnungen herausgelöste Datensatz mit 140 Studien das für diese Autoren wichtigste Merkmal, die Asymmetrie des Funnelplots im unteren Abbildungsbereich der "small studies", mit etwa gleicher Stärke aufweist wie der volle Datensatz mit 380 Studien.

Doch ist zu allererst empirisch zu zeigen, dass N_B , die von Bösch et al. gewählte Größe, kein Indikator des Studienumfangs darstellt, dass $U = f(N_B)$ nicht gilt. Die Autoren führen ihre Interpretation der Asymmetrie ihres Plots mit bemerkenswerter Sicherheit vor:

"Publication bias appears to be the easiest and most encompassing explanation for the primary findings of the meta-analysis" : "... a striking variability of effect size a clearly visible small study effect.. the publication process was clearly selective..." (p. 517)

Dies kontrastiert mit den Warnungen, die vorsichtige Methodenexperten in einem Artikel "The case of the misleading funnel plot" (Lau et al. 2006) aussprechen:

"[There are] many reasons why funnel plot asymmetry may not reflect publication bias"; "enormous uncertainty and subjectivity remains in the visual interpretation of the same plot by different researchers" (Lau et al., p. 597 f.).

Die folgende Reanalyse der Bösch et al.-Daten zeigt, dass es sinnvoll gewesen wäre, hätten sich die Autoren solchen Warnungen entsprechend verhalten.

4. Das Bittempo: Eine zentrale, von den Autoren vernachlässigte Variable.

Es ist schwer zu verstehen, warum die Autoren *FREQ*, ihre Variable des Bit-Tempos (die zeitliche Bitdichte), nicht mit ausgewertet haben. Hätten sie dies getan, dann hätten sie darauf aufmerksam werden müssen, dass sich ein Funnel-Plot, bei dem auf der Y-Achse die Angaben für das **Bittempo** abgetragen werden, vom Funnelplot ihrer *Figure 2* nicht unterscheidet, in welcher auf der Y-Achse die **Bitmenge** N_B als Maß für den Versuchsumfang (*study size*) eingetragen wurde. Dies zeigt ein Vergleich unserer Abbildungen #2 und #1. In beiden Funnelplots liegt eine Tendenz der Effektgrößen vor, mit höheren Einträgen auf der Y-Achse in die negative Richtung abzuweichen. Auf der Skala des Bitempos *FREQ* (Abbildung 2) verteilen sich die Effektgrößen der Studien genauso asymmetrisch wie auf der Skala der Bitmenge N_B (Abbildung 1). Wie ist das möglich?

Das **Bittempo** ist das primäre Merkmal, die in einer Untersuchung abgerufene **Bitmenge** ist ein sekundäres Merkmal. Die Bitmenge hängt vom Bittempo ab, nicht umgekehrt, ebenso wie die Fahrstrecke eines Automobils von der Fahrgeschwindigkeit abhängt und nicht die Fahrgeschwindigkeit von der Fahrstrecke. Die Asymmetrie des Funnelplots der Abbildung 1 beruht also auf der Asymmetrie des Funnelplots der Abbildung 2 und nicht umgekehrt. Wenn man Bit überhaupt als Maß für den Versuchsumfang in Betracht ziehen sollte, dann müsste man das primäre Merkmal des Bitempos und nicht das davon abhängige Merkmal der Bitmenge verwenden. Dieses aber verbietet die Logik: Das Bittempo ist als Maß für den Versuchsumfang ebensowenig verwendbar wie das Fahrtempo eines Autos als Maß für eine zurückgelegte Fahrstrecke verwendbar ist.

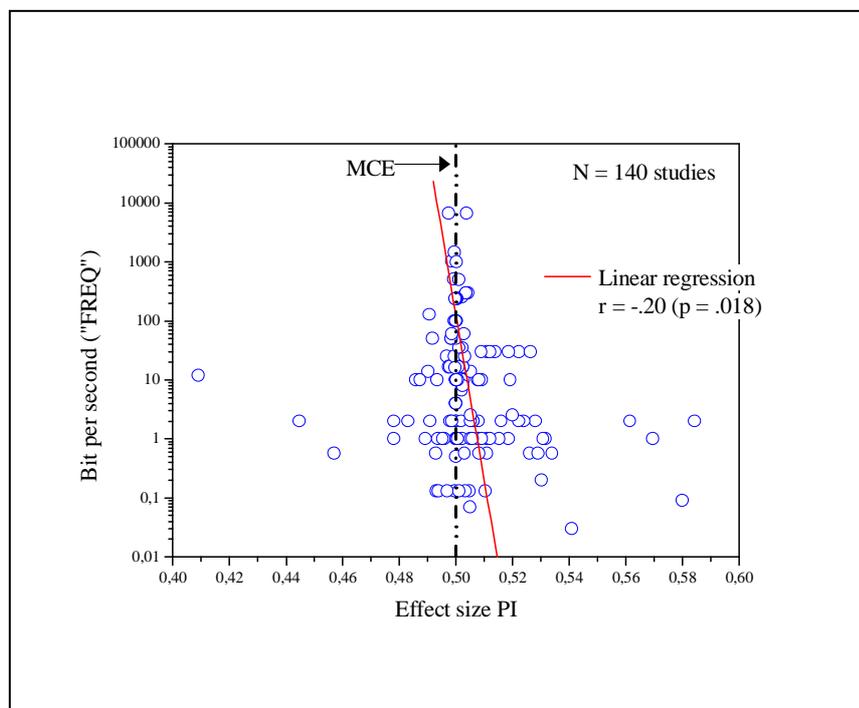


Abbildung 2

PK-Effektgrößen PI von 140 Studien in Abhängigkeit von *FREQ* (logarithmiert), d. h. vom Bit-Tempo oder der zeitlichen Bit-Dichte, die Bösch et al. als informative Variable übersehen

Erste Schlussfolgerung: Die Asymmetrie der Effektgrößen im unteren Y-Achsenbereich des Funnelplots, sowohl die der Abbildung 1 als auch der Abbildung 2, lässt sich als Indikator eines Publikationsbias nicht verwerten. Zweite Schlussfolgerung: Die Behauptung, der Datensatz der RNG-Untersuchungen sei durch einen Entzug von Forschungsberichten mit nicht-signifikanten Ergebnissen beeinträchtigt, verliert ihre vermeintliche Evidenz.

Die Beobachtung, dass das Bittempo (die Bitdichte) mit den Effektgrößen der RNG-PK-Studien zusammenhängt, ist - soweit man eine Prozessanalyse des PK-Geschehens zum Thema machen will - durchaus interessant. Auf einen analogen Zusammenhang sind erstmals Dobyns et al. (2004) aufmerksam geworden. Die Autoren hatten entgegen anfänglicher Erwartungen feststellen müssen, dass bei Verwendung einer extrem hohen Bitdichte die mental zu beeinflussende Größe den Beeinflussungswünschen der Testpersonen nicht folgt, dass sie dann sogar überzufällig stark zur Gegenrichtung hin umschlägt. Dieses unerwartete Verhalten des zu beeinflussenden Objekts wussten sie nicht zu erklären.

Immerhin wurde ihr Ergebnis in einer Folge-Untersuchung von ihnen selbst schon repliziert. Das hier in Abbildung 2 dargestellte Ergebnis mit den Bösch et al.-Daten ist eine weitere Replikation ihres Befundes. Je höher das Bit-Tempo gewählt wird, umso größer werden offenbar anteilmäßig die PK-Effekte mit intentionswidriger Richtung. Das zeigt die Regressionslinie, wie sie Abbildung 3 verdeutlicht. Die Y-Achse repräsentiert den prozentualen Anteil der Psi-Hitting Effekte an der Gesamtzahl der Effekte. Der prozentuale Anteil an Psi-Hitting sinkt, wenn das Bittempo zunimmt, während der Anteil der Psi-Missing-Fälle entsprechend zunimmt, d. h. die PK-Effekte zeigen dann öfter die wunschkonträre Richtung. Solange man diesen Zusammenhang nicht erklären kann, mag eine anthropomorphe Analogie die Lücke füllen: *Psi reagiert auf den allzu dichten Bit-Output der RNG-Geräte tendenziell ähnlich wie eine Frau, der ein Freier zu dicht auf die Pelle rückt. Sie tut das Gegenteil von dem, was er und auch sie vielleicht sich wünschen. Sie sucht Distanz.*

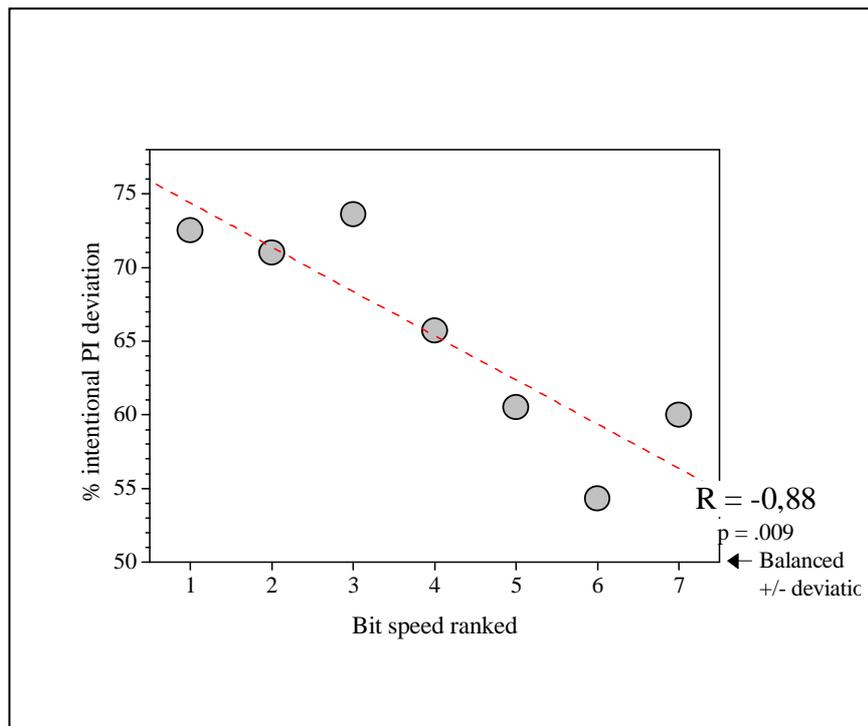


Abbildung 3

Prozentsatz der Effektgrößen, deren Richtung der Intention der Testpersonen entsprechen, im Verhältnis zur Gesamtheit der Effektgrößen, die der Intention entsprechen und nicht entsprechen. intentionswidrigen Effektgrößen.

Das Ergebnis der Abbildung 3 allein genügt nicht als Stütze der Annahme, dass es bei Verwendung eines RNG-Geräts mit überhöhtem Bit-Abruf zum Psi-Missing kommt. Doch handelt es sich hier um eine spezielle Hypothese einer PK-Prozess-Forschung, die im vorliegenden Zusammenhang nicht das Thema ist.

5. Versuchsdauer : eine sinnvolle Operationalisierung des Versuchsumfangs.

Zurück zum Versuchsumfang (*study size*), an dem Bösch et al. so wie die meisten Bias-Detektive unter den Meta-Analytikern sehr interessiert sind (Begg, 1994). Hängt der Versuchsumfang, wenn man ihn sachgerecht operationalisiert, mit der PK-Effektgröße vielleicht doch zusammen? Wie ist es um den behaupteten Zusammenhang zwischen dem Versuchsumfang und einer verstärkten Publikationsasymmetrie bei den umfangskleineren Studien tatsächlich bestellt?

Der Umfang einer Studie, der Testaufwand, den die Beteiligten leisten, lässt sich durch die Gesamtdauer der in der Studie durchgeführten Tests definieren. Die Zeitdauer kann als Maß verwendet werden, weil die Zeiteinheit konstant ist, sofern die in den verschiedenen Studien verwendeten Uhren gleich schnell und nicht wie "Bit-Uhren" mit extrem verschiedener Geschwindigkeit ablaufen. Erfreulicherweise steht zur Berechnung im Datensatz von Bösch et al. die Variable *FREQ* zur Verfügung, die Zahl der Bits pro Sekunde. Die für eine Untersuchung aufgewendete Zeit ergibt sich aus der in ihr aufsummierten Bitmenge (N_B) dividiert durch die zugehörige Bitzahl pro Sekunde (*FREQ*). *Eine Analogie: Die Fahrdauer eines Automobils ergibt sich aus Fahrstrecke und km/h. Legt man im PKW 600 km mit einem Tempo von 60 km pro Stunde zurück, dann benötigt man dazu 10 Stunden.*

Wertet man die Bösch-Daten mit Verwendung der Versuchsdauer als Indikator des Versuchsumfangs aus (anstatt mit der Bitmenge), dann verteilen sich die Effektgrößen der Untersuchungen im Funnel-Plot so wie dies Abbildung 4 zeigt: Versuchsumfang und PK-Effektgröße hängen **nicht** miteinander zusammen, nicht einmal andeutungsweise. Die Regressionslinie läuft mit der Y-Achse parallel, für *small*, *middle* und *big studies* liegen PK-Effekte der gleichen Größenordnung vor.

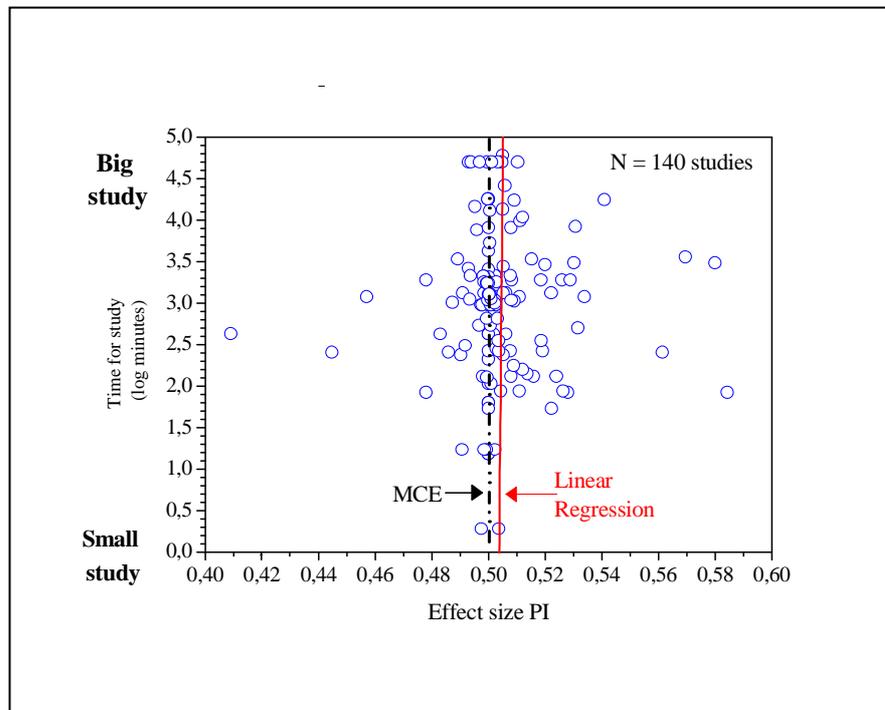


Abbildung 4

PK-Effektgrößen PI von 140 Studien
in Abhängigkeit vom zeitlichen Testaufwand (logarithmiert), d. h. der Variablen,
die Bösch et al. eigentlich als Maß für den Studienumfang (den Versuchsaufwand)
hätten zugrunde legen sollen.

Es ist verwunderlich, dass Bösch et al. auf ihre zweifelhafte Operationalisierung des Versuchsumfangs (durch die Bitmenge N_B) nicht schon beim Anblick ihrer *Figure 1* (p. 505) aufmerksam wurden. Dort wird ein Funnelplot abgebildet, in welchem die Y-Achse N repräsentiert, die Anzahl der an den Untersuchungen jeweils beteiligten Testpersonen. Die Zahl der verwendeten Testpersonen bestimmt den Versuchsumfang in erster Linie, sie wird zu seiner Ermittlung in jedem Fall benötigt (s. oben: $U = f(J, K, N)$). Auffällig ist, dass der Personenfaktor im Funnelplot der *Figure 1* der Autoren keinerlei Anzeichen für einen *small study* Effekt erkennen lässt. Die Autoren ignorieren diesen wichtigen Befund, der einem Meta-Analytiker beim Betrachten der *Figure 1* nicht entgehen sollte.

Da festzustellen ist, dass zwischen Versuchsumfang (*study size*) und PK-Effektgröße kein Zusammenhang vorliegt, auch nicht nach Korrektur des Berechnungsfehlers der Autoren, fehlt ihrer Behauptung eines Publikationsbias jede Evidenz. Einen Publikationsbias nur spekulativ vorauszusetzen, wäre nicht legitim. Nicht für alle Bereiche wissenschaftlicher Forschung liegt bei Publikationssammlungen ein File-drawer-Defizit vor:

"We have examined many hundreds of unpublished studies in the personnel selection research domain. We have found that it is not true that data were suppressed or omitted." (Hunter & Schmidt, 2004, p. 508).

Überdies ist hier zu bedenken, dass die *Ethical and Professional Standards of the Parapsychological Association* (P. A.) seit 1986 folgende Regel enthalten:

"The importance of full publication for a proper representation of the state of scientific knowledge has become clear because of Meta-Analyses. The P. A. encourages members to publish all legitimate research whether the hypothesis was supported or rejected. Both significant and non-significant results in experiments can provide useful information."

Dass Parapsychologen einen Anti-Filedrawer-Appell dieser Art verbreiten, lässt darauf schließen, dass man sich in ihren Reihen dieser Fehlerquelle besonders bewusst ist und dass man sie vermeiden möchte - eine ausgeprägte Furcht vor der Majorität der Psi-Ungläubigen wird das Motiv sein. Allenfalls lässt sich bei Parapsychologen gelegentlich ein File-drawer-Verhalten *mit umgekehrtem Vorzeichen* beobachten. J. B. Rhine behielt seine ersten PK-Ergebnisse, die er mit Würfeln gewann, neun Jahre lang im *file drawer*, wie Bösch et al. bemerken (p. 498) - offenbar aus Furcht vor zuviel Opposition, die zu erwarten wäre, wenn nicht massenhaft Replikationen des Effekts vorliegen. Heute noch findet sich eine ähnlich motivierte Zurückhaltung bei Herausgebern von parapsychologischen Fachzeitschriften: Forschungsberichte mit ungewöhnlich starken Psi-Effekten werden von ihnen und ihren Reviewern oft misstrauisch entgegen genommen und eher abgelehnt als Artikel mit schwachen oder fehlenden Effekten. Eigene Erfahrungen liegen hier vor. Diese Einstellung kontrastiert auffällig mit Einstellungen von Zeitschriften-Herausgebern im Mainstream der Psychologie, bei welchen man ein Editorial wie das von von Melton (1962) finden kann, in welchem empfohlen wird, bei nichtsignifikanten Ergebnissen auf eine Publikation der betreffenden Studie zu verzichten.

Hinzu kommt, dass Parapsychologen seit dem Bekanntwerden der oft auftretenden *Psi-Missing*-Tendenz (Abweichung von der Zufallserwartung in der von den Testpersonen nicht gewünschten Richtung) auf signifikant negative Resultate in der Regel durchaus offen und interessiert reagieren (Rhine, 1952, 1969, Timm, 1968, Kennedy, 1979, s. oben auch Dobyns et al., 2004). Ergebnisse mit wunschwidriger Abweichungsrichtung werden mit der von Rao (1965) so genannten *Bi-Direktionalität*, einem merkwürdigen, aber vielfach replizierten

Prozessmerkmal der Psi-Phänomene in Zusammenhang gebracht. Vergleichen kann man das Psi-Missing mit der normalpsychologischen *Reaktanz* von Probanden bei Persuasionsversuchen: Die Zielpersonen tun und denken u. U. das genaue Gegenteil von dem, was man ihnen zu tun und zu denken nahe legt.

Die Einstellung von Psi-Forschern gegenüber Effekten, die den Intentionen und Wünschen der Testpersonen widersprechen, ist somit eine völlig andere als etwa die von Medizinern, die einen erwartungswidrigen Ausgang einer Behandlung etwa von Krebspatienten, deren Sterbensrate durch eine getestete neue Therapieform nicht vermindert, sondern erhöht wird, kaum als interessanten Zusatzbefund betrachten. Für Ärzte wären dies echte Misserfolge, die der Fachöffentlichkeit eher vorenthalten werden, vermutlich mit Begründungen wie: *"Irgendetwas muss da nicht bedacht worden sein; man würde sich mit einem veröffentlichtem Misserfolg ja blamieren usw."*. Ein File-drawer-Verhalten von Parapsychologen ließe sich, da für sie eine wunschwidrige Abweichung oft ebenso willkommen ist wie eine wunschgemäße, allenfalls für *nicht-signifikante* Befunde ausdenken, nicht für Befunde mit signifikant negativer Richtung.

Obgleich unbestritten ist, dass zumindest in der orthodoxen psychologischen Forschungspraxis tatsächlich erzielte nichtsignifikante Ergebnisse häufiger nicht veröffentlicht werden als signifikante (dies wurde vielfach belegt, Begg, 1994) und eine, wenn auch schwächere Tendenz dieser Art ebenfalls bei Parapsychologen denkbar ist (für sie wurde dies noch nicht belegt), verbieten sich blinde Verallgemeinerungen. Es gibt Bedingungen, unter denen die Publikation signifikant *negativer* Ergebnisse sogar wahrscheinlicher ist als die Publikation signifikant positiver Ergebnisse (Begg, 1994, p. 401).

6. Die Selektion psi-begabter Testpersonen: Eine stark varianzaufklärende Variable.

Nachdem auf den *Small-study*- und *File-drawer*-Alarm Entwarnung folgen darf, lässt sich unter bereinigten Bedingungen die Hauptfrage der Bösch et al.-Untersuchung aufgreifen:

Wird die PK-Hypothese, wonach sich die RNG-gesteuerten Bewegungen eines Objekts durch bloßes Wünschen mental beeinflusst werden, durch eine meta-analytische Auswertung aller bislang vorliegender Forschungsergebnisse geschwächt oder gar widerlegt? Ergibt sich als Resultat ein begründeter Zweifel?

Ein Blick auf die Verteilung der Z-Werte der RNG-Untersuchungen, der Originalwerte aus dem Bösch et al.-Datensatz, zeigt auffällige Unterschiede zwischen den Studien, für welche Testpersonen entweder selektiert oder nicht selektiert wurden (s. Abbildung 5). Bei Studien mit selektierten Personen sind REG-PK-Effekte öfter und höher signifikant als bei Studien mit nicht-selektierten Personen. Selektierte Testpersonen sind solche, die schon in vorausgehenden Untersuchungen psi-erfolgreich waren, ferner solche, die sich selbst als besonders psi-begabt (als "psychics") ansehen bzw. angeben, öfter spontane paramentale Erlebnisse gehabt zu haben.

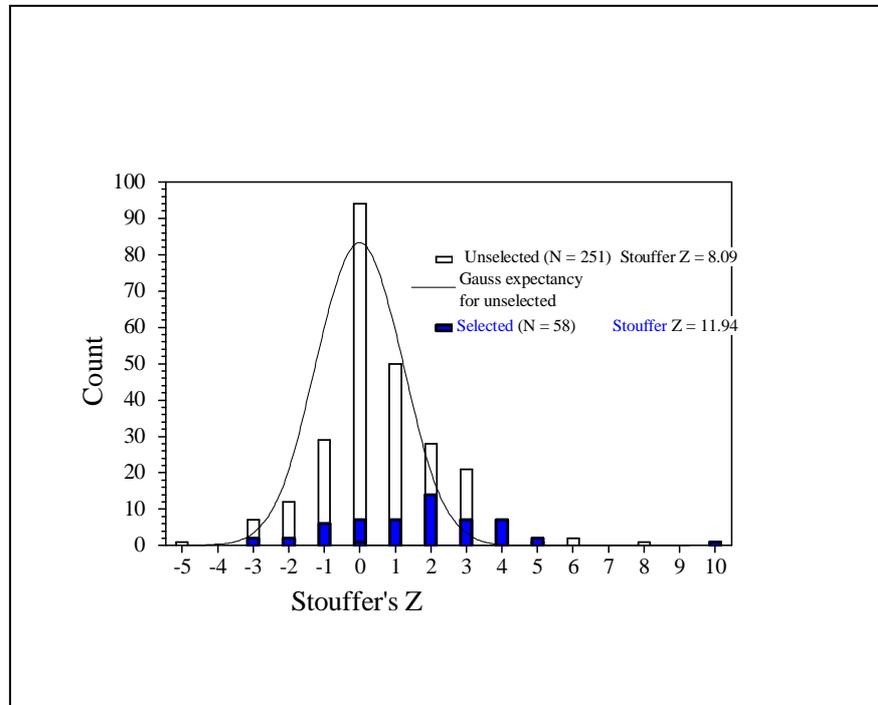


Abbildung 5

Z-Wertverteilung der kleinen Stichprobe der Studien, für die die Testpersonen selektiert wurden und der großen Stichprobe der Studien, für die die Testpersonen nicht selektiert wurden. in Abhängigkeit von „N“ (logarithmiert), d. h. der Variablen, die Bösch et al. als Maß für den Studienumfang (den Untersuchungsaufwand) zugrunde legen.

Dieser Befund wird von Bösch et al. selbst beschrieben:

"As can be seen in Table 6, the effect size of studies with selected participants is considerably larger than that of studies that did not select their participants ... The difference ... is highly significant" (p. 509).

Doch wird er von den Autoren kurzerhand umgedeutet:

"However, the two subsamples differ considerably in sample size. Studies using selected participants were considerably smaller... (p. 509).

Der Leser soll denken, es handele sich auch hier um einen Effekt der Bitmengen. Die von den Autoren zum Indikator für die *sample size* hochstilisierte Variable N_B soll wie ein *deus ex machina* nun auch noch den Zusammenhang der PK-Effektgrößen mit Persönlichkeitsmerkmalen der Testpersonen erklären und damit aus dem Weg räumen. Wie soll man das verstehen? Bösch wurde persönlich in einem Interview mit Helmut Schmidt darüber informiert, dass es vor allem darauf ankommt, für REG-Experimente Personen „mit besonderen Fähigkeiten“ auszuwählen. *„In den meisten meiner Experimente war der Psi-Effekt viel größer als nur ein Bit von 10000, wie es von ‚Massenexperimenten‘ mit unausgewählten Versuchspersonen berichtet wird. Meiner Meinung nach muss man nur die richtigen Vps finden“* (Bösch, 2005, p. 239). Weder diese Aussagen von Helmut Schmidt noch dessen respektable Effektgrößen in Bösch et al.'s Datenbank veranlassende Meta-Analytiker, ihren signifikanten Moderatorbefund mit der Selektion von Testpersonen als das zu nehmen, was er ist.

Eigentlich hätten die Autoren noch einen anderen Grund zuzugeben, dass die Stabilität der Effekte bei Psi-Begabten für die Existenz und Echtheit der Effekte spricht. Sie vertreten dieses Argument im Prinzip sogar ausdrücklich selbst, sie verwenden es gegenüber Wilson & Shadish (2006): *"If the dependencies ... really did exist they would point to the existence of psi"* (p. 534). Zu den *dependencies* - das sind die Moderatorvariablen, die sich als signifikant erwiesen, gehört die Selektion psi-begabter Personen an erster Stelle.

Frage: Was hindert die Autoren daran, das, was sie *in abstracto* richtig aussprechen, auf ihren Fall konkret anzuwenden? Antwort: Würden sie die Verknüpfung der Effektgrößen mit Psi-Sensitivität als ein Anzeichen für die Echtheit der RNG-Effekte anerkennen, würden sie sich ein Hindernis schaffen für ihr abschließendes Fazit, auf das hin ihr Artikel von Anfang an zusteuert.

"The statistical significance of the overall database provides no directive as to whether the phenomenon is genuine... one must ask whether the findings are artifactual or indicative of a genuine effect. Publication bias appears to be the easiest and most encompassing explanation for the primary findings of the meta-analysis" usw. (p. 517).

Die Abhängigkeit der Effektgröße von der Selektion der Testpersonen lässt sich graphisch kompakter darstellen (s. Abbildung 6) und zusammen mit vergleichbaren Ergebnissen aus zwei anderen Meta-Analysen interpretieren (s. Abbildung 7). Für die Ergebnisse der Meta-Analysen zu ESP (Honorton & Ferrari, 1989) und Würfeltest-PK (Radin & Ferrari, 1991) ist die Bitmenge N_B , die nach Bösch et al. auch die Effektdifferenzen zwischen Testpersonen erklären soll, allemal irrelevant. Die Unterschiede der Testergebnisse zwischen psi-sensitiven und nicht-sensitiven Versuchsteilnehmern lassen sich nicht als Artefakte wegdiskutieren.

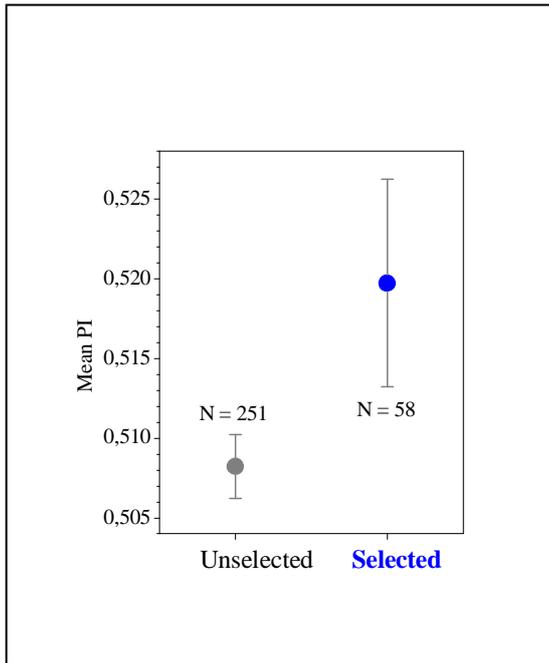
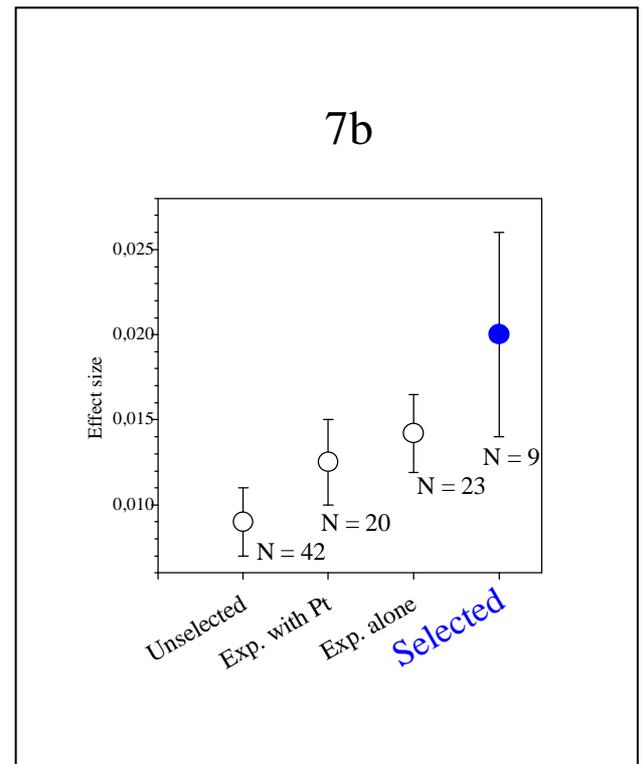
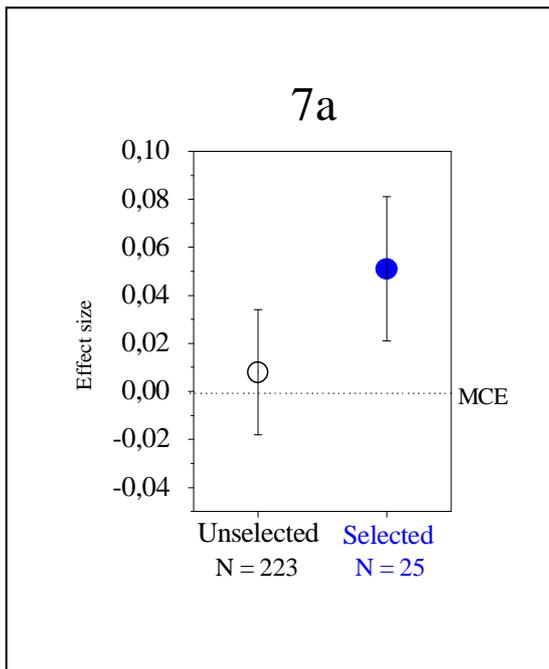


Abbildung 6

Mittlere PK-Effektgrößen PI der kleinen Stichprobe der Studien, für die die Testpersonen selektiert wurden und der großen Stichprobe der Studien, für die die Testpersonen nicht selektiert wurden (Daten von Bösch et al., 2006).



Abbildungen 7a und 7b

Effektgrößen der Studien, für die die Testpersonen selektiert bzw. nicht selektiert wurden. Entnommen einer Meta-Analyse zur Präkognition von Honorton & Ferrari (1989) (7 a) und einer Meta-Analyse zur PK mit der Würfelmethode von Radin & Ferrari (1991) (7 b).

Ich kann hier ein eigenes passendes Ergebnis hinzufügen aus einem multiple Choice-Test der ASW-PK-Kategorie. Ermittelt wurden die Rate- und Ziehquoten einer Groß-Stichprobe von unausgelesenen Probanden mit den Quoten einer Klein-Stichprobe von Probanden, die als erfolgreiche Testpersonen aus der Groß-Stichprobe ausgewählt wurden (s. Abbildung 8). Die Daten wurden mit dem Pingpongball-Test gewonnen, an dem die Anfängerstudenten in Göttingen zur Erlangung einer Bescheinigung für Versuchspersonenstunden teilnahmen. Dieser Test wurde zuhause ohne Kontrolle absolviert. Der Test mit den im Heimtest erfolgreichen Probanden fand dann unter meiner Kontrolle im Institut für Psychologie statt. In den hier verwendeten Daten der ausgewählten Probanden sind ihre Heimtest-Ergebnisse nicht enthalten. Die Ergebnisse der beiden Kohorten 2003 und 2005 wurden getrennt ausgewertet. Die Psi-Effekte bei den selektierten Probanden sind gegenüber denen der Heimtest Probanden wie bei den selektierten Testpersonen der Abbildungen 6 und 7 wesentlich stärker ausgeprägt.

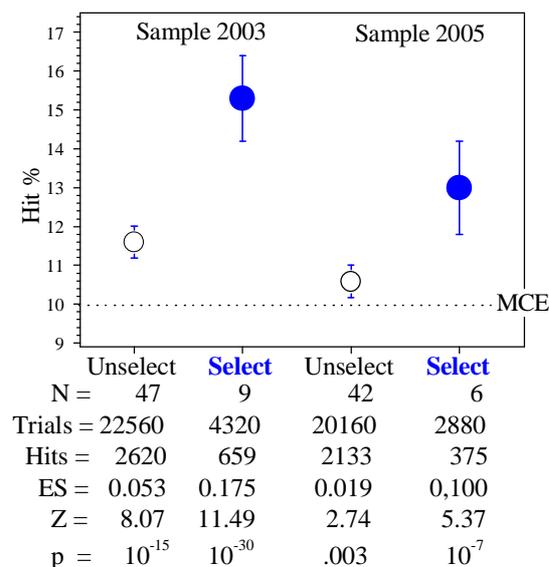


Abbildung 8

Trefferprozentage von zwei Kohorten von Studienanfängern 2003 und 2005, die einen Pingpongballtest zuhause ohne Kontrolle absolvierten („unselect“). Von ihnen wurden einige mit hohen Trefferzahlen eingeladen, den Test nochmals unter Kontrolle durchzuführen („select“). Die unter Kontrolle gewonnenen Trefferprozentage dieser Probanden liegen bedeutend höher als die der unausgewählten Probanden.

Nachwort:

Das Unglück mit der Publikation von Bösch-Steinkamp-Boller war vermeidbar.

Wiederholt brachten Bösch et al. zum Ausdruck, dass sie den PK-Effekt nicht bestätigt gefunden haben. So schreiben sie in (2006b, p. 534) :

„We do not share Radin et al. 's (2006) opinion that our meta-analysis ...confirms the existence of a PK-effect“ .

Die Behauptung von Bösch et al. , ihre gesammelten Daten würden die PK-Hypothese nicht bestätigen, lässt sich angesichts der Ergebnisse, die durch eine fehlerbereinigte Auswertung ihrer Daten gewonnen werden und angesichts der Ergebnisse früherer Meta-Analysen nicht begründen. Dies selbst zu erkennen, wäre Holger Bösch ohne weiteres möglich gewesen.

Ich verstehe nicht recht, warum Bösch mein Angebot von vor zwei Jahren, ihm bei seiner Forschung mit Rat und Tat beizustehen, nach hoffnungsvollen Anfängen in den Wind geschlagen hat. Teilnehmer am Offenburg-Workshop 2004 können mein Bemühen um ein Gelingen seiner Meta-Analysen bezeugen (Ertel, 2004, Bösch, 2004). Auch hatte Ulrich Timm, wie ich von ihm hörte, schon 2002 anlässlich eines Workshops in Offenburg bei Bösch ein Interesse an seinen Daten angemeldet. Sein statistisches-methodisches Knowhow nicht zu nutzen, das vor Ort zur Verfügung stand, ist mir unverständlich. Merkwürdig auch, dass einem anderen Kollegen, der das noch nicht eingereichte Manuskript der drei Autoren zu lesen wünschte, wie er mir mitteilte, dieses vorenthalten wurde. Der Schaden, der dem Ansehen unseres so sehr verwundbaren Grenzgebiets der Forschung durch seine Publikation entstanden ist, war nach meiner Ansicht vermeidbar.

Hat Bösch nicht gesehen, dass die Meta-Analyse als Methode auch missbraucht werden kann, worauf u. a. Stanley (2001, p. 146) warnend hinweist (Warnung vor "misuse")? Sind er und seine Mitautoren sich nicht der besonderen Verantwortung bewusst gewesen, die sie auf sich genommen haben, indem sie sich die Rolle von Schiedsrichtern über Arbeitsergebnisse ganzer Forschergenerationen zu eigen machten? Sie haben den Gesamtertrag bisheriger Forschung zur P. K. einschließlich der von hochkarätigen Wissenschaftlern in den Akzeptanzbereich der Nullhypothese hinein manövriert, so die Ergebnisse von Helmut Schmidt mit 57 Studien im Bösch et al-File, von Chuck Honorton mit 42, Dick Bierman mit 23, Bob Morris mit 18, Richard Broughton mit 14 Studien, William Braud mit 12 Studien u. a. . Die genannten Studien dieser sechs Autoren allein machen 50% des Bösch et al.-Datensatzes aus. Kann eine einzelne Forscher-Persönlichkeit es in Zukunft noch wagen, Arbeit und Zeit zur Lösung grenzwissenschaftlicher Probleme aufzuwenden, wenn damit zu rechnen ist, dass ihre Ergebnisse, zumal wenn sie die Psi-Hypothese stützen, früher oder später in die Hände von Meta-Analytikern geraten, die ihr Handwerk fehlerhaft und tendenziös verrichten? Wozu der Forschungseinsatz, wenn die Ergebnisse der besten Forscherpioniere von zukünftigen Publikationssammlern abgeerntet, verfälscht und in den Augen potentieller Interessenten entkräftet werden, womit sie die Chance einer fruchtbaren und fortschrittlichen Rezeption verlieren.

Bösch et al. führen am Anfang ihres Artikels die Leser von *Psychological Bulletin*, also typischerweise Laien auf dem Gebiet der Parapsychologie, mit der Ausführlichkeit einer ganzen Druckseite in "Psi Research", "Verifying Psi" und "Psychology and Psi" ein, treten also gegenüber dem Mainstream auf als Repräsentanten der gesamten Grenzdisziplin. Doch in der Schlussdiskussion ihrer Ergebnisse, die sich auf die PK-Forschung mit RNG-Geräten beschränkt, vergessen sie, ihren ohnehin fragwürdigen Zweifel auf das von ihnen behandelte Teilgebiet der Disziplin zu begrenzen. Die Leser ihres Artikels werden nach der anfänglichen Vorbereitung auf *Psychology and Psi* mit ihrer Hoffnung, Zweifel an Psi überhaupt möglichst bestätigt zu finden, kein Problem damit haben, das Fazit aus dem Ergebnis dieser Meta-Analyse zur PK auf die Parapsychologie insgesamt zu übertragen.

Doch ist diese Unterlassung von Bösch et al. eine nur fahrlässige? Die Autoren haben ins Feuer des Mainstream Öl gegossen, der brennend daran interessiert ist, seine Vorurteile gegenüber dem unbequemen Grenzgebiet bestätigt zu sehen ("*Die jagen doch nur einem Phantom hinterher*"). War ihnen das nicht bewusst? Die mangelnde Bereitschaft der orthodoxen akademischen Verantwortlichen zur Genehmigung von Fördermitteln und zur Aufnahme von Untersuchungsberichten zum Thema Psi in Fachzeitschriften, die die Bearbeiter parapsychologischer Probleme bislang beklagen müssen, wird längerfristig festgeschrieben werden, sobald sich die Ergebnisse aus Bösch et al.'s Meta-Analyse herumgesprochen haben. Wurde das nicht bedacht?

Der sanft-kritische Kommentar von Radin et al. (2006) im gleichen Heft der Zeitschrift wird durch das "*blowing trumpets to the tulips*" der Kommentatoren Wilson & Shadish (2006) übertönt. Diese Autoren verschärfen das Fazit von Bösch et al., indem sie ihren Aufweis des *sample size*-Effekts loben, dessen Interpretation aber als noch zu schwach, weil widersprüchlich hinstellen. Die Autoren "*would place ..[their] bets on the null hypothesis*" (p. 527), was Bösch et al. (2006b) wohl gerade recht ist: "*Wilson and Shadish (2006) correctly note that the observed effects are very small...*"(p.534). Ein Tenor des In-der-Schwebe-lassens herrscht in ihrem Schlusswort vor, womit sie einer Auseinandersetzung mit substantiellen Einwänden beider kommentierender Artikel aus dem Wege gehen: "*... the answers will remain in the eye of the beholder...*" p. 536). Der wissenschaftssoziologische "Erfolg" ihres Artikels, mit dem die Autoren von vorne herein rechnen konnten, wird durch das Nachspiel dieser Kommentare eher erhöht als vermindert.

In den **Science News**, publiziert im Internet, war unmittelbar nach der Publikation von Bösch et al. (July 22, 2006, Vol. 170, No. 4) eine Mitteilung zu lesen, die meiner Befürchtung entspricht:

From Mind to Matter: Data analysis challenges psychokinesis

"Numerous experiments in which volunteers mentally attempt to influence the output of computers that generate random sequences of 1s and 0s have failed to show that individuals can use their minds to manipulate the physical world."

References: Bösch, H., F. Steinkamp, and E. Boller. 2006. Examining psychokinesis: The interaction of human intention with random number generators—a meta-analysis. Psychological Bulletin 132(July):497-523. Abstract available at <http://content.apa.org/journals/bul/132/4>.

Bruce Bower

Holger Bösch wäre gut beraten gewesen, ein *Editorial von Harris Cooper, des Herausgebers des Psychological Bulletin, ernst zu nehmen:*

"Today, it is widely accepted that those who accumulate and integrate other people's data ought to be held to similar standards of methodological rigor as the researchers whose evidence forms the bases of their review. [They will be] influencing ... how future research will proceed " (Cooper, 2003, p. 3) (hervorgehoben nur hier).

Nun, kann man für die Zukunft der Parapsychologie anderes erwarten als was ihre Vergangenheit ihr beschert hat: Endlose Auseinandersetzungen mit Bollwerkswächtern der Orthodoxie, die, gepolstert mit dem sicheren Rückhalt der akademischen Mehrheitsmeinungen, sich dem wissenschaftlichen Fortschritt entgegen stellen? Holger Bösch kann sich zurücklehnen und mit nachhaltigen Folgen seiner Einflussnahme rechnen: mit weiteren *News* in den Wissenschaftsjournalen und am Ende mit Zitaten in den Lehrbüchern der Psychologie, die die Kluft zwischen dem vorgeschobenen Forschungsposten der Grenzgänger der Wissenschaft und dem Hinterland, mit dem sie engstens verbunden bleiben möchten, vergrößern.

Literatur

- Begg, C. B. (1994). Publication bias. In H. Cooper & L. V. Hedges (Eds). The handbook of research synthesis. New York, NY: Russel Sage.
- Bösch, H.(2004). 'Extra-sensory perception after sixty Years' (1940) - Eine Reanalyse der ersten Meta-Analyse in der Wissenschaftsgeschichte (Vortrag gehalten auf dem jährlichen Workshop der WGFP).
- Bösch, H. (2005). Zur systematischen Auswahl von Versuchspersonen. Ein Interview mit Helmut Schmidt. Zeitschrift für Anomalistik,5,238-243.
- Bösch, H., Steinkamp, F. & Boller, E. (2006a). Examining psychokinesis: The interaction of human intention with random number generators - A meta-analysis. Psychological Bulletin, 132, 497-523.
- Bösch, H., Steinkamp, F. & Boller, E. (2006b). In the eye of the beholder: Reply to Wilson and Shadish (2006) and Radin, Nelson, Dobyons & Houtkooper (2006). Psychological Bulletin, 132, 533-537.
- Cooper, H. (2003). Editorial. PsychologicalBulletin , 129, 3-9.
- Dobyons, Y. H., Dunne, B. J., Jahn, R. G., & Nelson, R. D. (2004). The MegaREG experiment: Replication and interpretation. Journal od Scientific Exploration, 18, 369-397.
- Ertel, S. (2004) Ließen die Parapsychologen der Frühzeit nichtsignifikante Befunde in der Schublade liegen? Holger Bösch ist wohl auf falscher Fährte (Korreferat zu einem Vortrag von Holger Bösch, gehalten auf dem jährlichen Workshop der WGFP in Offenburg).

- Honorton, C. & Ferrari, D. C. (1989). "Future telling": A meta-analysis of forced-choice precognition experiments, 1935-1987. *Journal of Parapsychology*, 62, 255-276.
- Hunter, J.E. & Schmidt, F. L. (2004). *Methods of Meta-Analysis. Correcting Error and Bias in Research Findings*. New York, NY: Sage.
- Greenhouse, J. B. & Iyengar, S. (2004). Sensitivity analysis and diagnostics. In H. Cooper & L. V. Hedges (Eds). *The handbook of research synthesis*. New York, NY: Russel Sage.
- Kennedy, J. E. (1979). Consistent missing. A type of information processing error in ESP. *Journal of Parapsychology*, 43, 113-128. Zugänglich auch über <http://jeksite.org/psi/jp79b.pdf>
- Lau, J., Ioannis, J. P.A., Terrin, N., Schmid, C. H., & Olkin, I. (2006). The case of the misleading funnel plot. *British Medical Journal*, 333, 597-600).
- Melton, A. (1962). Editorial. *Journal of Experimental Psychology*, 64, 553-557.
- Nelson, R. D. (2006). Time-normalized yield: A natural unit foreffect size in anomalies experiments. *Journal of Scientific Exploration*, 20, 177-179.
- Rao, K. R. (1965). The bidirectionality of psi. *Journal of Parapsychology*, 29, 230-250.
- Radin, D. & Ferrari, D. C. (1991). Effects of consciousness on the fall of dice: A meta-analysis. *Journal of Scientific Exploration*. 5, 61-83.
- Radin, D., Nelson R. and Dobyns, Y. & Houtkooper, J. (2006). Reexamining psychokinesis: Comments on Bösch, Steinkamp, and Boller (2006). *Psychological Bulletin* 132, 529-523.
- Radin, D., Nelson, R., Dobyns, Y. & Houtkooper, J. (2006). Assessing the evidence for mind-matter interaction effects. *Journal of Scientific Exploration*, 20, 361-374.
- Radin, D. (2006). Experiments testing models of mind-matter-interaction. *Journal of Scientific Exploration*, 20, 375-401.
- Rhine, J. B. (1952). The problem of psi-missing. *Journal of Parapsychology*, 16, 90-129.
- Rhine, J. B. (1969). Psi-missing re-examined. *Journal of Parapsychology*, 33, 1-38.
- Schub, M. H. (2006). A critique of the parapsychological number generator meta-analysis of Radin & Nelson. *Journal of Scientific Exploration*, 20, 402-419.
- Stanley, T. D. (2001). Wheat from chaff: Meta-analysis as a quantitative literature review. *Journal of Economic Perspectives*. 15, 131-150).
- Timm, U. (1969). Mixing-up of symbols in ESP card experiments as a possible cause for psi-missing. *Journal of Parapsychology*, 33, 109-124.
- Wilson, D. B. & Shadish, W. R. (2006). On blowing trumpets to the tulips: To prove or not to prove the null hypothesis: Comment on Bösch, Steinkamp & Boller (2006). *Psychological Bulletin*, 132, 524-528.