

Außersinnliche Wahrnehmung unter Kontrolle organisierter Skeptiker

SUITBERT ERTEL¹

Zusammenfassung – Der Ballzieh-Test ist ein Multiple-Choice-Verfahren, das zur Prüfung einer Kommunikationsanomalie entwickelt wurde, die traditionell als „Außersinnliche Wahrnehmung“ (ASW) bezeichnet wird. Im Rahmen der öffentlichen Auslobung eines Preisgeldes von 10.000 Euro wurde das Verfahren von organisierten Skeptikern eingesetzt. In einem von zwei Testläufen wurden signifikante Abweichungen vom Zufall ermittelt, die weitere Forschungsanstrengungen sinnvoll erscheinen lassen. Das Vorgehen der organisierten Skeptiker wird kritisch untersucht und evaluiert.

Schlüsselbegriffe: Parapsychologie – Ballzieh-Test – Skeptiker

Extrasensory perception under control of organized skeptics

Abstract – The ball drawing test, a multiple choice technique developed to probe for a communication anomaly traditionally called ‘Extra Sensory Perception’ (ESP), was adopted by organized skeptics within the scope of a public challenge to win 10,000 euros. In one of two test series deviations from chance were significant, calling for further research. The course of action of the organized skeptics is evaluated and examined critically.

Keywords: parapsychology – ball drawing test – skeptics

Einleitung

Nach acht Jahren Erfahrung mit dem 1997 eingeführten Pingpongball-Test (vgl. Ertel 2004a, 2004b, 2005a), einem Test zur Erfassung von Dispositionen zur außersinnlichen Wahrnehmung (ASW), und nach einem bis jetzt noch unpublizierten erfolgreichen Replikationsversuch mit drei ausgewählten Testpersonen durch skeptische Versuchsleiter am Institut für Grenzgebiete der Psychologie und Hygiene in Freiburg (IGPP) hielt ich es für angezeigt, einen weiteren Replikationsversuch im Rahmen des Preisausschreibens der GWUP („Gesellschaft

1 Prof. Dr. Suitbert Ertel ist emeritierter Professor am Georg-Elias-Müller Institut für Psychologie der Universität Göttingen. Anschrift: Institut für Psychologie, Gosslerstr. 14, D-37073 Göttingen. E-Mail: sertel@uni-goettingen.de.

zur wissenschaftlichen Untersuchung von Parawissenschaften“) zu wagen. Die GWUP bietet seit 2004 alljährlich Bewerbern einen einmaligen Preis von 10.000 € an, die es fertig bringen, unter ihrer Kontrolle eine paramentale Leistung vorzuführen, die sich einer alternativen Erklärung und dem Zufall mit an Sicherheit grenzender Wahrscheinlichkeit entzieht. Das Preisausschreiben der GWUP erfolgt unabhängig vom bekannteren JREF-Preisversprechen von James Randi, der bei Gelingen einer außersinnlichen Leistung eine Million Dollar zu zahlen verspricht, ist diesem aber ähnlich.

Interessiert war ich an einer Teilnahme, obwohl ich damit rechnete, dass ich im Falle eines Scheiterns der vermuteten Tendenz der GWUP zur Abwertung parapsychologischer Forschung in der Öffentlichkeit Veranlassung geben würde. Doch hielt ich es für wahrscheinlich, dass Testpersonen, die am Georg-Elias-Müller-Institut für Psychologie (GEMI) in Göttingen unter meiner eigenen Kontrolle in diesem Test relativ stabile hohe Trefferquoten erzielten, von ihren Fähigkeiten auch unter der Kontrolle von Vertretern der GWUP noch einiges unter Beweis stellen würden. Trotz einer zu erwartenden Minderung der Trefferquoten – bedingt durch erhöhte Anspannung in Gegenwart einer Mehrheit von Beobachtern, denen ein Misserfolg willkommen wäre – hielt ich ein statistisch signifikantes Ergebnis für nicht unwahrscheinlich. Die hohe Trefferquote zu erreichen, die ein Gewinn des GWUP-Preises voraussetzt, war nicht das primäre Ziel. Es würde genügen, wenn die beim Test anwesenden GWUP-Mitglieder der Feststellung einer signifikanten Trefferleistung im Balltest nichts entgegen zu setzen hätten. Der erwartbare Gewinn an Aufmerksamkeit, der bei Mainstream-Forschern gegenüber dem Paranormalen geweckt werden könnte, wenn organisierte Skeptiker die statistische Nullhypothese abzulehnen haben, rechtfertigte in meinen Augen das Risiko.

Um festzustellen, ob das Verhalten der GWUP-Versuchsleiter gegenüber den Testpersonen hinreichend entgegenkommend sein würde, nahm ich mit Erlaubnis der GWUP-Verantwortlichen an einem Tag im Jahre 2004 als Beobachter an ihren Prüfungen eines Wünschelrutengängers, eines Pendlers und eines „Wassermagnetisierers“ teil. Die Versuche finden am Biozentrum der Universität Würzburg in einem Praktikumsraum statt. Am Versuchsleiterverhalten der Hauptverantwortlichen, PD Dr. Rainer Wolf, Dr. Martin Mahner und der übrigen mitwirkenden GWUP-Mitglieder war nichts zu bemängeln (Ertel 2005b), so dass ich mich für August 2005 zur Vorführung des Balltests mit einer ausgewählten Testperson anmeldete.

Da die erzielte Trefferquote im ersten Anlauf statistisch nur mäßig signifikant war, meldete ich mich nach weiteren Testläufen in Göttingen 2006 noch einmal zu Versuchen unter GWUP-Kontrolle in Würzburg an, die von zwei anderen ausgewählten Göttinger Teilnehmerinnen bestritten wurden. Im Folgenden wird über beide Würzburg-Versuche berichtet. Diese sind indessen in der Testhistorie der Versuchsteilnehmerinnen ein jeweils nur kleiner Ausschnitt. Über die Tests, die vor und nach den Würzburger Versuchen unter meiner Kontrolle bzw. bei den Teilnehmerinnen zuhause ohne Kontrolle stattfanden, wird der Vollständigkeit halber ebenfalls berichtet, ebenso wie über die Versuche, die 2005/2006 in Göttingen zur

Vorbereitung des Durchgangs von 2006 in Würzburg durchgeführt wurden. Die Frage, inwieweit die Göttinger Versuchsergebnisse als Nachweis für die Existenz von Psi-Effekten betrachtet werden dürfen, steht nicht im Vordergrund. Das Hauptziel des Berichts ist eine Demonstration der Versuchsbedingungen, die vonseiten der GWUP gefordert und durchgesetzt wurden, und deren Problematisierung.

Projekt Würzburg 2005

Als Versuchsperson lud ich Silke ein,² eine Psychologie-Studierende im 8. Semester, die ich vier Jahre zuvor mithilfe des Standard-Balltests, der bei Studienanfängern vor Ort routinemäßig durchgeführt wird, als außergewöhnlich erfolgreiche Testperson entdeckt hatte und die inzwischen an einigen Fortsetzungsversuchen teilgenommen hatte. Der letzte Test mit ihr lag allerdings ein Jahr zurück. Zudem hatte sie durch einen Praktikumsaufenthalt im Ausland viel Zeit für ihre Diplomarbeit verloren, so dass sie wegen zeitlicher Engpässe vor Würzburg nicht mehr zu Versuchen am GEMI eingeladen worden war.

Es ist mir nicht gelungen, noch einen anderen Studierenden, Achim, zur Teilnahme am Versuch in Würzburg zu bewegen. Achim hatte in mehreren Testserien herausragende Trefferquoten demonstriert. Er gehörte wie Silke zu den wenigen Probanden, bei denen sich die Trefferquoten, die sie zuhause erreichten, unter meiner Kontrolle erheblich steigerten. Sein Mitwirken in Würzburg erschien auch deshalb wünschenswert, weil sich eine Testperson in Gesellschaft einer zweiten im Allgemeinen wohler fühlt, was einen erwarteten Psi-Einfluss vermutlich begünstigt. Außerdem können sich mehrere Teilnehmer nach einigen monotonen Testdurchläufen einander ablösen, so dass für den jeweils nicht Tätigen Erholungspausen entstehen. Doch Achim sagte nach einem längeren Entscheidungskonflikt seine Teilnahme an der Würzburg-Exkursion und gleichzeitig an allen weiteren Psi-Tests ab. Er ließ mich wissen, dass sich nach möglichem Bekanntwerden seiner Mitwirkung bei parapsychologischen Tests, zumal bei einem Test der GWUP, an dem die Öffentlichkeit interessiert ist, seine Chancen für eine spätere akademische Forscherkarriere, die er für sich erhoffte, vermindern könnten. Obgleich die GWUP die Wahrung seiner Anonymität versprach, blieb er bei seiner Entscheidung.

Versuche mit Silke am GEMI in Göttingen

Das Standard-Testverfahren

Der Pingpongball-Test wird routinemäßig am GEMI in zwei Versionen eingesetzt. Bei An-

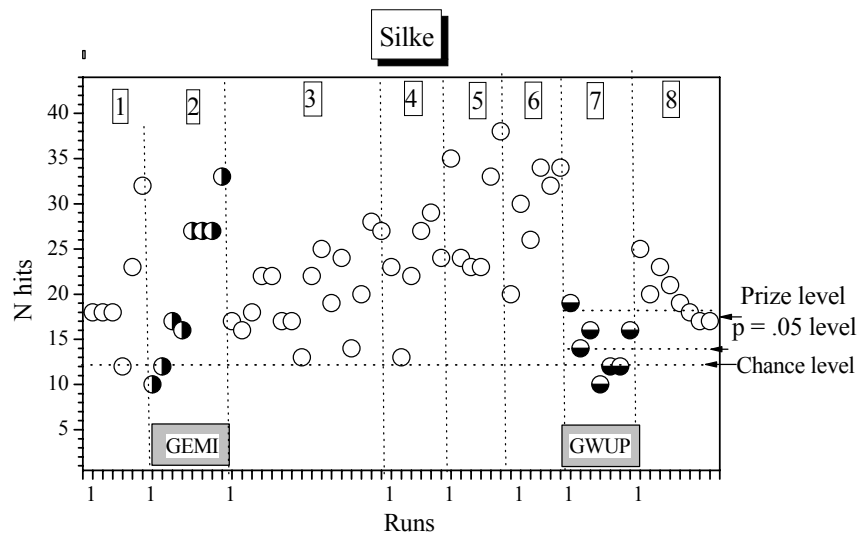
2 Die im Bericht verwendeten Vornamen der Versuchsteilnehmer, deren Zustimmung ich nicht einholte, ob sie mit der Verwendung ihrer Vornamen einverstanden sind, sind willkürlich gewählt.

wendung der Version I wird der Testperson ein undurchsichtiger Beutel ausgehändigt, der kommerziell als Sportbeutel erhältlich ist (zum Aufbewahren von Sportschuhen, Turnzeug etc.). Im Beutel befinden sich 50 Pingpongbälle von Trainingsqualität in Standardgröße. Auf die Bälle werden ringsum auf die Oberfläche Zahlen geschrieben ähnlich wie bei den Bällen des Zahlenlottos. Je zehn Bälle tragen die Zahlen 1, 2, 3, 4 und 5. Die Aufgabe der Testperson ist es, zunächst den Beutelinhalt durcheinander zu bringen: Der vor der Testperson auf dem Tisch liegende flache Beutel wird von ihr hochgenommen und wie ein halbfertiger Pfannekuchen durch Umschwenken gewendet. Sodann soll die Testperson mit geschlossenen Augen einen Ball aus dem Beutel ziehen. Vorher soll sie die Zahl raten, von der sie meint, dass sie sie mit dem nächsten Ball ziehen wird. Wird der Versuch ohne Aufsicht zuhause durchgeführt, soll die Testperson die geratene Zahl in ein Protokollblatt eintragen. Wird der Versuch unter Kontrolle durchgeführt, ist es der Versuchsleiter, der die von der Testperson jeweils geratene Zahl in das Protokollblatt einträgt. Ebenfalls wird die daraufhin gezogene Zahl von der Testperson (im Heimversuch) oder vom Versuchsleiter (im Versuch unter Kontrolle) in das Protokollblatt eingetragen. Der gezogene Ball wird in den Beutel zurückgelegt, bevor dieser für das nächste Trial (Zahlraten mit Ballziehen) gewendet wird. Ein Run (Versuchsdurchgang) besteht aus 60 Trials (Einzelziehungen). Die Göttinger Probanden werden vor ihrem Heimversuch darauf hingewiesen, dass sie, wenn sie auffällige Treffer-Ergebnisse haben sollten, zur Fortsetzung des Versuchs, dann unter Kontrolle, ins Institutslabor gebeten würden. (Die Testversion II wird unten beschrieben).

Silkes Testergebnisse am GEMI

Die Balltest-Historie von Silke vor ihrer Würzburg-Exkursion zeigt der Abschnitt links in Abbildung 1. Die X-Achse repräsentiert eine Zeitreihe der geleisteten Runs, die Y-Achse skaliert die Trefferzahlen pro Run. Silke absolvierte vor dem Würzburg-Test in der Zeit zwischen 2001 und 2006 sechs Versuchsserien. Die Serien 1, 4, 5 und 6 bestanden aus je sechs Runs, die übrigen vier aus je acht Runs. Da ein Run aus 60 Trials besteht, beträgt bei fünf Wahlalternativen die mittlere Zufallserwartung in einem Run 12 Treffer. Im ersten Heimtest (Serie 1, durchgeführt 2001) ergab sich aus Silkes Testprotokoll eine weit überzufällige Anzahl von Treffern (121 Treffer bei 360 Trials, erwartet 72 Treffer, das ergibt $Z = 6,4$, $p = 10^{-10}$).

Ich bat Silke zur Fortsetzung des Tests unter meiner Kontrolle. Unter Kontrolle (Serie 2) lagen ihre Trefferzahlen der ersten zwei Runs bei der Zufallserwartung. Doch nahmen im Laufe der Kontrollsitzen bis zum achten Run die Treffer rasant und kontinuierlich zu und übertrafen sogar im Mittel ihre Heimtest-Leistung (169 Treffer bei 480 Trials, erwartet 96 Treffer, $Z = 8,3$, $p = 10^{-15}$).



**Abbildung 1: Trefferzahlen in den Balltestserien von Teilnehmerin Silke.
Serien unter Kontrolle sind durch partielle Kreisfüllungen markiert.**

Silke absolvierte sodann weitere Heimtests. In Serie 3 (Heimtest), durchgeführt 2002, hatte sie die Zahlen 1, 2, 3, 4, 5, 1, 2, 3, 4, 5 usw. in dieser Folge zu ziehen, die Freiheit der Zahlenwahl war ihr jetzt genommen. Ergebnis: 321 Treffer bei 960 Trials, erwartet 192 Treffer, $Z = 10,4$, $p = 10^{-25}$. In Serie 4 (Heimtest), ebenfalls aus 2002, war die Ziehfolge 5, 4, 3, 2, 1, 5, 4, 3, 2, 1 vorgegeben. Ergebnis: 138 Treffer bei 360 Trials, erwartet 72 Treffer, $Z = 8,6$, $p = 10^{-17}$. Interessanterweise war Silkes Trefferleistung unter den beiden Bedingungen der Zielvorgabe im Ganzen nicht geringer als bei freier Zahlenwahl. Die Treffer im Durchschnitt pro Run, ohne Zielvorgabe, waren für Serie 1 und 2: 20,1 bzw. 21,1, mit Zielvorgabe für Serie 3 und 4: 20,0 bzw. 23,0.

Für Serie 5 (Heimtest), wieder mit freier Zahlenwahl, durchgeführt in 2003, waren die Zahlen auf die Balloberflächen farbig geschrieben worden. Zum Beispiel war die 1 immer mit rotem Stift geschrieben worden, die 2 immer mit grünem usw. Silke durfte bei ihrer Rate-Aufgabe die Farben ignorieren. Die Unterschiede zwischen den Zahlen, die schon numerisch als Werte und graphisch als Zeichen vorliegen, wurden durch farbliche Verschiedenheit lediglich visuell verstärkt, was dem ASW-Effekt zugute kommen sollte. Die Trefferquote Silkes lag tatsächlich sehr hoch (176 Treffer bei 360 Trials, erwartet 72 Treffer, $Z = 13,6$, $p = 10^{-41}$), be-

deutend höher als bei der Darbietung der Zahlen in monoton gleicher Farbe, wie es in der Standardbedingung (Serie 1 und 2) praktiziert wird. In den Serien 1 und 2 waren die Trefferhäufigkeiten pro Run 20,1 bzw. 21,1, in Serie 5 nun 29,3. Der Unterschied ist sehr signifikant ($Z = -4,2$, $p = .00001$). Sollte ein Einfluss mit positivem Vorzeichen durch farbliche Differenzierung der Multiple Choice-Ziele eines ASW-Tests generell regelmäßig auftreten, könnte man auf diesem Wege in Zukunft die Bedingungen in diesem und vielleicht auch in anderen ASW-Tests (z. B. beim Zener-Symbol-Test) routinemäßig verbessern.

In Testserie 6 (Heimtest), durchgeführt 2004, wurden wieder Standardbälle mit nur schwarz geschriebenen Zahlen verwendet. Indessen waren über die Balloberflächen zwischen den Zahlen entweder rote oder grüne Punkte verteilt worden. Die Testperson hatte neben den Zahlen (*Mean Chance Expectancy* MCE = 20%) auch die Farben zu raten (MCE = 50%), für Volltreffer (Zahl und Farbe richtig) war MCE = 10%. Diese Testanordnung ist im Göttinger Gesamtprojekt als Standardverfahren II eingesetzt worden. In Abbildung 1 sind nur Silkes Zahlentreffer eingetragen, die hier primär interessieren. Sie erreichen wieder eine beachtliche Höhe (176 Treffer bei 360 Trials, erwartet 72 Treffer, $Z = 13,6$, $p = 10^{-41}$).

Die Farbtrefferquoten lagen ebenfalls, doch weniger extrem, über der Zufallserwartung ($Z = 4,6$ bei Farben gegenüber $Z = 13,6$ bei den Zahlen). Man möchte gern wissen, ob die relative Minderleistung beim Farbenraten durch die für Psi-Effekte generell ungünstigere MCE von 50% bedingt ist (50% ist für Psi-Effekte die bei Multiple Choice-Aufgaben ungünstigste Bedingung, siehe Rosenthal & Rubin 1989) oder ob bei Silke vielleicht zusätzlich eine generelle Psi-Präferenz für numerisches Multiple Choice-Material besteht.

Für Letzteres spricht das Ergebnis einer Ergänzungsserie zu Serie 5, das in Abbildung 1 nicht berücksichtigt wird. Bei Verwendung der farbig geschriebenen Zahlen wurde Silke auch gebeten, nur die Farben zu raten und dabei zu ignorieren, dass mit den Farben Zahlen geschrieben worden waren. Sie erzielte mit 360 Trials 148 Treffer beim *Farbenraten*, während sie mit demselben Versuchsmaterial in Serie 5 beim *Zahlenraten* 176 Treffer erzielt hatte (erwartet in beiden Fällen 72 Treffer). Der Proportionsunterschied ist signifikant ($Z = 2,10$, $p = 0,018$, einseitig).

Zur Komplettierung der Information für Serie 6 sei noch ergänzt: In den sechs Runs dieser Serie mit insgesamt 360 Trials wurden 116 Volltreffer gezählt (MCE = 10%, Erwartung: 36 Treffer, $Z = 1,3$, $p = 10^{-80}$).

Versuch mit Silke in Würzburg (2005)

Testbedingungen

Die Testbedingungen in Würzburg wichen von den Göttinger Bedingungen erheblich ab, da der Hauptverantwortliche der GWUP, Rainer Wolf, mit dem Standardverfahren nicht einverstanden war. Zum Teil wurde in langwieriger Email-Korrespondenz hierzu meine Zustim-

mung verlangt, zum größeren Teil wurden die Bedingungen vor Ort ad hoc festgelegt. Der Göttinger Sportbeutel wurde nicht akzeptiert. Stattdessen wurde ein Stoffsack ähnlich einer Miniatur-Hängematte in einen Kasten aus Karton eingehängt und in ihn hinein wurden die Pingpongbälle gegeben. Die Bälle waren durch ein am Kasten seitlich angebrachtes Loch erreichbar. Die Testperson, vor dem Kasten sitzend, war durch einen Sichtschirm den Blicken der beiden Versuchsleiter entzogen – ein GWUP-Mitglied und ich hatten diese Funktion.

Die Aufgabe der Testperson war es, wie im Standardversuch, eine Zahl zu raten, doch durfte sie diese vorerst nicht mitteilen, sondern musste sie auf einen Zettel schreiben. Nach dem Raten und Aufschreiben der Zahl griff sie mit ihrer rechten Hand – sie hatte einen dünnen Baumwollhandschuh zu tragen – durch das seitliche Loch im Kasten in die Hängematte, zog einen Ball und hatte diesen durch ein Loch an der Rückwand des Kastens – von ihr aus gesehen 'Rückwand' – durch ein Loch in eine Röhre fallen zu lassen, durch die der Ball den beiden Versuchsleitern, die hinter dem Sichtschutz saßen – von ihr aus gesehen 'dahinter' –, zugeleitet wurde. Diese lasen vom Ball die gezogene Zahl ab und notierten sie.

Das Schreiben der geratenen Zahl auf einen Zettel, das Ziehen eines Balls, der durch die Rückwand den Versuchsleitern zugeleitet wurde, und das Notieren der gezogenen Zahl durch die Versuchsleiter wiederholte sich für fünf Trials. Daraufhin händigte die Testperson über den Sichtschutz hinweg den Versuchsleitern ihren Zettel mit den fünf geratenen Zahlen aus. Diese schrieben die geratenen Zahlen neben die jeweils gezogenen Zahlen in ihr Protokollblatt. Sie durften vorerst weder die geratenen noch die gezogenen Zahlen laut vorlesen.

Ein Feedback war erst gestattet, nachdem jeweils fünf weitere Zahlen geraten, gezogen und notiert waren. Dann erst bekam die Testperson eine Mitteilung darüber, wie viele Treffer sie bei den fünf Trials gehabt hatte, die den jeweils vorletzten Fünferblock bildeten (eine verzögerte Rückmeldung von Treffer-Zwischensummen).

Die Bälle des vorletzten Fünferblocks wurden alsdann vom GWUP-Versuchsleiter in den Behälter im Kasten zurückbefördert und von ihm mit den dort verbliebenen Bällen vermischt. Ein Berühren der Bälle war mir als Versuchsleiter nicht gestattet.

Die Versuchsanordnung war von der GWUP gegenüber der Göttinger Standardbedingung also beträchtlich verfremdet worden. Die Göttinger Bälle wurden nicht akzeptiert, nur GWUP-Bälle durften ins Spiel kommen. Der Göttinger Ballbehälter wurde abgelehnt, es musste ein GWUP-Behälter her. Dieser wurde in einen Karton der GWUP eingelassen. Zwei Sichtschutz-Vorrichtungen, die ich aus Göttingen mitgebracht hatte (ich hatte der Sichtschutz-Idee zugestimmt), wurden nicht akzeptiert, weil noch die Haare der hinter dem Göttinger Sichtschutz Sitzenden zu sehen waren. Stattdessen wurden höhere Pappwände herbeigeschafft, so dass eine Art Kabine entstand. Dünne Handschuhe für Münzensammler, die ich mitgebracht hatte (die Handschuhbedingung hatte ich akzeptiert), wurden ignoriert, GWUP-Handschuhe wurden verwendet. Die in Göttingen übliche Kommunikation zwischen Testperson und Versuchsleiter über geratene und gezogene Zahlen war beseitigt und durch die ver-

spätete Mitteilung der jeweils vorletzten Treffersumme ersetzt worden.

In einer ausgiebigen Email-Korrespondenz mit Rainer Wolf begründete dieser die Maßnahmen wie folgt: Es sei möglich, dass eine Testperson, wenn sie einen Ball in der Hand hält und erfährt, welche Zahl auf ihr steht, diesen Ball markiert, so dass sie die betreffende Zahl anhand dieser Markierung in der Folge wieder finden könne. Ich hatte Dr. Wolf wiederholt gebeten, die Möglichkeit eines Markierens der Bälle durch hypothetische Beispiele einer konkreten Ausführung plausibel werden zu lassen. Wolfs Antworten auf meine Fragen in dieser Richtung seien wörtlich wiedergegeben:

Mein Einwand: Die Testpersonen tragen doch Handschuhe und können Markierungen nicht wahrnehmen. *Antwort* (5.2.2005): „mit hochempfindlichen Fingerspitzen kann man die Zahlen vielleicht doch anhand der Rauigkeit der Oberfläche ein klein wenig erfühlen“. *Antwort* bei anderer Gelegenheit: „Eine – bewusst oder auch unbewusst! – eingesetzte Trickmöglichkeit könnte darin bestehen, dass man taktile Unterschiede zwischen den Bällen herausfindet: z.B. in ihrer Elastizität.“

Mein Einwand: Da Unterschiede in der Elastizität von Tischtennisbällen nicht erst beim Betasten der Bälle erkannt werden, sondern sich schon längst auf die Leistung von Tischtennispielern ausgewirkt haben müssten, wären nach Rückmeldungen an die Industrie ein solches Abweichen von der Norm, wenn es denn trotz maschineller Herstellung überhaupt vorkommen kann, die Mängel längst abgeschafft worden. Selbst wenn sie nicht abgeschafft worden wären, ist es doch unsinnig zu denken, dass Unterschiede zwischen den Bällen hinsichtlich Elastizität mit den von uns aufgeschriebenen Zahlen systematisch korrelieren, denn nur bei einer Korrelation zwischen z.B. drei Elastizitätsgraden Nr. I, II, III und unseren aufgeschriebenen Zahlen 4, 1, 5 z.B. könnte der Testteilnehmer daraus einen Vorteil ziehen. *Antwort* (2.5.2005): „Selbst wenn alle Bälle aus der selben Charge eines einzigen Herstellers stammen (was ich als selbstverständlich voraussetze), könnte der Proband (beim Ziehen und vor allem beim Zurücklegen) die Bälle je nach Zahl, die er ja dank des feedback kennt, unterschiedlich fest drücken und damit vielleicht taktil unterscheidbar machen. Ich kann mir durchaus vorstellen, dass man dies sogar unbewusst tun könnte, ohne dabei einen Trick im Sinn zu haben.“

Mein Einwand: Solange keine genauere Information über die Möglichkeiten von Markierungen von Pingpongballen gegeben werden, die man blind ertasten kann, kann dieser Einwand keine Anerkennung einer Gültigkeit erwarten. *Antwort* (10.8.2005): „Wenn – wie auch immer – der Proband erfährt, welche Zahl er gerade gezogen hat, könnte er diesen Ball leichter wiederfinden, wenn er ihn zuvor, sozusagen auf Verdacht hin, auf irgendeine Weise markiert hat, die wir als Versuchsleiter nicht bemerken. Randi kennt solche Methoden, und er würde es nicht akzeptieren, wenn wir bei unserem Vortest solche möglichen sensory leaks akzeptierten.“

Mein Einwand: Solange diese phantastische Leistung eines heimlichen Ballmarkierens Ih-

rerseits nicht demonstriert wurde, darf ich daran ebenso stark zweifeln, wie Skeptiker ohne Demonstration an der Existenz von Psi-Effekten zweifeln. *Antwort* (16.8.2005): „Es gibt viele Möglichkeiten, Bälle mit Tricktechniken zu markieren. Einige davon zu demonstrieren, ist überflüssig, weil es viele andere gibt.“ (Ich erkenne die Logik nicht. Mit dem Argument eines Staatsanwalts, es gebe viele Tatbeweise, kann er doch nicht die Demonstration wenigstens eines Beweises für überflüssig erklären).

Mein Einwand: Man kann nach Abschluss der Testserie die Bälle zusammen sorgfältig visuell und taktil ohne Handschuhe inspizieren. Wir könnten bei Feststellung einer Druck- oder sonstigen Spur den ganzen Versuch für ungültig erklären. *Antwort* (16.8.2005): „Und ich bin sicher: Es gibt viele andere Trickmöglichkeiten! Mit anderen Worten: Hätte ich Interesse an einem positiven Ausgang des Test, würden mir als Proband viele andere Tricks einfallen“.

Mein Einwand: Man könnte mehrere gleiche Beutel aus dem Göttinger Testmaterial verwenden. Nach jedem Zug würde die Testperson den gerade benutzten Beutel mit dem gezogenen Ball dem am gleichen Tisch gegenüber sitzenden GWUP-Versuchsleiter reichen. Dieser kann den gezogenen Ball in den Beutel zurücklegen, mit den übrigen vermischen und den Beutel unter dem Tisch verschwinden lassen, von wo er für den nächsten Trial einen jeweils anderen Beutel hervorzieht. *Antwort* (29.4.2005): „Für unseren PSI-Test können wir Ihren interessanten Kugelzieh-Versuch nur dann akzeptieren, wenn die oben genannten Bedingungen realisiert werden.“

Mein Einwand: Man könnte eine größere Menge Bälle verwenden und für die Dauer eines Runs die gezogenen Bälle eine Zeitlang nicht in den Beutel zurücklegen. Dass die Testperson Vorteile daraus zieht, dass sie sich merkt, wie oft die 1, die 2 usw. in einem Run bislang gezogen wurde und die weniger häufig gezogenen Zahlen rät, weil sie etwas mehr Trefferchancen haben, ist unwahrscheinlich. Zudem kann man ganz verhindern, dass das Nichtzurücklegen der Bälle irgendeinen Vorteil für sie bringt, indem man sie die Zahlen nicht raten lässt, sondern ihr die zu ziehenden Zahlen vorgibt, z.B. 1, 2, 3, 4, 5, 1, 2, 3, 4, 5 usw. *Antwort* (2.5.2005): „Die Möglichkeit einer taktil spürbaren Markierung einzelner Bälle bleibt bestehen“.

Rainer Wolf wollte unter allen Umständen verhindern, dass die Testperson eine Rückmeldung darüber erhält, welche Zahlen sie jeweils gezogen hat, um ihr die Möglichkeit zu nehmen, die von mir nicht vorstellbaren und auch von ihm selbst nicht konkretisierten Markierungen an den Bällen vorzunehmen. Er ging davon aus, dass die Testperson vor den Augen von zwei Versuchsleitern, die ihr gegenüber sitzen, auf den Bällen, die sie nach dem Ziehen für einen Augenblick in der Hand hält, unbemerkt Markierungen anbringen kann, die sie in die Lage versetzen, beim späteren Ziehen Bälle z.B. mit einer 1-Markierung von Bällen mit einer 2-Markierung usw. zu unterscheiden. Der Sichtschutz, den er schließlich forderte und bis zur Verdeckung der Haarspitzen durchsetzte, wurde damit begründet, dass vor allem ich als Versuchsleiter, der die Zahlen auf den gezogenen Bällen protokolliert, nicht nur bewusst (bei einer gezogenen 5 könne ich vielleicht fünfmal husten), sondern durch „Körpersprache“ auch „unbewusst“ der Testperson Informationen über die numerischen Werte der gezogenen

Bälle zukommen lassen könne. Auch konnte mein Einwand an Wolfs Entscheidung über den Sichtschutz nichts ändern, dass doch die Testperson Informationen durch Körpersprache selbst dann, wenn sie vorkämen und von ihr richtig dekodiert würden, nichts anfangen kann, da sie ja darüber erst Rückmeldungen erhalten würde, nachdem sie den jeweils gezogenen Ball aus ihrer Kabine hinaus befördert hat und an ihm also keine Markierungen mehr anbringen kann.

Mein eigenes Bestreben war es, psi-hemmende Einfluss-Faktoren abzuwenden, die mit der Verkünstlichung der Testsituation durch das Aufstellen von Barrieren für die Testperson und für mich als Versuchsleiter entstehen. Auch wollte ich den zu befürchtenden psi-abträglichen Experimentatoren-Effekt, der sich bei asymmetrischer Rollendominanz zugunsten der GWUP-Vertreter ergeben konnte, so weit wie möglich eindämmen. Dies ist mir kaum gelungen. Ich war schon gegen Ende der zähen Vorverhandlungen mit Rainer Wolf über die Testbedingungen geneigt gewesen, den Plan eines Balltests unter GWUP-Kontrolle aufzugeben, das ließ ich in der Korrespondenz mit ihm anklingen. Doch angesichts des immer noch erhofften Erfolgs mit dem Balltest auch mit den Abwandlungen des Designs hatte ich am Ende trotz Befürchtungen dem zugestimmt, was dann testmethodisch realisiert wurde. Nur die extremsten Deformationen des Göttinger Standard-Verfahrens konnte ich verhindern.

Nicht verhindern konnte ich eine befindlichkeitsstörende Bedingung, die durch folgendes Vorkommnis hervorgerufen wurde: Während der anfänglichen Beschäftigung mit den Versuchsaufbauten in Würzburg wurde uns von einem anwesenden Anwalt der GWUP eröffnet, dass wir noch einen Vertrag zu unterzeichnen hätten. In diesem sollte in erster Linie der Anspruch des Preises von 10.000 € geregelt werden, der von der GWUP zu zahlen wäre, wenn Silke das festgelegte und von mir akzeptierte Treffersoll erreichen würde. Da ich als Bewerber um den Preis mit der GWUP verhandelt hatte, war ich der Meinung gewesen, dass für den Fall eines Erfolgs in Würzburg der Preis mir als Bewerber ausbezahlt wäre. Den potentiellen Zuwachs an finanziellen Mitteln wollte ich nicht nur für studentische Hilfskräfte und Testpersonen zum Ausbau der Göttinger parapsychologischen Forschung verwenden, die bislang keinerlei finanzielle Unterstützung erhalten hatte. Ich wollte auch Silke als erfolgreiche Testperson angemessen belohnen, was ich ihr angekündigt hatte, ohne schon eine genaue Summe zu nennen. Durch den Vertragstext nun wurde indessen auch Silke als potentielle Preisgewinnerin juristisch eingesetzt, ohne dass der Text eine genaue Aufteilung festlegte. Über diese sollten wir uns, so der Anwalt, ad hoc einig werden.

Dadurch entstand eine von uns völlig unerwartete, potentiell dissonanz-trächtige Situation. Diese war von der Gegenseite vielleicht mit einkalkuliert worden. Denn man hatte mir gegenüber, der ich mit Martin Mahner und bis zum Vortage mit Rainer Wolf über den Versuch korrespondiert hatte, einen Vertragstext nicht erwähnt. Aufschlussreich war ein Satz im Vertrag, mit dem wir durch Unterschrift bestätigen sollten, dass wir genügend Zeit gehabt hätten, den Vertragstext zu lesen. Unnötig wäre es ja, sich dies eigens bestätigen zu lassen, wenn der Einwand eines Zeitdrucks nicht nahe gelegen hätte. Missklang erzeugte nicht etwa die Auftei-

lungsbedingung beim Preisgewinn überhaupt, sondern die Tatsache, dass mir diese ohne Ankündigung plötzlich auferlegt wurde. Das Gefühl des Manipuliertwerdens setzte sich bei mir im anschließenden Testverlauf und darüber hinaus lange fort.

Am Kommunikationsstil der GWUP-Vertreter war indessen nichts auszusetzen. In der vorausgehenden Korrespondenz und vor Ort beflößigten sich die Beteiligten mir und der Testperson Silke gegenüber einer Freundlichkeit, an der sich jeder ein Beispiel nehmen kann, der sich gegenüber einem Meinungsgegner unter allen Umständen durchsetzen will.

Versuchsergebnisse

In Testserie 7 mit sieben Runs (420 Trials) erzielte Silke 99 Treffer, erwartet waren 84 Treffer ($Z = 1,77$, $p = 0,038$, siehe Abbildung 1).³ Ihre Trefferleistung ist somit statistisch signifikant. Zwar konnte sie nicht 122 Treffer vorweisen, damit hätte sie die Preiszahlung der GWUP erreicht. Auch liegt sie gegenüber ihrer eigenen Göttinger Leistung, die sie unter Kontrolle erreichte, weit zurück (Serie 2, dort 8 Runs, auf 7 Würzburg-Runs umgerechnet hatte sie dort ein Trefferäquivalent von 152 Treffern). Doch hat sie mit bescheidenen 99 Treffern zumindest das *primäre* Ziel eines statistisch signifikanten Trefferüberhangs erreicht, und dies trotz der zahlreichen psi-hemmenden Bedingungen, die von der GWUP durchgesetzt worden waren.

Hinzu kommt ein die Psi-Hypothese stützender uneingeplanter Zusatzbefund. Zwei Studenten, Thomas S. und Michel K., hatten sich im Hintergrund des Praktikumsraumes, angefangen mit Run 3 bis Run 7 (also über 5 Runs), am Zahlenraten beteiligt. Wenn Silke signalisierte, dass sie die jeweils nächste zu ziehende Zahl riet und diese stumm notierte, notierten die beiden Studenten im Hintergrund ebenfalls, voneinander unabhängig, jeder eine Zahl, von der sie dachten, dass sie Silke anschließend ziehen würde. Bei den je 300 Voraussagen hatte Thomas 71 und Michel 70 Treffer, erwartet wurden 60 Treffer. Beide Ergebnisse zusammen ergeben bei 600 Trials 141 Treffer mit einer Erwartung von 120 Treffern, ein ebenfalls signifikantes Ergebnis ($Z = 2,09$, $p = 0,018$). Silkes Ergebnis und das der Studenten darf man zusammenfassen und erhält als Summen 1020 Trials mit 240 Treffern (erwartet 204 Treffer), was mit $Z = 2,78$ und $p = 0,002$ sehr signifikant ist.

Die Teilnahme der Studenten geschah heimlich. Ich hätte dem nicht zugestimmt, da es sein kann, dass sich mehrere Psi-Prozesse, die bei verschiedenen Personen im gleichen Raum zur gleichen Zeit ablaufen, in Wechselwirkung gegenseitig stören. Tatsächlich hatte Silke in den fünf Runs während der Beteiligung der beiden Studenten nur 66 Treffer, weniger als die Studenten. Hätte sie das Trefferniveau ihrer ersten beiden Runs, als sich die Studenten noch nicht beteiligten, beibehalten, dann hätte sie 115 Treffer erzielt. Hätte sie ähnlich wie in einigen vorhergehenden Göttinger Serien, zumal in Serie 2 (unter Kontrolle) ihre Trefferleistung

3 Alle p-Werte stammen von einseitigen Tests, da bei einer Zufallsabweichung ein Trefferüberhang erwartet wird.

im Laufe der Serie gesteigert – durch die ratenden und vielleicht „psi-abzapfenden“ Studenten wurde dies möglicherweise verhindert –, wäre von ihr u. U. sogar das Preislevel von 122 Treffern erreichbar gewesen. Mir ist nicht bekannt, ob Rainer Wolf und Martin Mahner vom Mitraten der Studenten, die der GWUP nahe standen, wussten und dies uns vorenthielten. Zugute halten muss man ihnen, dass sie dies und die Ergebnisse uns nachträglich bekannt gaben.

Nachuntersuchungen in Göttingen

Die nur mäßig hohe Trefferquote Silkes in Würzburg könnte alternativ erklärt werden als Ergebnis eines Würzburg-unabhängigen generellen Abfalls ihrer Psi-Fähigkeit (Decline). Die letzte Balltest-Serie hatte sie ein Jahr vor Würzburg absolviert. Um sicher zu gehen, bat ich Silke nicht lange nach der Würzburg-Exkursion, den Standard-Balltests noch einmal daheim durchzuführen (wie in ihrer Serie 1).

Das Ergebnis ist in Abbildung 1 als Serie 8 wiedergegeben. In acht Runs mit zusammen 480 Trials erzielte sie 160 Treffer (erwartet 96), $Z = 7,25$, $p = 10^{-12}$. Silke hätte mit diesem Ergebnis, wenn es in Würzburg erzielt worden wäre, die Preisanforderung der GWUP weit überschritten. Die Trefferquote entspricht etwa der von Silkes Serien 1 und 3. Einen Decline-Effekt kann man nicht erkennen.

Diskussion

Die wichtigste Erkenntnis aus den Ergebnissen des Projekts Würzburg 2005 ist, dass es möglich ist, mit einer geeigneten Probandin trotz psi-dämpfender Umstände, die durch weitreichende und zum Teil unverständliche Eingriffe der Skeptiker in die Versuchsanordnung des Balltests hervorgerufen werden, einen zumindest schwach signifikanten Treffer-Effekt zu erzielen. Bei einer Fortsetzung der Versuche unter der Kontrolle der Skeptiker, der wünschenswert ist, sollte versucht werden, ihren Bedingungen, soweit sie auf diesen unverrückbar bestehen, nachzukommen, auch wenn sie unvernünftig erscheinen, allerdings nur solange dadurch die psi-förderliche Natürlichkeit und die Lockerheit der Testsituation nicht allzu sehr beeinträchtigt werden.

Untersuchungen zur Vorbereitung des Projekts Würzburg 2006

Zur Fortsetzung der Experimente zum nächsten Würzburg-Termin (August 2006) trainierte ich Catalina, eine Göttinger Studentin im höheren Semester, als Versuchsleiterin. Sie sollte mit den Studienanfängern des Jahrgangs 2005/6, soweit diese obligatorische Versuchspersonen-Stundennachweise erwerben wollten, einen Balltest durchführen. Am Test nahmen 48 Studierende teil. Im Unterschied zu früheren Tests mit Studienanfängern wurden die Ballzie-

hungen also nicht zuhause, sondern unter Catalinas Kontrolle im GEMI durchgeführt, während ich selbst während des ganzen Semesters verreist war.

Trommelbedingung

In der Vorbereitung zu diesem Test entwarf ich eine neue Testanordnung, die ohne die unnatürliche Kabinenbedingung auskommen, doch von der GWUP noch akzeptabel erscheinen sollte. Anstelle eines Stoffbeutels wurde ein fester zylindrischer Behälter verwendet, eine Art Trommel. Tatsächlich war es ursprünglich ein Papierkorb (Höhe 44 cm), dessen obere Öffnung (Durchmesser 22 cm) geschlossen wurde, so dass dem Boden eine gleichartige Decke gegenüber lag. In Boden und Decke wurden mittig Löcher von 12 cm gesägt, so dass man bei horizontaler Lage der Trommel seine Hände von links und rechts ins Innere hinein stecken konnte. An die Zugangslöcher wurden Ärmel einer Jacke geklebt, die die Sicht ins Trommelinnere verdeckten und die Reibung der hinein zu steckenden Arme an den Schnittstellen der Löcher abschwächten. In die Trommel wurden 100 Tischtennisbälle gekippt (man konnte zum Beladen der Trommel ihre Decke oder den Boden abheben).

Der Testperson wurde überdies eine zylindrische Glasvase gegeben, die sie mit der linken Hand durch die Ärmelöffnung in den Behälter führen konnte. Die Vase diente zum sukzessiven Aufnehmen von fünf Pingpongbällen. Mit der rechten Hand, die wie in Würzburg mit einem dünnen Handschuh bekleidet war, wurde ein Ball ergriffen und in die Vase gesteckt, die mit der linken Hand gehalten wurde, dies alles geschah im Innern der Trommel.

Die zu ziehenden Zahlen wurden in Fünfergruppen vorgegeben. Die Testperson hatte immer unter 10 verschiedenen Zielvorgaben mit den Zahlen 1 bis 5 eine auszuwählen. Zum Beispiel gab es die unregelmäßige Zielvorgabe 3, 2, 5, 1, 4 oder die regelmäßige 5, 4, 3, 2, 1. Immer kamen in einer Vorgabe alle fünf Zahlen vor. Nach Abarbeiten der gewählten Fünfergruppe von Zahlen sollte sie die nächste wählen.

Nach Wahl einer Zielvorgabe hatte die Testperson nacheinander fünf Bälle zu ziehen und in die Vase zu stecken, wobei sie sich vorstellen sollte, die Bälle würden die gewünschten Zahlen beim jeweiligen Ziehen tragen. Wenn fünf Bälle gezogen und in die Vase gesteckt waren, zog die Testperson die Vase aus der Trommel heraus. Erst jetzt war also eine Rückmeldung über ihre Ziehergebnisse gegeben, die dann von der Versuchsleiterin notiert und neben die „Sollzahlen“ der jeweiligen Zielvorgaben notiert wurden. Die Rückmeldung der Zieh-Ergebnisse war also gegenüber der Bedingung Würzburg 2005 erheblich verbessert.

Waren fünf Bälle von der Testperson gezogen worden, kippte Catalina sie aus der Vase in einen Korb. War ein Run von 60 Ziehungen beendet, wurden die Bälle aus dem Korb in die Trommel zurückgekippt. Die Testperson hatte also keine Möglichkeit, Bälle zu markieren. Denn während sie einen Ball in der Hand hielt, konnten weder sie noch die Versuchsleiterin wissen, welche Zahl der Ball trug. Auch war eine Veränderung von Balleigenschaften etwa durch die Wärme der Hand, die einen Ball gehalten hatte, wegen des Zwischenlagerns der gezogenen Bälle ausgeschlossen. Diese Testanordnung wurde mit Rainer Wolf nicht disku-

tiert. Ich hoffte, er würde mit ihr später einverstanden sein. Vorher aber wollte ich noch in Erfahrung bringen, ob mit der neuen Anordnung Psi-Effekte zu erreichen waren, die mit den Effekten, die seit 1998 alljährlich unter Standardbedingungen gewonnen wurden, vergleichbar sind.

Abbildung 2 zeigt auf der X-Achse in chronologischer Folge Testeinheiten von fünf Studenten-Kohorten 1998 bis 2002, die den Standardtest daheim (Beutelbedingung) absolvierten. Auf der Y-Achse wird die Trefferleistung als Effektstärke angegeben ($ES = Z/\sqrt{N_{\text{trials}}}$). Die übrigen zur Beurteilung der Psi-Effekte notwendigen Daten sind in der Tabelle unterhalb der Abbildung abzulesen. Es zeigt sich, dass die Trefferüberhänge der Kohorten 1998-2002 bei unsystematischer Streuung alle hochsignifikant ausfielen. Dagegen liegen die Trefferquoten in der Trommelanordnung von 2006 auf dem Zufallsniveau, die Abweichung vom MCE hat hier sogar eine numerisch negative Richtung.

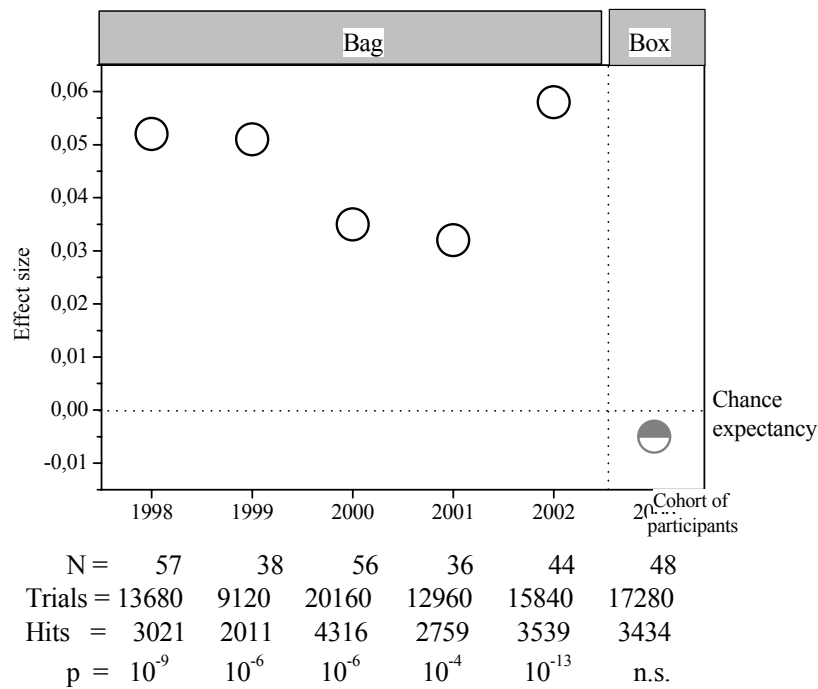


Abbildung 2: Effektgrößen der Trefferabweichungen von der Zufallserwartung für fünf Kohorten von Studienanfängern, die den Balltest mit einem Beutelbehälter für die Bälle (Bag), und einer Kohorte, die einen modifizierten Balltest mit einem trommelartigen Behälter durchführten (Box).

Nur zwei Anzeichen einer denkbaren Psi-Einwirkung ergaben sich für den Test mit der Trommel:

(1) Festgestellt wurde ein hochsignifikantes Defizit von wiederholten Ziehungen der gleichen Zahl in unmittelbarer Folge. Wird zum Beispiel die 4 zweimal hintereinander gezogen, ist dies *eine* Wiederholung, wird sie dreimal hintereinander gezogen, sind dies *zwei* Wiederholungen. Dabei bleibt unberücksichtigt, was die Testperson geraten hat. Erwartet wurden bei 17280 Trials insgesamt 3456 Sofortwiederholungen ($17280/5$), gezählt wurden nur 3277, $Z = -3,4$, $p = 0,007$). Da in den Vorgaben von fünf Zielzahlen aufeinander folgende Zahlen immer wechselten, kann es sein, dass sich nur die Eigenschaft des Wechsels der Zahlen als abstraktes Muster in den Ziehungen der Probanden auf paramentalem Wege durchgesetzt hat. Möglich ist dies, auch ohne dass beim einzelnen Trial die gezogenen Zahlen überzufällig oft treffen, doch diese Frage kann offen bleiben.

(2) Die Treffersummen hatten in der Stichprobe der 48 Teilnehmer eine überdurchschnittlich große Streuung (es resultierte eine flache Kurtosis mit einer nicht unbeträchtlichen Abweichung von $Z = -2,82$, $p = 0,002$). Die Tendenz von drei Teilnehmern zu signifikant erhöhten Trefferzahlen (andeutungswise Psi-hitting) wurde durch die Tendenz von drei anderen Teilnehmern zu signifikant verminderten Trefferzahlen (Psi-Missing) ausgeglichen, so dass in dieser Stichprobe insgesamt keine Treffersumme mit Überhang zustande kam.

Ist der seit 1997 regelmäßig beobachtete Trefferüberhang diesmal ausgeblieben, weil zur Trommelbedingung gewechselt worden war? Man weiß es nicht, denn es kann sein, dass die Kohorte der Studenten von 2006 entweder nur zufällig oder aufgrund eines natürlichen „Generationen-Declines“ an Psi nichts mehr erkennen ließ. Deshalb wurden die Testteilnehmer im darauf folgenden Semester zu einem weiteren Ballzieh-Experiment, wieder unter Kontrolle, gebeten. Als Experimentator fungierte diesmal ich selbst, der ich nach Göttingen zurückgekehrt war.

Diesmal wurde wieder der bewährte Beutelbehälter verwendet, zu raten waren neben fünf Zahlen die zwei Farben rot und grün, die auf den Balloberflächen als Punkte verteilt waren (Standardbedingung II, Silkes Bedingung in Serie 6). An diesem Experiment nahmen 42 der ursprünglich 48 Studierenden teil, in zwei Testsitzungen wurden von jedem Teilnehmer 8 Runs absolviert.

Ergebnis: Ein sehr signifikanter Trefferüberhang ist festzustellen (2133 Treffer bei 20160 Trials, $Z = 2,74$, $p = 0,003$). Das Ergebnis lag minimal unter dem der Kohorte von 2001, es wäre dem von 2001 gleich gekommen, wäre unter den Probandinnen nicht eine gewesen, die eine phantastische Psi-Missing Leistung von nur 6 Volltreffern erbrachte, bei 480 Trials wurden 48 Treffer erwartet). Ein Decline-Effekt ist demnach nicht erkennbar. Die Schlussfolgerung ist erlaubt, dass die in der Untersuchung von Experimentatorin Catalina aufgetretene Minderleistung beim Ballziehen durch den Wechsel von der Beutel- zur Trommelbedingung verursacht worden war.

Letzteres wurde erhärtet durch Silke, die ich bat, den Versuch mit dem Trommelbehälter

unter meiner Kontrolle durchzuführen. Silke blieb bei dieser Bedingung mit 91 Treffern (erwartet bei 480 Trials waren 96 Treffer) unter der Zufallserwartung. Das war die bislang niedrigste Trefferquote in ihrer langen Balltestkarriere.

Picknick-Box-Bedingung

Da die gleichen Versuchsteilnehmer in der Beutelbedingung (unter meiner Kontrolle) bedeutsame Trefferüberhänge, in der Trommelbedingung (unter Catalinas Kontrolle) nur Zufallstreffer zustande brachten, lag es nahe, die Minderleistung der Trommelbedingung zuzuschreiben.⁴ Ich vermutete, dass sie den Probanden zu unbequem und unnatürlich vorkam, für den bald anstehenden Versuch in Würzburg kam diese Anordnung nicht infrage.

In Eile wurde ein anderer, bequemerer Behälter ausfindig gemacht, ein kleiner, bei Draufsicht nahezu quadratischer, freundlich aussehender blauer Picknick-Behälter, der mit Stoffen zum Kühlhalten von Getränken weich ausgepolstert war (Höhe 40 cm, Breite und Länge 22 cm unten, 29 cm oben). Er stand auf einem kaum sichtbaren Trolley-Fahrgestell. Seine Form war leicht konisch, die Ecken waren gerundet. Die Picknick-Box wurde zu einem Pingpongball-Behälter umfunktioniert. Der Testperson wurde Zugang zum Inneren des Behälters ermöglicht, indem an den Boxseiten links und rechts mittig runde Öffnungen von 12 cm Durchmesser eingeschnitten und daran Hemdenärmel angeklebt wurden, so dass sie mit beiden Armen gleichzeitig hineingreifen konnte. Der normale Zugang zum Boxinneren, der durch eine Deckenklappe möglich war, die man mit Reißverschluss öffnen und schließen konnte, konnte zum Hineinkippen von Bällen verwendet werden und während der Balltest-Durchführung geschlossen bleiben. Der Zugang zur Picknick-Box war gegenüber dem Zugang zum Innern der Trommel weit bequemer, die Arme waren nur 25 cm auseinander zu halten, bei der Trommel waren es 44 cm.

Die Idee mit der Auffangvase, in welche im Innern des Behälters nacheinander fünf Bälle hineingesteckt werden, wurde fallen gelassen. Ähnlich wie 2005 in Würzburg sollten die Bälle durch einen Ausgang an der Hinterseite der Box über eine schräg nach unten gerichtete durchsichtige Ablaufröhre hinaus befördert werden. Dies schloss eine Ballmarkierung durch die Testperson aus, da diese ja nicht wissen konnte, welche Zahlen die jeweils von ihr hinaus beförderten Bälle trugen. So konnte die Testperson ihre Ballziehungen mit den Versuchsleitern am gleichen Tisch vornehmen, ohne Sichtschutz, da ein solcher nicht mehr zu begründen war. Die Verwendung vorgegebener Zielzahlen konnte beibehalten werden, da in früheren Versuchen mit erfolgreichen Ballziehern, zum Beispiel mit Silke (siehe Abbildung 1, Serien 3 und 4) kein merklicher Nachteil gegenüber der freien Zahlenwahl erkennbar war.

Wegen der Kürze der Zeit bis zum Würzburg-Termin konnte mit den an diesem Versuch

4 Ein Experimentatoren-Effekt kann theoretisch auch vorliegen, doch das negative Trommel-Ergebnis mit Silke, bei dem ich wie bei den übrigen Göttinger Versuchen mit ihr Experimentator war, spricht dagegen.

sich beteiligenden zwei Probandinnen nur je ein kurzer Vortest am GEMI durchgeführt werden. Das Ergebnis war nicht ermutigend (siehe weiter unten in Abbildung 2) und ließ kaum erkennen, ob der Wechsel von der Trommel zur Picknickbox den erhofften Vorteil bringen würde. Doch reichte die Zeit nicht für eine weitere Bedingungsoptimierung.

Projekt Würzburg 2006

Ein nicht geringes Problem für die Teilnahme am Würzburg-Test 2006 ergab sich bei der Rekrutierung der Testpersonen. Silke, die erfolgreiche Ballzieherin von 2005, wollte ein zweites Mal nicht mitwirken. Ihr hatten, wie sie sagte, die überraschenden Vorfälle dort mit ihrer Intransparenz atmosphärisch nicht behagt.

Als Ersatz für sie gewann ich eine Studentin der Sozialwissenschaft, Veronika, als Testperson, die sich unter den Studienanfängern von 2003 im Balltest zuhause und im anschließenden Kontrollversuch hervorgetan hatte. Veronika sollte eigentlich durch Vanya unterstützt werden, einen hervorragenden 12-jährigen Balltest-Teilnehmer aus Kiew, der sich vor drei Jahren beim Balltest am IGPP in Freiburg vor unabhängigen skeptischen Testleitern bewährt hatte und in Begleitung seiner Mutter der Einladung folgen wollte. Doch ergaben sich für Vanya unvorhergesehene bürokratische Probleme bei der Beschaffung der notwendigen Visa, die seine Teilnahme vereitelten.

So wick ich in letzter Minute auf Katarina aus, einer im Semester zuvor diplomierten Studentin der Psychologie, die sich an den Göttinger Balltests seit 1997 immer wieder mal beteiligt hatte, aber aus Gründen persönlicher Belastung und wegen Umzugs in eine andere Stadt schon lange keine Tests mehr durchgeführt hatte. Sie nahm meine Einladung zum Test nach Würzburg sehr zögerlich an, da sie sich, wie sie in längeren Telefonaten erläuterte, in einer persönlich kritischen Verfassung befand. Leider war ein anderer testbewährter Teilnehmer in den laufenden Semesterferien kurzfristig nicht kontaktierbar.

Die Tests am GEMI in Göttingen

Die Abbildungen 3a und 3b geben die Ergebnisse der beiden Testpersonen wieder, die lange Zeit vor Würzburg am GEMI in Göttingen begonnen hatten. Die X-Koordinate repräsentiert wieder die chronologische Folge von Runs. Ein Run besteht wieder aus 60 Trials. Die Runs werden in zwei übergreifende Testabschnitte (*Bag* und *Home*) und in vier Unterabschnitte (Serien) eingeteilt. Im ersten Testabschnitt wurde der Sportbeutel als Ballbehälter verwendet (*Bag*-Bedingung). Im zweiten Abschnitt befanden sich die Bälle im oben beschriebenen Picknickbehälter (*Box*-Bedingung). Der Beutel-Versuch wurde zunächst unter der Heimbedingung (*Home*, Serie 1, acht Runs) ohne Kontrolle durchgeführt. Darauf folgte der Beutel-Versuch unter Kontrolle mit Experimentator Ertel (Bedingung „laboratory“ oder kurz *Lab*,

Serie 2, mit acht Runs für Veronika, mit sechs Runs für Katarina). Dann wurde zur Box-Bedingung gewechselt, die nur unter Kontrolle durchgeführt wurde, zweimal unter Kontrolle von Ertel in Göttingen (Serie 3).

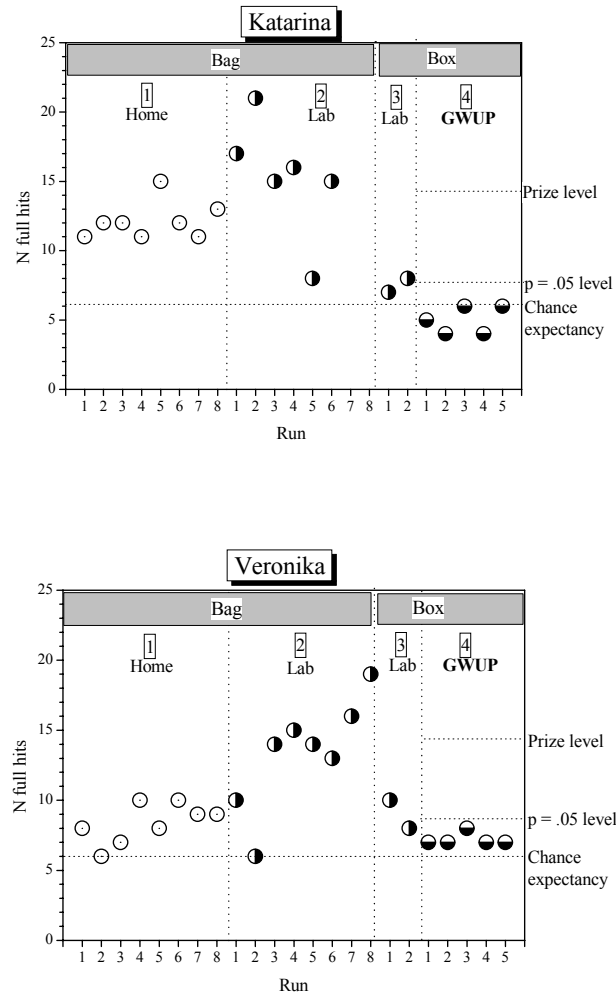


Abbildung 3 (a und b): Trefferzahlen in den Balltestserien von Katarina (3a) und Veronika (3b). Testserien unter Kontrolle sind durch partielle Kreisfüllungen markiert.

In der Beutel-Bedingung wurden Bälle verwendet, auf denen die Zahlen von 1 bis 5 in schwarzer Farbe geschrieben waren, unter die entweder rote oder grüne Punkte verteilt waren (wie in Silkes Serie 6, eine Ausnahme in Katarinas Serie 1 wird unten erläutert). In der Box-Bedingung waren die Zahlen von 1 bis 5 auf die Bälle in entweder roter oder grüner Farbe geschrieben worden.

In der Beutel-Bedingung wurden Zahl und Farbe, die für den einzelnen Zug zu ziehen waren, vor jedem Zug von der Testperson frei gewählt. In der Box-Bedingung waren die Zahlen und Farben für die Ziehungen der Testperson von der Versuchsleitung im Voraus festgelegt worden.

Die Box-Bedingung in Würzburg unterscheidet sich von der Boxbedingung in Göttingen nur dadurch, dass in Würzburg die einmal gezogenen Bälle nicht mehr verwendet wurden. In der Göttinger Box-Bedingung wurden die gezogenen Bälle zwar vom Versuchsleiter zunächst auch nicht zurückgelegt, wie in Würzburg, doch wenn 60 von 120 Bällen gezogen worden waren, wurden sie in die Box zurückgegeben.

Die Y-Koordinate zeigt für die einzelnen Runs die Zahl der *Volltreffer* beim Ballziehen. Als Volltreffer gilt, wenn mit einem Ball die Zielzahl (1, 2, 3, 4 oder 5) und die zugehörige Zielfarbe (rot oder grün) von der Testperson gezogen wurde. Die Zieh-Ergebnisse, bei denen nur die Zahl oder nur die Farbe getroffen wurde, werden in den Abbildungen nicht berücksichtigt.

Die durchgezogene waagerechte Linie zeigt die mittlere Zufallserwartung (MCE) an, die bei 60 Zügen und der Wahrscheinlichkeit von 10% 6 Treffer beträgt.

Für die GWUP-Bedingung wird durch eine kurze horizontale Linie rechts das Trefferniveau angezeigt, das in den fünf Runs einer Testperson in Würzburg zu einem signifikanten Gesamtergebnis führen konnte (7,8 Treffer im Durchschnitt für eine Testperson). Auch wird durch eine Linie angezeigt, bei welchem Trefferniveau der von der GWUP ausgesetzte Preis (10000 €) gewonnen werden konnte (11,6 Treffer im Durchschnitt in den 10 Runs der beiden Testpersonen, also fast das Doppelte der MCE).

GEMI-Ergebnisse Katarinas

Serie 1: Bag-Home

Die Trefferzahlen liegen weit über dem Signifikanzniveau (107 Treffer bei 480 Trials, erwartet wurden 48 Treffer, $Z = 8,9$, $p < 10^{-18}$, $ES = 0,41$), würden aber das von der GWUP festgesetzte Preisniveau, wenn es darum gegangen wäre, noch nicht erreicht haben.

Methodisch zu ergänzen ist, dass die Ergebnisse der acht Runs von Katarina die letzten einer Serie von 96 Runs sind, die sie mit der Aufgabe durchführte, die zu ziehenden Zahlen vorauszusagen (MCE 20%). Ein zusätzliches Voraussagen und Ziehen von Farben war nicht gefragt (MCE der Zahlen-plus-Farb-Treffer ist 10%). Um die Ergebnisse der 20%-Bedingung in die vorliegende Abbildung mit den übrigen Ergebnissen der 10%-Bedingung einzupassen, wurden die Treffersummen Katarinas über die Effekt-Berechnung nach Rosenthal und Rubin

(PI) entsprechend umgerechnet. Die Daten der Heimbedingung stammen aus Versuchen, die mit ihr 1997/98 durchgeführt wurden.

Serie 2: *Bag-Lab*

Die Trefferzahlen liegen noch höher als in der Heimbedingung (92 Treffer bei 360 Trials, erwartet 36 Treffer, $Z = 9,75$, $p = 10^{-22}$, $ES = 0,52$). Dies ist bemerkenswert, weil bei vielen Probanden unter Kontrolle die Trefferzahl meist mehr oder weniger geringer ist als unter der Heimbedingung. Dies war bei einem früheren Ball-Versuch (Nur-Zahlen-Bedingung) auch bei Katarina der Fall (in Abbildung 3 nicht berücksichtigt). Sie hatte jedoch nach anfänglich starkem Treffer-Abfall im Laufe von acht Runs unter anhaltender Kontrolle zu ihrem hohen Heim-Trefferüberhang zurückgefunden.

Zu beachten bleibt noch, dass sich diesmal innerhalb der sechs Runs unter Kontrolle ein Absinken anzudeuten scheint. Zumindest liegt in der Bedingung Bag-Lab-Göttingen eine größere Streuung vor mit einem positiven Ausreißer im zweiten Run und einem negativen Ausreißer im vorletzten Run.

Die tendenziell höhere Trefferquote von Katarina unter der Kontrollbedingung widerspricht der Erwartung eines Decline-Effekts, wonach mit der Zeit eine geringere Trefferquote zu erwarten wäre. Die Bag-Lab-Versuche Katarinas wurden im Februar 2005 durchgeführt, fünf Jahre nach ihrem letzten Bag-Home-Versuch.

Serie 3: *Box-Lab in Göttingen*

Die Trefferzahlen in den beiden Versuchen mit dem Picknick-Behälter, durchgeführt in Göttingen kurz vor der Würzburg-Exkursion im August 2006, ebenfalls unter Kontrolle, liegen weit unter denen, die mit dem Beutelbehälter gewonnen worden waren (15 Treffer bei 120 Trials ergibt $Z = 0,76$, $ES = 0,001$), sie liegen nicht signifikant über der erwarteten Zahl von 12 Treffern. Sehr signifikant ist der Unterschied der Trefferproportionen zwischen Bag-Lab und Box-Lab: $Z = 3,38$, $p = 0,0007$, zweiseitig.

Für die Minderleistung in der Generalprobe vor Würzburg bieten sich zwei Erklärungen an. Zum einen befand sich Katarina nach eigenen Aussagen in einer kritischen Lebenssituation, eine Bedingung, die von ASW-Forschern generell als psi-abträglich erachtet wird. Zum anderen können sich trotz aller Verbesserungsabsicht die Eigenschaften des Picknick-Behälters gegenüber denen des Beutels ungünstig auf Psi-Faktoren ausgewirkt haben. Letzteres ist nahe liegend, weil auch Veronika mit der Picknick-Box geringere Trefferzahlen erzielte als mit dem Beutel-Behälter.

GEMI-Ergebnisse Veronikas

Serie 1: *Bag-Home*

Sieben von acht Trefferzahlen liegen bei Veronika über der Zufallserwartung. Das Gesamter-

gebnis unter dieser Bedingung ist sehr signifikant (67 Treffer bei 480 Trials, erwartet 48 Treffer, $Z = 2,81$, $p = 0,002$), jedoch weniger stark abweichend von der Zufallserwartung ($ES = 0,13$) als bei Katarina ($ES = 0,34$).

Serie 2: *Bag-Lab*

In den beiden ersten Runs unter Kontrolle scheint sich Veronikas mäßige Trefferleistung der Heim-Runs fortzusetzen. Doch mit dem dritten Run machen die Trefferzahlen einen deutlichen Sprung nach oben bis zu dem Niveau, das Katarina unter der gleichen Bedingung erreicht hatte. Insgesamt liegt der Effekt fast auf gleicher Höhe (107 Treffer bei 480 Trials, erwartet 48 Treffer, $Z = 8,9$, $p = 10^{-18}$, $ES = 0,41$). Der Unterschied der Trefferproportionen bei Veronika zwischen Kontroll- und Heimbedingung ist hochsignifikant ($Z = 3,35$, $p = 0,0008$, zweiseitig). Warum auch Veronika unter der Kontrolle bei sonst gleicher Testbedingung einen hohen Trefferanstieg zeigt, ist ebenso rätselhaft wie das gleichartige Ergebnis bei Katarina. Wenn die Anwesenheit eines Experimentators der entscheidende Faktor gewesen wäre, hätte der Anstieg unter dieser neuen Bedingung eigentlich schon mit dem ersten Run erfolgen sollen.

Serie 3: *Box-Lab in Göttingen*

In den beiden Box-Versuchen in Göttingen sinken die Trefferzahlen Veronikas auf ein mäßig hohes Level ab (18 Treffer bei 120 Trials, erwartet 12 Treffer, $Z = 1,67$, $p = 0,05$, $ES = 0,31$). Das Sinken der Trefferzahlen lässt sich nicht sicher erklären, doch liegt es nahe, dieselben Gründe anzunehmen wie die, die zum Absinken der Trefferzahlen unter der Box-Bedingung bei Katarina geführt hatten. Die Hypothese, dass die Box-Bedingung generell eine psi-hemmende bzw. weniger psi-unterstützende Bedingung ist im Vergleich mit der Beutel-Bedingung, erfährt eine erneute Stütze.

Testbedingungen in Würzburg 2006

In der E-Mail-Korrespondenz zur Vorbereitung des Würzburg-Tests 2006 war Rainer Wolf mit dem Picknick-Behälter und mit fast allen anderen Vorschlägen, die ich machte, einverstanden. Er bestand allerdings darauf, dass ein von der Testperson einmal gezogener Ball von ihr nicht ein zweites Mal verwendet werden dürfe. In der Erwartung, am Testtage mit meinen Testpersonen etwa 10 Runs durchführen zu können, besorgte ich 600 Pingpongbälle, schickte sie an Rainer Wolf, der sie mit Zahlen beschriften ließ (eine Beschriftung durch Göttinger Studenten erschien ihm zu unsicher). Fünf Farbstifte legte ich bei, mit denen Wolfs Helfer dann allerdings nicht die farbige Punktbesprenkelung unter schwarz geschriebene Zahlen einfügten, was in Göttingen Usus war (Standardversion II). Sie schrieben die Zahlen in verschiedenen Farben auf die Bälle, damit hatte ich mich einverstanden erklärt. Die Testperson konnte zum Beispiel eine „rote 2“ oder eine „grüne 2“ raten – die Zahl konnte stimmen aber die Farbe nicht, die Farbe konnte stimmen aber die Zahl nicht, und beides konnte stimmen:

ein Volltreffer, oder nicht stimmen, ein trefferloses Ergebnis. Silke hatte unter der genau gleichen Bedingung auch hohe Trefferleistungen erzielt (siehe ihre Serie 6, Abbildung 1). Vereinbarung wurde, dass nur Volltreffer berücksichtigt wurden, sie hatten in Göttinger Versuchen auch die höchste Effektstärke gezeigt. Den aus der Picknickbox entstandenen Ballbehälter ließ ich Rainer Wolf vor dem Testtermin zur Prüfung postalisch zukommen.

Nach Ankunft des Göttinger Versuchsleiters am Biozentrum in Würzburg in Begleitung der Testpersonen Veronika und Katarina wartete auf uns wieder eine Überraschung, die im Vergleich zu der von 2005 aber weniger gravierend war. Wolf hatte am Ballbehälter auszusetzen, dass am angeklebten Eingangsärmel eine Lücke verblieben war. Sie war klein und nur erkennbar, wenn man sie mit gespreizten Daumen und Zeigefinger weitete. Wenn dabei das Licht günstig fiel und man bei verminderter Sichtdistanz aufmerksam ins Innere des Behälters hineinblickte, konnte man Zahlen auf Bällen erkennen. Ein unbemerktes Hineinsehen, ohne die Finger zu Hilfe zu nehmen, wäre unmöglich gewesen. Doch Wolf bestand darauf, die Lücke zu schließen. So überklebte er die linke Öffnung mit weißem Papier. Die Testperson hatte deshalb nur einarmig über den rechten Ärmel Zugang zu den Bällen, ich hatte die Box für einen Zugang links und rechts hergestellt. Zudem war der Anblick der blauen Box durch die weiße Überpflasterung unschön geworden.

Auch hielt Wolf meine durchsichtige Ausgangsöhre für zu unsicher, da, wie er meinte, der Versuchsleiter, der dieser Röhre mit dem Ausgabe-Loch an ihrem oberen Ansatz gegenüber saß, den von der Testperson herausgegebenen Ball mit ihrer jeweiligen Zahl zu Gesicht bekommen würde. Das wäre möglich in dem Augenblick, wenn die Testperson den Ball zum Ausgabeloch hin geführt, aber noch nicht fallen gelassen hatte und ihn für einen Sekundenbruchteil in den Händen hielt. Der Versuchsleiter könne in diesem Augenblick der Testperson durch seine Körpersprache den Zahlenwert des Balles schnell zurückmelden, so dass diese den Ball noch vor seiner Ablieferung markieren könne.

Rainer Wolf hielt sein Argument durch meinen Einwand nicht geschwächt, dass die Bälle nach dem vereinbarten Protokoll nur einmal verwendet werden sollten und somit Markierungen auf ihnen nutzlos wären. Mir erschien die Begründung der Zusatzröhre sinnwidrig, zumal die Versuchsleiterin, die der Röhre gegenüber saß und angeblich Signale durch Körpersprache geben könne, der GWUP nahe stand, so dass man die Zahlen am Boxausgang vor ihr nicht zu verbergen brauchte. Trotzdem schaffte Rainer Wolf eine GWUP-Pappröhre herbei, die er in die durchsichtige Plastikröhre aus Göttingen als Sichtschutz hineinsteckte.

Beide angeblichen Korrekturmaßnahmen Wolfs ließ ich kopfschüttelnd geschehen. Ich hätte gut daran getan, mit derselben Striktheit – in Freundlichkeit eingebettet –, die ich auf der Gegenseite kennen gelernt hatte, diese Eingriffe in die Versuchsanordnung abzuwenden.

Testergebnisse von Katarina in Würzburg

Katarinas Trefferzahlen liegen im Bereich der Zufallserwartung (25 Treffer bei 300 Trials,

erwartet 30 Treffer, $Z = -0,9$, n.s.). Zwei Trefferzahlen liegen auf der MCE-Linie, drei sogar darunter. Das Ergebnis unterscheidet sich nicht signifikant vom Box-Ergebnis Katarinas in der vorausgehenden GEMI-Kontrolluntersuchung.

Katarina hatte in den Testserien seit 1997 in über 104 Runs ausnahmslos Trefferquoten über der Zufallserwartung erzielt. Zur Erklärung der geringen Trefferquote in Würzburg stehen die beiden Einflussmöglichkeiten zur Debatte, die schon für ihre niedrigen Trefferquoten in der Bag-Lab Bedingung in Göttingen besprochen wurden (ungünstige Box als Behälter plus seelische Verfassung der Testperson). Hinzukommen kann ein psi-hemmender Effekt, mit dem man bei einer Umgebung von skeptischen Beobachtern rechnen muss (Experimentator-Effekt).

Testergebnisse von Veronika in Würzburg

Veronika hält in Würzburg in etwa das sehr mäßige Trefferniveau, das sie beim Box-Versuch schon in Göttingen gezeigt hatte. Die Trefferzahlen der fünf Runs liegen zwar alle über der Zufallserwartung, jedoch ohne statistische Signifikanz (36 Treffer bei 300 Trials, erwartet 30 Treffer, $Z = 1,06$, n.s., $ES = 0,06$). Signifikanz würde sich ergeben, wenn Veronika 15 Runs anstelle von nur 5 Runs durchgeführt und das Trefferniveau der ersten fünf Runs beibehalten hätte.

Weitere explorative Auswertungen

In Würzburg ist der Unterschied der beiden Teilnehmerinnen in den Trefferleistungen beachtlich, wenn man den t-Test für unabhängige Stichproben anwendet – der t-Test berücksichtigt nicht nur das Trefferniveau insgesamt, sondern auch die Trefferstreuung über die Runs ($t = 4,49$, $df=8$, $p = 0,002$). Man kann mutmaßen, dass Katarina aufgrund einer besonders unglücklichen Verfassung, von der sie schon Tage zuvor mehrmals sprach, in Würzburg zum Psi-Missing hin tendierte. Veronikas Befindlichkeit hatte offenbar keine negativen Vorzeichen.

Die Würzburger Treffer der beiden Testpersonen für die Zahlen allein (erwartet 60 Treffer) betragen 60 (für Katarina) und 66 (für Veronika), für die Farben allein (erwartet 150) betragen sie 138 (für Katarina) und 160 (für Veronika). Auch für die Zahlen und Farben allein ergeben sich also Treffervorteile für Veronika. Die Beobachtung ist im vorliegenden Zusammenhang nicht zentral. Sie wird erwähnt, weil generell in den Göttinger Untersuchungen zu außersinnlichen Fähigkeiten interindividuelle Unterschiede zu den am stärksten bestätigten Beobachtungen überhaupt gehören. Bei Psi-Testergebnissen entsteht nicht nur die Frage, ob zufallsabweichende Treffer vorkommen, sondern auch, ob zufallsabweichende Unterschiede in den Testergebnissen zwischen Testpersonen und innerhalb von Zeitreihen ein und derselben Testperson vorkommen.

Nachuntersuchungen in Göttingen

Nur mit Veronika konnten nach Rückkehr von Würzburg weitere Testsitzungen stattfinden. Dabei interessierte die Frage, ob sie den nur geringen Trefferüberhang, den sie unter der Picknickbox-Bedingung im Vorversuch am GEMI und anschließend in Würzburg gezeigt hatte, unter Kontrollbedingung am GEMI fortsetzen würde. Die Würzburg-Bedingungen wurden also beibehalten.

Dabei unterlief mir zunächst ein Fehler, der aufschlussreiche Folgen hatte. Ich hatte anfangs vergessen, die Probandin zu bitten, wie in Würzburg die bereitliegenden dünnen Handschuhe anzuziehen. Nachdem sie diese ab dem vierten Run trug, sanken die Trefferzahlen plötzlich auffallend. So kam bei mir der Verdacht auf, dass sich das Tragen von Handschuhen auf die Trefferquote generell nachteilig auswirken könnte. Veronika absolvierte insgesamt achtzehn Runs, dann abwechselnd mal ohne und mal mit Handschuhe (je 540 Trials) und erzielte die Trefferzahlen, die Tabelle 1 zeigt.

Tabelle 1: Treffer-Ergebnisse unter den Bedingungen ohne und mit Handschuhen bei Veronika (Würzburg-Teilnehmerin) und Carolin (Nicht-Teilnehmerin)

	Veronika		Carolin	
	Ohne Handsch.	Mit Handsch.	Ohne Handsch.	Mit Handsch.
N Trials	540	540	480	480
Zufallsproportion für Treffer	.10	.10	.20	.20
Treffer erwartet	54	54	96	96
Treffer beobachtet	115	68	159	132
Z	8.7	1.94	7.1	4.0
p Treffer	$<10^{-15}$.03	$<10^{-14}$.00002
p Unterschied	$<10^{-13}$.03	

Der Unterschied entspricht dem, was vermutet wurde, in hohem Maße. Die Trefferzahlen sinken bei Benutzung von Handschuhen. Um zu erfahren, ob sich der Effekt generalisieren lässt, wurde die Mit-und-ohne-Handschuh-Variation, diesmal beim Standard-Beutelverfahren, mit einer anderen viele Treffer erzielenden Probandin (Carolin) wiederholt. Der Unterschied hatte die gleiche Richtung und war ebenfalls signifikant ($p = 0,03$), wenn auch nicht so extrem.

Schlussdiskussion

Woran liegt es, dass die beiden Test-Teilnehmerinnen von 2006 in Würzburg keine signifikanten Trefferüberhänge zustande brachten?

Box statt Beutel

Nach den Ergebnissen der Untersuchungen mit der Kartonbox in Würzburg, mit der Trommel in Göttingen und der Picknickbox in Göttingen und Würzburg liegt die Schlussfolgerung nahe, dass Box-Bedingungen beim Ballversuch generell potentielle Psi-Effekte vermindern oder sogar zum Verschwinden bringen. Die Beutel-Bedingung ist insgesamt offenbar wesentlich psi-freundlicher.

Diese Erklärung fände vom Erleben der Probanden her Unterstützung: In der Beutel-Bedingung werden nicht nur die Bälle bewegt, auch der Behälter wird gerüttelt und gewendet. Durch das Rütteln von Pingpongbällen im Beutel entsteht ein angenehmes Geräusch. Die Testperson greift in den Beutel hinein, so wie sie im alltäglichen Leben in Beutel, Taschen und Säcke hinein greift. Der ergriffene Ball wird entnommen, in der Hand gehalten, angeschaut und zurück- oder weggelegt, wiederum ähnlich wie im täglichen Leben. Die Testperson erlebt, dass sie über das Testmaterial verfügt, eine natürliche *internale Kausalattribution* dominiert. Auch wenn dieses Gefühl bei der blinden Auswahl eines Balls aus einer größeren Menge von Bällen keinen konventionswissenschaftlich erklärbaren Vorteil für das Trefferziel erkennen lässt, könnte es zur Aktualgenese von Psi-Effekten mit beitragen.

In der Box-Bedingung steht der Testperson ein starrer Behälter gegenüber, in dem die Bälle wie in einem Sicherheitssafe aufbewahrt werden. Sie gelangt zu den Bällen nur durch eine enge Öffnung, an welcher sich Arm und Hände reiben. Überdies hat die Testperson die ausgewählten Bälle wie Fremdkörper sofort wieder zu entfernen, indem sie sie durch eine Öffnung nach außen in einer Ausgangsröhre ablaufen lässt. Sie darf die Ergebnisse ihres Wählens und Greifens nicht in den Händen halten und nicht anschauen. Die Situation ist unnatürlich verfremdet, sie trägt den Stempel misstrauischer Einengung und Behinderung. Die hypothetischen Psi-Effekte könnten auf subtile psychische Bedingungsveränderungen generell empfindlich ansprechen und durch die Boxbedingung gehemmt werden. Mit Einführung von Trommel und Picknickbox, womit ich den Vorstellungen der GWUP entgegen kommen wollte, habe ich Fehlentscheidungen getroffen.

Handschuhe

Möglicherweise ist auch die von der GWUP geforderte Benutzung von Handschuhen eine abträgliche Bedingung, die Befunde bislang sprechen für einen psi-dämpfenden Einfluss. Dem sollte man durch Forschung weiter nachgehen. Werden die Trefferquoten der Testpersonen durch die Verwendung von Handschuhen generell vermindert? Wenn ja, werden

durch den Handschuh-Stoff sensorische Informationen ausgefiltert? Diese Hypothese halte ich für schwach, manche Nebenfunde sprechen dagegen (z. B. Verschiebung der Treffer auf Zahlen, die der gewünschten Zahl benachbart sind). Für wahrscheinlicher halte ich, dass die Gefühle des unmittelbaren Hautkontakts mit den Bällen einen Psi-Effekt begünstigen, und dass das Tragen von Handschuhen diesen Vorteil zunichte macht. Dies wird durch Aussagen von Veronika und Carolin nahe gelegt, die beide sagten, dass sie beim Tragen der Handschuhe das Gefühl hatten, behindert zu werden und den Bezug zu den Bällen zu verlieren.

Seelische Grundverfassung

Zu den Faktoren, die die Testergebnisse beeinflussen können, gehört die innere Verfassung der Teilnehmer, die sie aus ihrem Lebenskontext in die Testsituation mitbringen und die dort mit hineinspielt. Bei höheren Spannungsgraden können paranormale Effekte verschwinden oder vermindert werden. Dies war Katarinas vermutete Erklärung für ihr niedriges Trefferniveau, das sie schon in Göttingen zeigte. Sie berichtete, dass sie nach meiner Mitteilung über ein erwünschtes Mitwirken beim GWUP-Test den Beutel mit Bällen genommen habe, der aus früheren Heimversuchen stammte, um ihre aktuellen Psi-Fähigkeiten zu testen. Nach fünfzehn Zügen habe sie keinen einzigen Treffer gehabt, woraufhin sie den Run, der eigentlich 60 Trials erfordert, aus Enttäuschung nicht beendete. Außerdem sei sie in den Wochen vor dem Würzburg-Test von persönlichen Problemen belastet gewesen. Sie stehe derzeit in Scheidung und habe berufliche Zukunftsängste. Sie berichtete, vor Sorgen zwei Nächte vor dem Termin kaum geschlafen zu haben. Sie rief mich am Abend vor der Reise an und bat mich, sie am nächsten Tag früh um 5.30 Uhr anzurufen, damit sie das Aufstehen für die Zugfahrt nach Würzburg nicht verpasse. Sie sei wegen ihrer Schlaflosigkeit übermüdet und habe Angst, durch anomal tiefen Schlaf ihren Wecker zu überhören. Katarina hätte die Würzburg-Exkursion am liebsten kurzfristig abgesagt, schrieb sie mir am Tage nach dem Test.

Die Trefferzahlen der Versuchsteilnehmerinnen waren am GEMI in Göttingen (Silke, Katarina und Veronika) bedeutsam größer als im Biozentrum in Würzburg. Woran liegt das?

Sensorische Schlupflöcher?

Würde ein Einfluss durch *sensory leakage* vorliegen, dann würde dieser unter Kontrollbedingungen, wenn überhaupt, vermindert auftreten. Eine Testperson zeigt gegenüber einem Experimentator i. a. mehr Selbstkontrolle und Aufmerksamkeit. Die im Alleinversuch sich möglicherweise einschleichenden Tast- oder Wärmeempfindungen hätten unter den Augen von Versuchsleitern weniger Chancen, gespürt und beim Suchen eines Balls im Behälter genutzt zu werden. Doch nahmen bei Katarina und Veronika im Beutel-Test unter meiner Kontrolle die Trefferzahlen nicht ab, sie nahmen zu.

Ein Einfluss sensorischer Faktoren müsste, wenn er vorhanden war, auch *innerhalb einzel-*

ner Runs einen allmählichen Anstieg der Treffer hervorrufen. Sensorische Rückmeldungen können sich nicht schon bei den ersten Ballzügen auswirken, sie müssen bei den ersten Zügen eines Runs gewonnen und gespeichert werden, damit sie sich bei späteren Zügen auswirken können. Doch die Trefferfolgen von Katarina und Veronika zeigen keinen Lernanstieg. Addiert man die Treffer aus den Versuchen am GEMI für jeweils sechs aufeinander folgende Züge (zuerst für Zug 1, 2, 3, 4, 5, 6 dann für 7, 8, 9, 10, 11, 12 usw. bis 55, 56, 57, 58, 59, 60), dann ergeben sich für Katarina und Veronika innerhalb eines Runs von 60 Zügen 10 aufeinander folgende Treffer-Teilsummen (siehe Tabelle 2). Hinzugefügt wurden in Tabelle 2 die Treffer-Teilsummen von 47 Studierenden, die den Beuteltest im Jahre 2003 als Heimtest durchgeführt hatten. Für ein Lernen aufgrund unbewusster Auswertung sensorischer Informationen, das einen ansteigenden Trend von Treffern zur Folge haben müsste, ergeben sich keine Hinweise, weder bei Katarina und Veronika noch in der Gruppe der 47 Studierenden. Die Trendkorrelationen sind tendenziell sogar negativ, bei Katarina $r = -0,41$, bei Veronika $r = -0,45$, bei den 2003 getesteten Studenten $r = -0,27$. (Die Korrelationen wären bei $df=8$ erst mit $r = 0,63$ signifikant.)

Tabelle 2: Treffer-Teilsummen für je sechs aufeinander folgende Ziehungen (Trials) innerhalb eines Runs von 60 Ziehungen.

Teilsummen von Treffern für je sechs Trials											
	N	1-	7-	13-	19-	25-	31-	37-	43-	49-	55-
	Trials	6	12	18	24	30	36	42	48	54	60
Katarina	840	12	23	22	19	15	13	10	16	14	15
Veronika	960	16	21	22	17	18	15	15	16	19	15
Stud2003	22560	258	272	267	270	265	242	261	262	271	252

Für die Stud2003-Stichprobe wurde zusätzlich eine größere Auflösung der Trefferfolge von der ersten bis zehnten Ballziehung vorgenommen, für die zehn ersten Trials wurden 37, 45, 48, 40, 51, 37, 49, 40, 36, 45 Treffer gezählt. Von einem Anstieg der Trefferzahlen vom ersten bis zehnten Trial, der auf Einflüsse eines sensorischen Lernens schließen lassen würde, ist auch hier nichts zu bemerken.

Lässt sich die niedrige Trefferleistung beim *Box-Test* in Würzburg dadurch erklären, dass sinnesdurchlässige Bedingungen, die in Göttingen nicht erkannt wurden, in Würzburg ausgeschlossen wurden? Dem widerspricht die Tatsache, dass ein Absinken der Trefferzahlen bei beiden Testpersonen in der Box-Bedingung schon in Göttingen, nicht erst in Würzburg beobachtet wurde. Auch lässt sich nicht ernsthaft annehmen, dass in Göttingen im Beutelversuch unbemerkte Wahrnehmungen, zum Beispiel durch Wärmestrahlung, vorkamen, die bei den

in den Beutel zurückgegebenen Bällen das weitere Auffinden von Bällen mit erwünschten Zahlen unterstützt hätten. Das Ergebnis von Katarina und Veronika spricht dagegen, da es bei ihnen innerhalb der 16 Beutel-Runs am GEMI zu großen Unterschieden in der Trefferleistung kam. Bei Effekten einer Wärmestrahlung und anderer sensorischer Informationen, wenn es sie gäbe, sollte man über die Versuchsdurchgänge hinweg Konstanz erwarten, die beobachtete überzufällige Streuung der Trefferzahlen sind so nicht zu erklären.

Eine Erklärung der niedrigen Trefferleistung in Würzburg durch unbekanntes sinnesdurchlässige Faktoren würde zudem nicht ausreichen zur Erklärung des sehr signifikanten Unterschieds der Würzburger Trefferleistungen zwischen Katarina und Veronika. Wenn die von der GWUP hergestellten Kontroll-Bedingungen optimal waren und nach Ansicht skeptischer Vertreter der Nullhypothese Abweichungen von der Zufälligkeit ausschlossen, dann dürften nicht nur keine signifikanten Trefferüberhänge, sondern auch keine signifikanten Unterschiede zwischen Testpersonen auftreten.

Tricks?

Sind die hohen Trefferüberhänge in Göttingen durch Tricks der beiden Testpersonen zustande gekommen und sind die geringen Trefferquoten in Würzburg dadurch zu erklären, dass Tricks, deren Anwendung der Testleiter in Göttingen nicht bemerkt hat, ausgeschaltet wurden? Dagegen spricht, dass die Trefferzahlen in Göttingen unter Kontrolle deutlich höher waren als daheim ohne Kontrolle. Die Testpersonen hätten unter Kontrolle, wo sie weniger tricksen konnten, weniger und nicht mehr Treffer erzielen müssen als unter der Heimbedingung.

Die GWUP war in erster Linie an der Ausschaltung möglicher Tricks interessiert, weshalb Rainer Wolf darauf bestand, Bälle, die einmal gezogen worden waren, kein zweites mal im gleichen Test zu verwenden. Bei nochmaliger Verwendung gezogener Bälle, so meinte er, könne sich eine Testperson an Markierungen orientieren, die sie mit Kenntnis der gezogenen Zahlen vor dem Zurücklegen in den Behälter an den Bällen anbringt.

Dies zu denken halte ich für abwegig, denn die technischen Probleme sind unlösbar: Wie soll die Testperson die Bälle markieren? Sie müsste dies mit nur einer Hand tun, und dazu einen heimlich in die Box oder den Beutel eingeschmuggelten Markierungsstift verwenden. Es müsste vor den Augen der Versuchsleiter geschehen und ein Ergebnis haben, das nach dem Versuch beim späteren kontrollierenden Abtasten der Bälle nicht bemerkt wird. Auch hätte das Betrugsmanöver von mir vorher einstudiert werden müssen. Zudem hätte ich mich, wenn ich ein solches geplant hätte, nicht darauf eingelassen, dass ein gezogener Ball nicht noch einmal verwendet wird bzw. ich wäre zu einem Versuch in Würzburg nicht angetreten, das einen Wechsel der Bälle nach einmaliger Verwendung vorsieht. Als Betrugsorganisator hätte ich ja vorausgesehen, dass der Test, in dem nicht gemogelt werden konnte, scheitern würde.

Fazit

Ein Betrugsverdacht erfordert eine Begründung

Die GWUP-Repräsentanten brachten ziemlich drastisch zum Ausdruck, dass sie einen Wissenschaftler-Kollegen, in diesem Falle mich, verdächtigten, er könne das Ergebnis des Experiments, das er ihnen mit seinen Mitarbeitern vorführen wollte, manipulieren. Sie hielten ein betrügerisches Verhalten bei ihm nicht nur für denkbar, sondern rechneten mit ihm so sehr, dass sie penible Schutzmaßnahmen ergriffen.

Nach meiner Auffassung ist ein offen zum Ausdruck gebrachter Betrugsverdacht gegenüber einem Wissenschaftler-Kollegen ein Verstoß gegen wissenschaftsethische Normen. Zwar darf man einen Forscher auch in der Öffentlichkeit verdächtigen, einen Betrug vorgenommen zu haben oder zu beabsichtigen, doch dies nur dann, wenn hinreichende Indizien vorliegen. Diese lagen nicht vor. Ohne Stützung durch Indizien verletzt man mit einem Betrugsvorwurf oder -verdacht die für einen kooperativen Umgang in der Wissenschaft unverzichtbaren Normen des Vertrauens.

Einen blinden Verdacht zu rechtfertigen mit dem Argument, es handele sich nur um einen „generellen“ Verdacht, nicht um einen persönlich gerichteten, ist nicht stichhaltig, denn die vermeintlich notwendigen Schutzmaßnahmen wirken sich einengend und u.U. entwürdigend auf die individuell betroffene Forscherpersönlichkeit aus.

Der Wissenschaftsprozess bedarf wie andere Formen des produktiven menschlichen Zusammenwirkens einer Basis des Vertrauens. Dies zu begründen, erübrigt sich. Detailliert zu begründen ist dagegen jede Abweichung von der Vertrauensnorm, wobei Evidenzen oder zumindest Fast-Evidenzen, die ein Misstrauen nahe legen, unaufgefordert vorzulegen wären.

Eine Analogie: Die Ablehnung eines Richters durch einen Anwalt mit dem generellen Argument, Richter können befangen sein, wäre vergleichbar mit der Ablehnung eines Psi-Forschers mit dem Argument, Parapsychologen können betrügerisch vorgehen. Natürlich ist vor Gericht das Argument des Verdachts auf Befangenheit erst diskutabel, wenn vom Anwalt stichhaltige Begründungen für Befangenheit eines betreffenden Richters vorgelegt werden. Einen solchen Antrag mit einem „generellen“ Verdacht gegenüber Richtern zu begründen, wäre geradezu lächerlich.

In Würzburg aber fühlte man sich zur Verdachtsbegründung im Einzelfall nicht verpflichtet. Ich hatte den Eindruck, dass man sich dort zu beliebiger Verdächtigung berechtigt und dazu befugt glaubt, das Forschungsdesign eines Wissenschaftlers, der Grenzphänomene untersuchen will, zu manipulieren. Offenbar erwartet man von ihm, dies als besondere Gewissenhaftigkeit in der Forschung entsprechend zu honorieren.

Zur Frage des GWUP-Preises

Die GWUP möge Überlegungen anstellen, ob sie weitere Forschung zu parapsychologischen Fragen auch außerhalb ihres Preisprojekts in Kooperation mit anderen Forschern durchführen würde. Aufgrund meiner Erfahrungen von 2005 und 2006 gewinnen die GWUP-Versuchsleiter über die Forscher, die sich auf die eine Preisbewerbung einlassen, über diese eine bemerkenswerte Autorität. Bei Verhandlungen über Durchführungsmodalitäten steht ständig das unausgesprochene Argument im Hintergrund: „Wenn du meinen Batzen Geld gewinnen willst, musst du schon meine Bedingungen akzeptieren.“ Wie ein Arbeitnehmer ohne Gewerkschaftsschutz hat sich der Bewerber dem Geldgeber, der den Lohn zahlt, unterzuordnen. Die im normalen Wissenschaftsprozess selbstverständliche Gleichberechtigung von Argumenten, der „herrschaftsfreie Diskurs“, ist bei aller Wahrung anständiger Umgangsformen in dem durch die ausgesetzte Prämie monetär definierten Gefälle unterschwellig ständig gefährdet.

Wünschenswert wäre auch, in gemeinsame Projekte namhafte Methodiker mit einzubeziehen, die zwar parapsychologischen Hypothesen auch skeptisch gegenüber stehen mögen, doch ihr Urteil aufgrund fehlender institutioneller Bindung unabhängig von verpflichtender Rücksichtnahme auf Meinungsmehrheiten der Vereinigung, der sie angehören, finden dürfen.

Zukünftige Forschungsziele

Weitere Versuche in Kooperation mit der GWUP sind wünschenswert, doch sollten Bedingungsverbesserungen vorgenommen werden, die die bisherigen Versuchsergebnisse optimal berücksichtigen. Mitglieder der GWUP haben im informellen Gespräch erfreulicherweise ein anhaltendes Interesse an einer Untersuchung paranormaler Phänomene bekundet.⁵ Folgende Projekte ließen sich diskutieren:

Fachliche Begutachtung der Beutelbedingung

Eine Begutachtung der Beutel-Bedingung durch unabhängige professionelle Fach-Magier wäre sinnvoll. *Erwartung:* Die Beutelbedingung wird, möglicherweise mit leichten Veränderungen, von ihnen als unbedenklich erklärt werden. Auf eine Box, auf Handschuhe, auf das Nichtzurücklegen der Bälle könnte man ggf. nach ihrem Urteil in Zukunft verzichten. Um einen vernünftig bleibenden letzten Zweifel an sauberen Versuchsbedingungen auszuräumen,

5 Ein Beispiel gab Chris French, einer der führenden britischen Skeptiker. Er hat zwei Studierenden die Überprüfung des Pingpong-Balltests zur schriftlichen Prüfungsaufgabe gemacht. Nach meiner Reanalyse der Daten der beiden BSc-Kandidaten (40 Probanden) ergibt sich eine Treffer-Abweichung vom Zufall mit $p = 0,002$ bzw. $p = 0,0003$.

könnte man mehrere gleiche Beutel verwenden, die von Trial zu Trial gewechselt werden.

Versuchsmaterial Beutel vs. Box (Replikationsversuch)

Wenn die Beutel-Bedingung von den Trick-Fachleuten für unbedenklich erklärt wird, könnte man den Ballversuch von denselben Test-Teilnehmern abwechselnd von Run zu Run unter der Beutel- und Box-Bedingung durchführen. *Hypothese:* Unter der Beutel-Bedingung werden mehr Treffer erzielt als unter der Box-Bedingung.

Anwesenheit von Beobachtern, die der Testperson Erfolg wünschen vs. Anwesenheit von Beobachtern, die ihr Misserfolg wünschen

Der Balltest wird von ausgewählten Psi-Begabten im Einzelversuch durchgeführt unter zwei Bedingungen: 1. Bei der Testdurchführung sind etwa vier andere Personen anwesend, die wünschen sollen, dass die Testperson eine Trefferrate erreicht, die ihre Spitzenleistung übertrifft. Sie sollen (monetär) belohnt werden, in dem Maße, wie ihnen dies gelingt. 2. Anwesend sind etwa vier andere Personen, die wünschen sollen, dass die Testperson eine geringere Trefferrate hat, als sie selbst bislang gezeigt hat. Sie sollen belohnt werden in dem Maße, wie ihnen dies gelingt. Die Testperson weiß nicht, dass die Anwesenden ihr Erfolg oder Pech wünschen. *Hypothese:* Die wünschenden Anwesenden beeinflussen die Trefferraten der Testperson.

Skeptischere vs. offenere Einstellung des Experimentators

Balltest-Versuche sollten abwechselnd durchgeführt werden von Versuchsleitern, die sich nach einem Fragebogen zur Okkultgläubigkeit als extrem glaubensabgeneigt erweisen (z.B. als Mitglieder der GWUP), ein anderes Mal von Versuchsleitern, die nach diesem Fragebogen überdurchschnittlich glaubensbereit sind. Die Glaubensgegner als Versuchsleiter sollten dieselben Testpersonen auf ASW hin untersuchen. *Hypothese:* Bei gleichen Testpersonen resultieren aus den Testserien mit glaubensabgeneigten Versuchsleitern signifikant weniger Treffer als aus den Testserien mit glaubensbereiten Versuchsleitern (Hypothese des „Experimentator-Effekts“).

Zurücklegen vs. Nicht-Zurücklegen der Bälle

Durchführung des Beutelversuchs, abwechselnd von Run zu Run, mit Zurücklegen der Bälle und ohne Zurücklegen der Bälle in den Beutel. *Hypothese:* Zwischen den Bedingungen mit und ohne Zurücklegen gibt es keine systematischen Treffer-Unterschiede.

Testdurchführung mit vs. ohne seelische Belastung

Durchführung des Beutelversuchs mit einer ausgelesenen Stichprobe von guten Ballziehern in einem Prüfungssemester, zunächst noch ohne Prüfungsdruck, ein zweites Mal etwa eine

Stunde vor einer mündlichen Prüfung, und ein drittes Mal bald nach der Prüfung, wobei noch zwischen gutem und schlechtem Bestehen bzw. ggf. Nichtbestehen zu unterscheiden wäre (ggf. genügt eine bevorstehende Statistik-Klausur als Druckbedingung). *Hypothese:* Die Trefferzahlen liegen niedriger in einer Situation mit Prüfungsdruck als ohne Prüfungsdruck, niedriger nach einer schlecht oder nicht bestandenen Prüfung als nach einer gut bestandenen Prüfung.

Trefferleistung einer Person vs. Unterschiede der Trefferleistung zwischen zwei oder mehr Personen

Bei einem nächsten Test in Würzburg könnten konstante Treffer-Unterschiede zwischen Testpersonen als Psi-Indikator dienen anstelle der bisher geforderten Treffer-Höchstleistungen. Katarina und Veronika z.B. hätten, trotz niedriger Trefferleistung mit einem signifikanten Ergebnis abgeschlossen, wenn ein Unterschied in ihren Leistungen als Kriterium vereinbart gewesen wäre. Man könnte in Zukunft zwei Personen oder Personengruppen auswählen, die in Vorversuchen relativ hohe und anhaltende *Unterschiede* in ihren Trefferzahlen zeigen. *Hypothese:* Testpersonen, die in Vorversuchen konstant hohe Trefferzahlen haben, unterscheiden sich von denen, die in Vorversuchen konstant niedrige Trefferzahlen haben, in gleicher Richtung, auch unter Kontrolle der GWUP.

Instruktionswechsel intraindividuell: Möglichst viele Treffer vs. möglichst wenig Treffer

Testpersonen mit hohen Trefferzahlen sind unter der Standardbedingung dazu anzuhalten, von Run zu Run abwechselnd mal möglichst viele Treffer und mal möglichst wenig Treffer zu erzielen. Balltests mit der Instruktion, möglichst wenig Treffer zu erzielen, sind mit sieben Testpersonen, die unter Standardbedingungen hohe Trefferzahlen erzielten, bereits durchgeführt worden (Balltest nur mit Zahlen, Zufallserwartung 20% Treffer). In insgesamt 104 Runs (6240 Trials) erzielten sie in der *Möglichst-wenig-Treffer*-Bedingung 862 Treffer, erwartet wurden 1248 Treffer. Das sind pro Run 8,28 Treffer statt zufallserwartete 12 Treffer ($Z = -12,2$, $p = 10^{-20}$). In der *Möglichst-viele-Treffer*-Bedingung, also in der Standardbedingung, erzielten die gleichen sieben Testpersonen in 80 Runs 2052 Treffer. Das sind pro Run 25,7 Treffer statt der zufallserwarteten 12 Treffer ($Z_{(\text{Binomial})} = 40,0$, $p < 10^{-100}$). Diese Versuche wurden ohne Kontrolle unter Heimbedingungen durchgeführt.

Der Vorteil wäre ein doppelter: (1) Zur Erklärung eines signifikanten Unterschieds zwischen hohen und niedrigen Trefferzahlen, die bei derselben Testperson vorkommen, kann man sinnesdurchlässige Faktoren oder Markierungen noch weniger heranziehen, als wenn nur hohe Trefferzahlen erzielt werden. (2) Man könnte die Trials der *Möglichst-wenig-Treffer*-Bedingung nur zur Festlegung des mittleren Erwartungswertes verwenden, ohne dass die individuellen Trials in der Endauswertung berücksichtigt werden. Der empirisch ermittelte Erwartungswert würde den mittleren Zufallswert (MCE) für den anschließenden Hauptversuch, bei dem es nur noch um hohe Trefferzahlen geht, ersetzen. *Hypothese:* Beim Wünschen

von vielen Treffern werden mehr stärkere Effekte ermittelt, wenn der Erwartungswert durch Ergebnisse beim Wünschen von wenigen Treffern definiert wird, als wenn er als MCE definiert wird.

Instruktionswechsel intraindividuell: Möglichst hohe vs. möglichst niedrige Zahlen ziehen

In einer Serie von Runs werden Testpersonen gebeten, möglichst die 5, alternativ die 4 usw., bzw. möglichst hohe Zahlen zu ziehen. In einer anderen Serie von Runs werden sie gebeten, möglichst die 1, alternativ die 2 usw., bzw. möglichst niedrige Zahlen zu ziehen. Für jeden Run wird die Summe der gezogenen Zahlen ermittelt. Der Vorteil ist u.a. ein psychologischer: Die gezogenen Zahlen werden nicht nur nach richtig und falsch, sondern auch nach den Zahlenwerten beurteilt, womit die Aufgabe abwechslungsreicher wird. *Hypothese*: In den Serien mit Wunsch nach hohen Trefferzahlen werden höhere Zahlensummen erzielt als in den Serien mit Wunsch nach geringen Zahlensummen.⁶

Beiträge anderer Forscher

Kritische Beobachter mit Skepsis, die für möglich halten, dass man die hohen inter- und intraindividuellen Varianzen der Trefferquoten im Balltest auch ohne den Psi-Begriff irgendwie alternativ erklären können müsste, mögen weitere Versuchsanordnungen entweder vorschlagen oder, besser noch, in eigene Forschung umsetzen.

Literatur

- Ertel, S. (2004a): The ball drawing test. Psi from untrodden ground. In: Thalbourne, M.A.; Storm, L. (Eds.): Parapsychology in the twenty-first century: The Future of Psychical Research. McFarland, Jefferson/NC.
- Ertel, S. (2004b): Astrologie und Psi. Eine Fallstudie verstärkt die Zusammenhangshypothese. *Zeitschrift für Anomalistik* 4, 52-68.
- Ertel, S. (2005a): Psi test feats achieved alone at home: Do they disappear under lab control? *Australian Journal of Psychology* 10 (2), 149-164.
- Ertel, S. (2005b): Run auf Randis Million – das Nadelöhr der GWUP in Würzburg. Beobachtungen und

6 Dieser Versuch wurde von mir im Wintersemester 2006/7 mit 38 Studienanfängern zunächst als Heimtest durchgeführt. Die Hypothese wurde bestätigt. Bedingung 1: möglichst niedrige Zahlen, a. M. = 2,92, SAW = 0,749. Bedingung 5: möglichst hohe Zahlen ziehen, a. M. = 3,00, SAW = 0,085. Der Unterschied ist nach dem Mann-Whitney-U-Test sehr signifikant: $Z = 3,06$, $p = 0,001$. Allerdings ist eine signifikante Abweichung vom Erwartungswert 3 nur bei der 1-Bedingung festzustellen ($t = 3,3$, $df = 37$, $p = 0,002$), nicht bei der 5-Bedingung ($t = 0,28$, $df = 37$, n.s.). Das zeigen die beiden Probandinnen mit höchsten Treffern bei Ziel 1 (85 bzw. 76 Treffer, erwartet 48). Beim Ziel 5 haben beide Probandinnen 50 Treffer (erwartet ebenfalls 48).

Reflexionen. *Skeptiker* 2/2005, 63-65.

Rosenthal, R.; Rubin, D.B. (1989): Effect size estimation for one sample multiple choice type data: Design, analysis, and meta analysis. *Psychological Bulletin* 106, 332-337.

Anmerkung der Redaktion

Die Redaktion der *Zeitschrift für Anomalistik* hat ein halbes Dutzend GWUP-Mitglieder, die an diesen Experimenten federführend oder als Helfer und Beobachter beteiligt waren, dazu eingeladen, an dieser Stelle Kommentare zum Aufsatz von Suitbert Ertel zu veröffentlichen. Leider hat binnen der üblichen Abgabefrist von sechs Wochen bis zum Redaktionsschluss kein einziges der betroffenen GWUP-Mitglieder eine Stellungnahme eingereicht. Wir verschieben deshalb die Veröffentlichung der gesammelten Diskussionsbeiträge zu diesem Aufsatz auf die nächste Ausgabe der *Zeitschrift für Anomalistik* (Heft 1/2008), um den GWUP-Vertretern nochmals Zeit zur Einreichung von Repliken zu geben. Der entsprechende Redaktionsschluss ist der 31. März 2008. Auch andere bis dahin noch eingeschickte Diskussionsbeiträge zu diesem Aufsatz werden dann mit veröffentlicht.