

Die drei Themen des Journal Clubs dieser Ausgabe sind durch drei unsichtbare Ariadnefäden miteinander verbunden. Wir besprechen die jüngsten Replikationsversuche des isopathischen Forschungsmodells in der klinischen Homöopathieforschung, das durch David Reillys erste Lancet-Publikation einer kontrollierten klinischen Homöopathiestudie weite Berühmtheit erlangte, zusammen mit zwei norwegischen Arbeiten, die man als konzeptuelle Replikation bezeichnen könnte. Als Interpretationsfolie wird uns eine noch unpublizierte Arbeit dienen, die im Frühjahr dieses Jahres auf der Konferenz «Evidence Based Complementary Medicine» vorgestellt wurde und als Abstract in dieser Zeitschrift publiziert wurde. Dies ist die ausführlichste Serie von Replikationen, die es in der Homöopathieforschung gibt, sieht man einmal von den Studien Wiesenauers mit Galphimia C6 ab, die ja aufgrund der tiefen Potenzierung weniger kontrovers sind. Als zweitem Thema wenden wir uns der Fernheilstudie aus Exeter zu, die geistiges Fernheilen bei Warzen untersuchte. Drittens gehen wir ein wenig über die Grenze der medizinischen Forschung hinaus und grasen im Niemandsland der Parapsychologie, indem wir die jüngsten Versuche Sheldrakes, seine Theorie der morphogenetischen Felder zu belegen, unter die Lupe nehmen.

Die Theorie der morphogenetischen Felder – ein erster verbindender Strang zwischen den verschiedenen Themenkreisen – wurde und wird verschiedentlich bemüht, um das mögliche Funktionieren komplementärmedizinischer Massnahmen, aber auch von Gebet und Fernheilung, theoretisch zu untermauern. Sie ist bei verschiedenen Proponenten alternativer Wissenschaftsansätze sehr beliebt und bei allen New-Age-Nahestehenden eine geläufige Erkennungsmerkmale. Sheldrake hatte vor Jahren postuliert, dass man bestimmte Phänomene im Zusammenhang mit der biologischen Formentwicklung, aber auch unkonventionelle Phänomene aus den Grenzgebieten der Wissenschaften nur dann gut verstehen könne, wenn man die Existenz sogenannter morphogenetischer Felder annehme. Darunter sind immaterielle Kraftfelder zu verstehen, die durch das Entstehen und Wiederholen bestimmter Formen und Zusammenhänge entstehen und auf das Auftreten ebendieser Formen und Zusammenhänge später und anderswo erleichternd wirken. Sie haben vage Ähnlichkeit mit dem aristotelischen Begriff der Form, die ja bekanntlich die konkrete Ausformung einer Materie, die reine Möglichkeit darstellte, erst Wirklichkeit werden lässt. Wichtig für das Verständnis von Sheldrakes Begriff der morphogenetischen Felder ist die Tatsache, dass diese Felder, ähnlich den bekannten Feldern, also etwa dem Gravitationsfeld oder dem elektromagnetischen Feld, unmittelbar mit materiell-physikalischen Ereignissen wechselwirken können sollen. Sie wären also eine Möglichkeit, den cartesianischen Schnitt zwischen materiellen und immateriellen Systemen zu überwinden. Wichtig ist ausserdem, dass diese Felder veränderlich sind. Sie entstehen, wenn eine kritische Masse von neuen Formen oder Zusammenhängen vorhanden ist. Wenn etwa eine bestimmte Anzahl einer biologischen Spezies ein neues Verhalten gelernt hat – z. B. Meisen in England, die Aludeckel von Milchflaschen aufzupicken –, dann entsteht ein morphogenetisches Feld, das wiederum anderen Mitgliedern der gleichen Spezies – Meisen in Holland – ermöglichen soll, dieses Verhalten leichter und schneller zu zeigen, als per Zufall zu erwarten ist. Morphogenetische Felder wären also eine Art überindividuelles Gedächtnis der Natur, weil sie die Lern- und Verhaltensgeschichte über Zeit und Raum hinweg repräsentieren. Gäbe es so etwas wie morphogenetische Felder, dann könnte man theoretisch ohne Probleme verstehen, wie etwa das Beten und Wünschen vieler

Menschen ein Feld aufbaut, das unmittelbare physikalische Wirkungen zeigt, unabhängig von zeitlichen und räumlichen Entfernungen. Geist- oder Gebetsheilung wäre dann nichts Mysteriöses mehr, sondern eine subtile Art physischer Einwirkung. Man könnte das Modell sogar dafür hernehmen, sich die Wirkungsweise der Homöopathie zu erklären. Meines Wissens ist zwar noch niemand explizit auf diesen Gedanken gekommen, aber in Ansätzen würde das etwa so funktionieren: Das ausführliche Beschäftigen mit einer Substanz und ihren Eigenschaften – über Arzneimittelprüfungen, durch vermehrte Berührung bei Vergiftungen, durch das Aufstellen eines Arzneimittelbildes und häufige klinische Verwendung, schliesslich durch den zeit- und energieaufwendigen Zubereitungsprozess bei der Herstellung – erzeugt ein morphogenetisches Feld. Dieses Feld tritt durch das homöopathische Ritual in Wechselwirkung mit einem Kranken, bei dem diese Substanz indiziert ist, und löst so biologische Prozesse aus, die die Selbstheilung anstossen. Die Theorie der morphogenetischen Felder und der morphischen Resonanz hat also durchaus Auswirkungen auf mögliche oder unmögliche Theoriebildungen innerhalb der Komplementärmedizin. Vor einigen Jahren hat Sheldrake konkrete Experimente vorgeschlagen, mit denen sein Modell testbar ist, und begonnen, solche durchzuführen. Zu diesen gehört das Experiment, bei dem andere Menschen von hinten angeschaut werden und abschätzen müssen, ob gerade jemand hinschaut oder nicht. Dies ist zwar ein Standardparadigma der Parapsychologie, das bereits vor Sheldrake durchgeführt wurde, erhält aber durch seine Publizität neue Brisanz.

Der zweite verbindende Strang ist also, so befremdlich das klingen mag, die Parapsychologie. Diese Disziplin beschäftigt sich schon seit ihren Anfängen im 19. Jahrhundert damit, ob es so etwas wie direkte Einflüsse mentaler Ereignisse auf physikalische Ereignisse geben kann. Sheldrakes Experimente sind Beispiele für die klassische Fragestellung, ob so etwas wie eine Intention, repräsentiert durch den Blick, irgend einen direkten Einfluss auf das Vermuten einer anderen Person haben kann. Für die Geistesheilung werden ebenfalls immer wieder direkte Verbindungen zwischen dem Geist einer Person – des Heilers – und dem Körper einer anderen reklamiert. Dass sich hinter diesen Konzepten oft hochbrisante Fragen verbergen – wie etwa die nach der Natur seelischer und körperlicher Vorgänge –, sei hier nur kurz erwähnt. Die Komplementärmedizin hat mit der Parapsychologie etwas gemeinsam und kann von dieser viel lernen: das Replizierbarkeitsproblem, den dritten, verbindenden Strang. Die Tatsache, dass Einzelbefunde nur unter grösster Schwierigkeit, wenn überhaupt, replizierbar sind, ist beinahe zu einem Markenzeichen parapsychologischer Befunde geworden. Neuere Modellbildungen, zu denen unter anderem einer der Autoren des Journal Clubs dieser Ausgabe, Walter von Lucadou, massgeblich beigetragen hat, sagen solche mangelnde Replizierbarkeit als notwendigen Bestandteil der vorliegenden Phänomene geradezu voraus. Möglicherweise bestünde die Lösung für das Replizierbarkeitsproblem der Homöopathie darin, nicht nach einem stabilen, replizierbaren Modell Ausschau zu halten, sondern sich dem Bemühen der Parapsychologen anzuschliessen und die mangelnde Replizierbarkeit in ein Metamodell einzubauen. Ob ein solches Replizierbarkeitsproblem auch bei der Homöopathie vorliegt und wie dieses möglicherweise lösbar wäre, werden wir bei der Besprechung der Isopathiestudien sehen. Wenn dies so ist, wenn also bestimmte Bereiche der Komplementärmedizin – etwa die Geistesheilung, die Homöopathie, möglicherweise auch andere Phänomene – mehr Ähnlichkeit mit der

Parapsychologie als mit der Pharmakologie haben, dann wäre zu erwarten, dass bei der Beforschung dieser Bereiche auch ähnliche Probleme wie in der parapsychologischen Forschung auftauchen werden. Dann ist zu vermuten, dass etwa die Serie positiver Befunde in der Geistheilforschung, wie unlängst hier

vorgestellt, abbricht, oder dass ein anfangs gefundener Effekt in der Homöopathie sich verliert. Dann wäre es an der Zeit, andere Modelle für die Erforschung dieser Phänomene zugrunde zu legen. Möglicherweise gibt dieser Journal Club ein paar Hinweise.

Sind homöopathische Effekte replizierbar? Das Isopathiemodell unter der (parapsychologischen) Lupe

Taylor MA, Reilly D, Llewellyn-Jones RH, McSharry C, Aitchison TC: Randomised controlled trials of homoeopathy versus placebo in perennial allergic rhinitis with overview of four trial series. *BMJ* 2000;321:471–476.

Objective: To test the hypothesis that homoeopathy is a placebo by examining its effect in patients with allergic rhinitis and so contest the evidence from three previous trials in this series.

Design: Randomised, double-blind, placebo-controlled, parallel group, multicentre study.

Setting: Four general practices and a hospital ear, nose, and throat outpatient department.

Participants: 51 patients with perennial allergic rhinitis.

Intervention: Random assignment to an oral 30c homoeopathic preparation of principal inhalant allergen or to placebo.

Main outcome measures: Changes from baseline in nasal inspiratory peak flow and symptom visual analogue scale score over 3rd and 4th weeks after randomisation.

Results: Fifty patients completed the study. The homoeopathy group had a significant objective improvement in nasal airflow compared with the placebo group (mean difference 19.8 l/min, 95% confidence interval 10.4–29.1, $p=0.0001$). Both groups reported improvement in symptoms, with patients taking homoeopathy reporting more improvement in all but one of the centres, which had more patients with aggravations. On average no significant difference between the groups was seen on visual analogue scale scores. Initial aggravations of rhinitis symptoms were more common with homoeopathy than placebo [7 (30%) versus 2 (7%), $p=0.004$]. Addition of these results to those of three previous trials ($n=253$) showed a mean symptom reduction on visual analogue scores of 28% (10.9 mm) for homoeopathy compared with 3% (1.1 mm) for placebo (95% confidence interval 4.2–15.4, $p=0.0007$).

Conclusion: The objective results reinforce earlier evidence that homoeopathic dilutions differ from placebo.

Aabel S, Laerum E, Dølvik S, Djupesland P: Is homeopathic 'immunotherapy' effective? A double-blind, placebo-controlled trial with the isopathic remedy *Betula* 30c for patients with birch pollen allergy. *Br Homeopathic J* 2000; 89:161–168.

Objective: To examine the effect of the homeopathic remedy *Betula* 30c vs. placebo for patients with birch pollen allergy.

Design: Double-blind, randomized, placebo-controlled trial with 3 days single-blind placebo run-in.

Setting: 4 general practices (3 homeopathic, 1 general) in Oslo, Norway, May 1995.

Participants: 70 patients with verified seasonal birch-pollen allergy. 3 were excluded after randomisation because of missing allergy symptoms, 1 dropped out. 34 patients were treated with placebo, 32 with homeopathy.

Intervention: *Betula* 30c or placebo one tablet daily for maximally 4 weeks or until clear improvement was experienced, during verified birch pollen season.

Main outcome measure: Total score of 17 different allergy symptoms, rated daily, and converted into a sum score. Point estimates of the median difference between the experimental and placebo groups with 95% confidence intervals were the main measure of effect.

Results: No statistically significant difference between the groups was found during the first and last period of May. However, from May 8 to 18, a clinically interesting difference was revealed between the groups, those receiving *Betula* 30c having fewer and less serious symptoms. For some days these differences were statistically significant. Surprisingly, this group reported more aggravation from the tablets than did the placebo group.

Conclusion: With a statistical power of 70% for a defined clinically interesting difference (25%), the present results indicate that treatment with *Betula* 30c during the pollen season deserves further attention.

Aabel S: No beneficial effect of isopathic prophylactic treatment for birch pollen allergy during a low-pollen season: A double-blind, placebo-controlled clinical trial of homeopathic *Betula* 30c. *Br Homeopathic J* 2000;89:169–173.

Objective: To determine if the homeopathic medicine *Betula* 30c is more effective than placebo at reducing symptoms of pollen allergy in patients sensitive to birch pollen.

Design: Double-blind, randomized, placebo-controlled trial. 4 weeks were prophylactic, with tablet intake once per week 1 tablet. After that, tablet intake as necessary, if allergy symptoms occur, up to 3 tablets daily, with 1 tablet per day as a rule.

Setting: General practice on Oslo, Norway, May 1996.

Participants: 80 patients (children, adolescents, and young adults) aged 7–25 with verified birch pollen allergy were randomized, 7 dropped out for various reasons.

Main outcome measure: Visual analogue scale of allergy symptoms, rated by patients or parents. Main criterion was the median symptom score with 95% confidence interval for all treated patients during a 10-day period, starting individually with the first allergy day.

Results: The pollen count was very low in 1996. Only on 3 days it was high enough to provoke allergy symptoms. Surprisingly, the verum-treated patients fared worse than the placebo group. They used more rescue medication, and had higher symptom scores during these days.

Conclusion: Homeopaths might attribute the findings to a putative aggravation response, but the results certainly do not lend support to the usefulness of the tested prophylactic approach, under conditions of low allergen exposure.

Kommentar – H. Walach, Freiburg

Als David Reilly (1986) seine mittlerweile klassische Studie zur Wirksamkeit potenzierten Graspollens C30 bei Heuschnupfen in Lancet publizieren konnte (Reilly et al. [5]), ging ein Triumphruf durch die Homöopathiegemeinde. Man meinte, endlich sei die Homöopathie bewiesen. Als schliesslich eine analoge Studie – potenziertes Antigen bei Asthma [4] – wiederum von Lancet publiziert wurde, in der noch dazu eine Meta-Analyse der drei Isopathiestudien eine klare Überlegenheit isopathischer Hochpotenzen über Placebo belegte, schien das Verdikt klar: Entweder Homöopathie – hier operationalisiert durch Isopathie – ist kein Placebo, oder etwas ist faul am Prinzip der klinischen Studie. In der Folge wurde der Ruf nach unabhängigen Replikationen laut. Heute sehen wir zu diesem Themenbereich nun gleich mehrere. Im Kopfschmerzbereich gibt es ebenfalls eine Reihe unabhängiger, konzeptueller Replikationen [6, 8, 10] einer ersten Studie [2], die zusammengekommen nicht positiv ausgefallen sind, so dass klassische Migräneforscher Homöopathie bei Kopfschmerz klar und deutlich belegt als unwirksam ansehen [3]. Am Modell der isopathischen Behandlung allergischer Probleme, so scheint es, könnte sich die Homöopathie entscheiden. Dabei ist es relativ unerheblich, ob wir es mit Asthma, Heuschnupfen oder anderen Formen allergischer Rhinitis zu tun haben. Sowohl Asthma als auch Heuschnupfen sind atopische Krankheiten und werden sowohl von der konventionellen Theoriebildung als auch von der homöopathischen Krankheitstheorie als zusammengehörig angenommen. Es scheint auch unerheblich zu sein, dass nicht die klassische Homöopathie getestet wird. Denn man kann argumentieren, dass bei jeder Untersuchung der

klassischen Homöopathie immer auch die Kompetenz des Homöopathen mitgetestet wird, falsch Negative also nie auszuschliessen sind. Das macht ja das isopathische Forschungsmodell so ansprechend: Die Arzneimittelwahl ist durch den isopathischen Ansatz definiert, und die beiden Wirkprinzipien der Homöopathie – Ähnlichkeit (in geringerem Masse), aber vor allem das kontroverse Potenzierungsprinzip – stehen in Reinform auf dem Prüfstand. Ist das Isopathiemodell Reillys replizierbar, auch in den Händen anderer Forscher? Ist es robust gegenüber leichten Abwandlungen der Durchführung? Ist es generalisierbar auch auf andere Arten der Allergie und andere Allergene? Dies sind die entscheidenden Fragen.

Wenden wir uns der ersten Studie zu, die von Reillys früherer Frau und Mitarbeiterin Morag Ann Taylor durchgeführt wurde. David Reilly war hier nicht der federführende Autor, sondern hat als methodischer Ratgeber zur Seite gestanden und dafür gesorgt, dass die Studie möglichst ähnlich jener war, die er selber durchgeführt hat. Die Studie stellt eine Mischung aller drei früheren Studien Reillys dar: Erstens ist sie eine Studie, die allergische Rhinitis zum Thema hat, wie die ersten beiden (ich rechne auch die Pilotstudie, die im British Homeopathic Journal erschienen ist, mit dazu). Zweitens wird nicht Pollenallergie zum zentralen Gegenstand der Behandlung gemacht, sondern andere Arten der Allergie. Der Arzt, der auf diese Art der Behandlung spezialisiert ist und die Behandlung der Asthmastudie realisierte, war wieder mit im Autorenteam. Drittens wird, anders als bei allen anderen Studien, nicht die visuelle Analogskala der subjektiven Beschwerden ins Zentrum gerückt, sondern ein objektives Mass. Dieses war als Peak Flow bei der Asthmastudie zwar miterfasst, aber als Nebenzielkriterium behandelt worden und nicht signifikant gewesen. Die früheren Studien hatten alle mit der visuellen Analogskala als Hauptzielkriterium gearbeitet und damit gute Erfolge erzielt. Bei der letzten Studie, der Asthmastudie, ist genau dies bemängelt worden, so dass ein Schwenk zu einem objektiven Parameter bei der neuen Studie folgerichtig und nachvollziehbar erscheint. Auf diese Art und Weise sind in den vier Studien kleine, aber möglicherweise wesentliche Variationen enthalten: Keine Studie war wirklich wie die andere, obwohl das Grundmodell der Isopathie bei atopischen Erkrankungen gleich geblieben ist. Eine beckmesserische Kritik der Studie selber erübrigt sich meines Erachtens. Auch die Durchführung dürfte über Kritik erhaben sein. Man könnte anmerken, dass 51 Patienten etwas wenig seien. Dies ist, vom Standpunkt einer pragmatischen klinischen Studie aus gedacht, zwar richtig. Denn man geht davon aus, dass prognostisch relevante Variablen erst ab einer gewissen Studiengrösse von mehreren hundert Patienten wirklich durch Randomisation gleich verteilt werden. Diese Kritik übersieht aber, dass wir es hier nicht mit einer Therapieform zu tun haben, deren breiter Einsatz auf robuste Wirksamkeit bei allen möglichen Patienten untersucht werden soll, sondern mit einem delikaten experimentellen Modell. Hier sind so viele gut ausgewählte Patienten ausreichend, wie benötigt werden, um einen Effekt zu demonstrieren.

ren. Dies ist den Autoren gelungen. Und so scheint die Studie auf den ersten Blick einen klaren Beleg für die Wirksamkeit des isopathischen Ansatzes zu liefern, was auch durch die Meta-Analyse aller vier vorliegenden Studien erhärtet wird. Bei näherem Hinsehen jedoch ergibt sich ein leicht anderes Bild: Warum ist bei dieser Studie die visuelle Analogskala (VAS) der subjektiven Beschwerden nicht signifikant geworden? Dies ist doch höchst merkwürdig! Denn wenn sich der objektive Befund bessert, dann sollte doch auch ein subjektiver Parameter, gerade auf diesem Gebiet, Veränderungen erkennen lassen?! Zwar lässt die zusammengefasste Analyse aller vier Studien auch bei der VAS einen signifikanten Effekt zugunsten der Isopathie erkennen, aber es ist unschwer feststellbar, dass dieser Effekt trotz und nicht wegen der letzten Studie sichtbar ist. Es ist also festzuhalten, dass bei aller deutlichen Signifikanz in dieser Studie auch ein grosses Fragezeichen bestehen bleibt: Warum ist hier der objektive Parameter deutlich signifikant und der subjektive nicht?

Näher an den ursprünglichen Heuschnupfenstudien von David Reilly sind eigentlich die neueren Studien von Siri Aabel, einer norwegischen homöopathischen Ärztin, die sich auf Reillys Modell beruft. Anstelle von gemischtem Graspollen C30 verwendete sie Birkenpollen, ebenfalls in der C30, aber bei Menschen, die auf Birkenpollen allergisch reagieren. Pollendichte wurde erfasst, und alle möglichen Symptome wurden in einen klinischen Index verpackt, der den gesamthaften subjektiven Belastungszustand eines Patienten widerspiegelt. Dies ist ein Mass, das der von Reilly verwendeten VAS sehr ähnlich ist. Während Reilly seine Patienten fragte: «Alles zusammengekommen, wie sind ihre Allergiebeschwerden?», liess Aabel ihre Patienten verschiedene Allergiesymptome einstufen und fasste die Masse hinterher selber zusammen. Psychometrisch läuft beides aufs gleiche hinaus. Im einen Fall führt der Patient eine intuitive Mittelung aller Symptome aus und gibt einen individuell-intuitiven Summenscore. Im anderen Fall führt der Forscher die Mittelung über verschiedene Items durch, die der Patient eingestuft hat. Vom Standpunkt der klassischen Testtheorie aus ist letzteres Verfahren das reliablere, aber auf jeden Fall hoch korreliert mit dem der intuitiven Mittelung durch eine VAS. Die beiden Masse können also als äquivalent angesehen werden. Aabel untersuchte 66 Patienten, nicht sehr viel weniger als Reilly in der ersten Lancet-Studie und mehr als Taylor. Die Effekte waren nur sehr klein und signifikant nur an einzelnen Tagen der Studie. Wenn man eine Korrektur für multiples Testen anlegte, wären die wenigen Signifikanzen verschwunden. Würde man die Werte phasenweise zusammenfassen, wären die Effekte gering. Siri Aabel hat, das weiss ich aus persönlichen Gesprächen mit ihr, verschiedene Auswertungsstrategien überlegt und sich ausführlich mit Biometrikern besprochen. Die Ergebnisse sind sicherlich nicht deswegen so schwach, weil schlecht gemessen oder gerechnet wurde. Man hätte durch das Zusammenfassen von täglichen Werten zu Wochenmittelwerten a priori weniger Testpunkte und damit eine höhere Wahrscheinlichkeit gehabt, einen Unterschied

sichtbar zu machen, so er vorhanden ist. Das Problem ist aber wohl eher, dass der Effekt zu klein war, um mit dieser Stichprobengrösse deutlich zu werden. Ein Powerproblem also: Grössere Stichprobe, und der Effekt ist im Kasten, klassisch pharmakologisch gedacht. Diese Interpretation ist nahelegend und ihre Richtigkeit ist auch nicht ausgeschlossen. Sie ist aber unwahrscheinlich aufgrund von drei Überlegungen:

1. Die bislang sichtbaren Effekte in den Studien von Reilly waren wesentlich grösser. Es ist ein Kennzeichen erfolgreicher Replikation, dass nicht nur ein Unterschied überhaupt statistisch signifikant sichtbar gemacht werden kann, sondern dass auch die absolute Grösse des Effekts einigermaßen konstant repliziert werden kann [7]. Genau das scheint nicht der Fall zu sein.

2. Wenn das ganze ein Powerproblem wäre, dann müsste sich mindestens die numerische Grösse des Effekts im gleichen Modell replizieren lassen, auch wenn eine Nachfolgestudie nicht signifikant wird. Genau das ist aber nicht der Fall gewesen, wie wir bei der Betrachtung der zweiten Studie sehen.

3. Noch etwas verschleiert von den Wirrungen des Publikationsprozesses ist das Ergebnis einer wirklich gross angelegten Replikation von David Reillys Asthmamodell. George Lewith hat in Southampton über 300 Patienten im Rahmen einer sehr sorgfältigen Studie dokumentiert und diese Daten auf dem Kongress «Evidence Based Complementary Medicine» in München vorgestellt. Ohne der Publikation vorgreifen zu wollen, kann man doch an dieser Stelle darauf hinweisen, dass sich der von Reilly berichtete Effekt nicht replizieren liess, weder in der Richtung noch in der Grösse. Was man sehen konnte, war, dass sich die Placebo- und die Verumgruppe in verschiedenen Messgrössen unterschiedlich über die Zeit verhielten. Varianzanalytisch gesprochen war eine signifikante Interaktion sichtbar, was mindestens den Schluss nahelegt, dass es sich bei Homöopathie und Placebo nicht um die gleichen Verläufe handeln kann. Aber man wird wohl die Publikation abwarten müssen, bevor man Endgültiges sagen kann.

Auf jeden Fall schliesst letzterer Befund einigermaßen sicher aus, dass das Replikationsproblem ein blosses Powerproblem darstellt. Dies zeigt sich auch an der zweiten Birkenpollenallergiestudie von Siri Aabel. Diese ist im Titel etwas missverständlich als Prophylaxe-Studie eingeführt. Dies ist teilweise richtig, aber etwas irreführend, da nicht nur Prophylaxe betrieben wurde, sondern auch während der Pollenflugzeit täglich Arznei eingenommen wurde. Es war also eine kombinierte Prophylaxe- und Behandlungsstudie. Insofern ist also diese Studie nur ansatzweise mit der Vorgängerstudie vergleichbar. In zwei Punkten wich sie möglicherweise entscheidend von der Vorgängerstudie ab: Die Arznei wurde prophylaktisch bereits 4 Wochen vor Beginn der Pollensaison einmal wöchentlich genommen. Und die Pollenflugsaison war extrem kurz, nämlich nur 3 Tage lang. Man könnte argumentieren, dass erstens die Vorbehandlung von Allergikern mit homöopathisch zubereiteten Arzneien eine Hypersensibilisierung auslösen könnte, even-

tuell sogar verbunden mit Arzneimittelprüfungssymptomen, und dass zweitens eine solche Vorbehandlung bislang noch gar nicht untersucht wurde. In einer vergleichbaren Studie, bei der *Oscillocochinum* prophylaktisch gegen Grippe eingesetzt worden sollte, ergab sich ebenfalls ein solches paradoxes Phänomen, dass die Verumgruppe mehr Nebenwirkungen hatte als die Placebogruppe [1]. Die leichte Tendenz zugunsten von Placebo in der Birkenpollenstudie von Aabel würde in dieses Bild passen. Wie auch immer: Die Daten erlauben uns nicht, zu argumentieren, dass die fehlende Signifikanz der ersten Studie nur ein Powerproblem darstelle. Es ist offensichtlich komplizierter. Möglicherweise hätten wir einen Effekt gesehen, wenn in der zweiten Studie nicht prophylaktisch, sondern nur therapeutisch gearbeitet worden wäre? Möglicherweise sind die besseren Werte unter Placebo zu verstehen als das Abbild einer Arzneimittelprüfung, die durch die prophylaktische Einnahme von *Betula C30* aufgetreten ist? Man sollte allerdings über solchen Überlegungen nicht vergessen, dass die Studien von Aabel der reflektierte Versuch waren, eine in Norwegen sehr übliche therapeutische Praxis, mit der die Autorin und andere Homöopathen gute Erfahrungen gemacht haben, als Forschungsmodell zu verwenden. Dieser Versuch ist offensichtlich fehlgeschlagen. Die Erwartungen der Forscher sind sicherlich ganz andere gewesen. Sie konnten aufgrund ihrer Erfahrungen und aufgrund der Ergebnisse der Vorgängerstudie, einschliesslich derjenigen von David Reilly, erwarten, einen positiven Effekt zugunsten der Homöopathie zu finden. Deshalb scheint mir die zweite Studie, bei aller Hoffnung, die man bei einer isolierten Betrachtung hegen könnte, doch ein klares Verdikt zu enthalten: Das isopathische Modell ist eigentlich, ausserhalb der Hände David Reillys, offenbar nicht replizierbar. Dieser Verdacht erhärtet sich, wenn man alle drei Studien und die noch nicht publizierte von George Lewith nebeneinanderlegt und kontrastierend überdenkt. Ein strikter Kritiker würde das Ergebnis der Taylor-Studie nicht anerkennen, weil sich das bisher verwendete Zielkriterium, die VAS, nicht als veränderungssensitiv erweist. Dazu kommt der Mangel an deutlichem Effekt in der ersten Birkenpollenstudie und der eigentlich negative in der zweiten. Zusammengefasst mit dem Befund der südenglischen Asthma-Studie eines fehlenden klinischen Unterschieds bei gemischtem Verlauf der Gruppen: Mit Verlaub, aber Homöopathie scheint wirklich mehr Ähnlichkeit mit Spuk aufzuweisen als mit Pharmakologie [9].

Literatur

- 1 Attena F, Toscano G, Aggozzino E, Del Giudice N: A randomized trial in the prevention of influenza like syndromes by homeopathic treatment. *Rev Epidemiol Santé Publique* 1995;43:380–382.
- 2 Brigo B, Serpelloni C: Homeopathic treatment of migraines: A randomized double-blind controlled study of sixty cases (homeopathic remedy versus placebo). *J Res Homeopathy* 1991;1:98–106.
- 3 Diener HC, Brune K, Gerber W-D, Pfaffenrath V, Straube A: Therapie der Migräneattacke und Migräneprophylaxe. Empfehlungen der Deutschen Migräne- und Kopfschmerzgesellschaft (DMKG). *Aktuelle Neurol* 2000;27:273–282.

- 4 Reilly D, Taylor MA, Beattie NGM, Campbell JH, McSharry C: Is evidence for homeopathy reproducible? *Lancet* 1994;344:1601–1606.
- 5 Reilly DT, Taylor MA, McSharry C, Aitchinson T: Is homeopathy a placebo response? Controlled trial of homeopathic potency with pollen in hayfever as a model. *Lancet* 1986;Oct.18:881–886.
- 6 Straumsheim PA, Borchgrevink CF, Mowinkel P, Kierulf H, Hafslund O: Homeopatisch behandling av migræne. En dobbelt-blind, placebonkontrollert studie av 68 pasienter. *Dynamis* 1997;2:18–22.
- 7 Utts J: Replication and meta-analysis in parapsychology. *Stat Science* 1991;6: 363–404.
- 8 Walach H, Gaus W, Haeusler W, Lowes T, Mussbach D, Schamell U, Springer W, Stritzl W, Haag G: Classical homeopathic treatment of chronic headaches. A double-blind, randomized, placebo-controlled study. *Cephalalgia* 1997;17: 119–126.
- 9 Walach H: Magic of signs: A non-local interpretation of homeopathy. *Br Homeopathic J* 2000;89:127–140.
- 10 Whitmarsh TE, Coleston-Shields DM, Steiner TJ: Double-blind randomized placebo-controlled study of homeopathic prophylaxis of migraine. *Cephalalgia* 1997;17:600–604.

Kontaktadresse: PD Dr. Dr. phil. Harald Walach, Universitätsklinikum Freiburg, Institut für Umweltmedizin und Krankenhaushygiene, Hugstetterstrasse 55, D-79106 Freiburg, E-mail walach@ukl.uni-freiburg.de

Kommentar – Walter v. Lucadou, Freiburg

Obgleich die Homöopathie in relativ grossem Umfang angewendet und akzeptiert wird und über eine gewisse Institutionalisierung innerhalb der Medizin verfügt und die Parapsychologie all dies nicht vorweisen kann, weisen beide Fächer doch eine Reihe von phänomenologischen, methodischen, theoretischen und wissenschaftssoziologischen Gemeinsamkeiten auf, die es wert sind, in Betracht gezogen zu werden. Da die Homöopathie allerdings selbst um ihre theoretische Begründung ringen und um ihren Status im Verschreibungswesen der Medizin kämpfen muss, wird sie über eine Sekundanz der Parapsychologie kaum erfreut sein, gilt die Parapsychologie doch gemeinhin als wissenschaftlich unseriös oder gar anti-wissenschaftlich. Allerdings ist es im allgemeinen kaum bekannt, dass es der wissenschaftlichen Parapsychologie in den letzten 15 Jahren gelungen ist, nicht nur in experimenteller, sondern vor allem in theoretischer Hinsicht bemerkenswerte Fortschritte zu erzielen, die auch für die Homöopathie von Bedeutung sein könnten. Dabei zeigt sich, dass die «paranormalen» Phänomene viel weniger «para»-normal sind, als es den Anschein hat.

Unter Parapsychologie [vgl. 1] wird die Anwendung empirischer Forschungsstrategien auf Erlebnisse und Verhaltensweisen des Menschen verstanden, die aus dem bisher bekannten Erklärungsrahmen der etablierten Disziplinen von Psychologie, Biologie und Physik herauszufallen scheinen. Traditionsgemäss umfasst die experimentelle Parapsychologie hauptsächlich zwei Forschungsbereiche:

1. Unter dem «kognitiven» Aspekt wird als »ausser sinnliche Wahrnehmung« (ASW) die Frage untersucht, ob und unter welchen Bedingungen Menschen in der Lage sind, Informationen ausserhalb bisher bekannter und definierter sensorischer Kanäle aufzunehmen und/oder abzugeben;
2. unter dem

«motorischen» Aspekt als «Psychokinese» (PK) wird die Frage untersucht, ob und unter welchen Bedingungen Menschen eine direkte «psychische» Wirkung auf physikalische Systeme ausüben können, die den bisher bekannten bzw. akzeptierten naturwissenschaftlichen Erklärungsmodellen zu widersprechen scheint.

Wie auch in der obigen Homöopathiestudie wurden in letzter Zeit die bisherigen Experimente der Parapsychologie Metaanalysen unterzogen, die die Frage klären, ob die experimentellen Ergebnisse «robuste Effekte» sind oder möglicherweise experimentelle Artefakte darstellen, die durch methodische Fehler und/oder Selektion zustande gekommen sind. Aufgrund dieser Studien werden folgende Merkmale heute weitgehend als gesichert angesehen:

(1) ASW und PK sind unter Laborbedingungen statistisch-experimentell nachgewiesene «robuste» Effekte. (2) Es handelt sich dabei um eine allgemein verbreitete, aber unterschiedlich ausgeprägte «Disposition» des Menschen, die als Psi bezeichnet wird. (3) Psi hängt von unbewussten Faktoren ab, tritt sporadisch und unvorhersehbar auf und ist nicht beliebig steuerbar. (4) Äussere physikalische Parameter (wie Beschaffenheit der Zielobjekte, räumliche und zeitliche Distanz, Abschirmungen usw.) spielen kaum eine Rolle. (5) Psychologische Variablen (überdauernde Persönlichkeitsmerkmale wie Extraversion oder emotionale Stabilität sowie die Einstellung der Vpn) korrelieren signifikant mit der «Psi-Leistung». (6) ASW und PK sind strukturell verwandt.

Dieser Merkmalskatalog, der aus hunderten experimenteller Untersuchungen destilliert wurde, stösst nicht nur auf Widerspruch seitens der Kritiker der Parapsychologie, sondern unterliegt auch einer Revidierung aufgrund theoretischer Überlegungen. Während die Rhinesche Schule hauptsächlich von Vorstellungen der klassischen Physik ausging («implizite Modelle» in der Art von Sender – Kanal – Empfänger, Reiz-Reaktions-Schema), zeichnet sich in den letzten Jahren ein neuer theoretischer Zugang zum Verständnis von Psi-Phänomenen ab. Ausgehend vom viel diskutierten «Messproblem in der Quantenphysik» wurden unterschiedliche Modellansätze entwickelt, die sogenannten «observational theories» (OTs). Diese weisen eine Reihe von Übereinstimmungen auf: (1) Die Verletzung der «üblicherweise» geltenden Naturgesetze ist schwach, d. h., Erhaltungssätze und Symmetrien werden nicht verletzt, nur quantenphysikalisch-stochastische Prozesse können «beeinflusst» werden; (2) ASW kann auf PK zurückgeführt werden; (3) der PK-Effekt ist zumindest in einem näher zu bestimmenden Rahmen Raum-/Zeit-unabhängig (nichtlokal); (4) PK «funktioniert» zielorientiert im Hinblick auf die gegebene Instruktion, wobei Feedback über den Erfolg einen essentiellen Bestandteil des Prozesses darstellt. Die OTs erlauben spezifische quantitative Vorhersagen über die fraglichen Psi-Effekte, können also experimentell überprüft werden.

Die letzten beiden Punkte können sicher auch für die Homöopathie in Anspruch genommen werden. Vor allem die Notwen-

digkeit von Feedback bringt Schwierigkeiten beim Doppelblinddesign mit sich, wie wir weiter unten sehen werden.

In bezug auf die Homöopathie scheint vor allem das «Modell der Pragmatischen Information» (MPI), eine systemtheoretische Version der OTs [vgl. 4, 5], von Interesse zu sein. Es kann in zwei «Hauptsätzen» dargestellt werden:

Erster Hauptsatz: Psi-Phänomene sind nichtlokale Korrelationen in psycho-physikalischen Systemen, die durch die pragmatische Information, die das (organisatorisch geschlossene) System erzeugt, induziert werden.

Eine der wichtigsten Annahmen des MPI besteht darin, dass selbstorganisierende psycho-physikalische Systeme in der Lage sind, organisch geschlossene Einheiten («organizational closure») zu bilden. Das Phänomen der «organisierten Geschlossenheit» kann nach Varela [7] geradezu zur Definition von lebenden Systemen benutzt werden. Gleichzeitig sind lebende Systeme nur in «thermodynamisch offenen Systemen» möglich. Dabei sind nichtlokale Korrelationen gleichzeitig Indizien und Konstituenten dieser Einheit. Die organische Geschlossenheit kann man sich in Analogie zur Physik wie die «Geschlossenheit» eines Atoms oder Moleküls bzw. einer Zelle oder eines Individuums vorstellen. Dabei spielen zwar »klassische« Anziehungskräfte durchaus eine wichtige konstituierende Rolle, die «Einheit» oder die Stabilität des Gebildes wird jedoch in erster Linie von «quantenmechanischen» Wechselwirkungen, den sogenannten «Austauschkraften», erzeugt, die kein klassisches Analogon besitzen. Sie entsprechen phänomenologisch nichtlokalen Korrelationen. Die Geschlossenheit des Systems drückt sich in der Quantentheorie durch Erhaltungsgrößen oder Symmetrien aus (z. B. Energieerhaltung, Drehimpulserhaltung usw.). Bei komplexen nichtklassischen Systemen spielt die «Pragmatische Information» die Rolle einer solchen Erhaltungsgrösse. «Pragmatische Information» ist ein Mass für die «Bedeutung» einer Information für – oder genauer ein Mass für die «Wirkung» einer Information auf – ein System. Daraus folgt, dass der Aufbau oder die Zerlegung eines organisiert geschlossenen Systems pragmatische Information «verbraucht» bzw. «erzeugt», um die nichtlokalen Korrelationen innerhalb des Systems «aufzubauen» bzw. «aufzulösen». «Psycho-physikalisch» sind aber alle diejenigen Systeme, für deren Beschreibung gleichermassen die Sprache der Physik wie diejenige der Psychologie benötigt wird, die also ohne fächerübergreifende Beschreibung nicht untersucht werden können. Hierzu gehört das bekannte «Leib-Seele-Problem» oder – so wollen wir hier annehmen – auch die homöopathische Behandlung.

Die «nichtlokalen Korrelationen» haben allerdings eine Reihe von merkwürdigen Eigenschaften, die den Physikern grosses Kopfzerbrechen bereiten und die bisweilen sogar als «paradox» oder «geisterhaft» bezeichnet werden. Trotzdem glaubt natürlich kein ernstzunehmender Physiker, dass es hier mit «übernatürlichen Dingen» zugeht, ganz im Gegenteil: Nicht-

lokale Korrelationen beschreiben die fundamentalen Eigenschaften der Materie.

Eine nichtlokale Korrelation ist ein Zusammenhang zwischen Ereignissen, ohne dass dabei eine Informationsübertragung beteiligt wäre. Dieser Zusammenhang ist «flüchtig», was im zweiten Hauptsatz des MPI ausgedrückt wird:

Zweiter Hauptsatz: Jeder Versuch, nichtlokale Korrelationen zur Signalübertragung zu verwenden, bringt diese zum Verschwinden oder ändert sie in unvorhersagbarer Weise.

Auf die «Verwertung» paranormaler Phänomene angewendet, heisst das allerdings: In dem Moment, wo man sich auf eine «Technik» wirklich verlässt, funktioniert sie nicht oder etwas anderes geschieht, als man erwartet. Das heisst nicht, dass es keine «paranormalen Phänomene» geben kann, nur kann man sich nicht auf sie verlassen, d.h., sie sind statistisch nicht reliabel. In der Parapsychologie weisen in der Tat fast alle statistischen Experimente einen deutlichen «Decline effect» (Absinkungseffekt) auf, d.h., bei (identischen) Replikationen zeigt sich, dass die Effektstärke abnimmt [vgl. 2]. Hier tritt nun eine erstaunliche Parallele zur Homöopathie zutage.

Es ist sicher nicht falsch, in der homöopathischen Behandlung einen komplexen Kommunikationsvorgang zwischen dem Arzt, dem Patienten und dem Medikament zu sehen. Manche theoretischen Modelle der Homöopathie nehmen sogar an, dass die Information über die Wirkung im Lösungsmittel der Verdünnung gespeichert sei. Im MPI muss man aber so weit gar nicht gehen.

Nach meiner Auffassung geht es hier in erster Linie um das therapeutische Ritual, das man aber als Gesamtsystem betrachten muss. Es macht also einen Unterschied, ob das Medikament einem meist langwierigen (!) Herstellungsprozess unterzogen wird oder ob bloss ein Placebo gegeben wird. Ich habe hierfür den Begriff der «Pseudomaschine» geprägt. Pseudomaschinen haben eine echte Wirkung, sie wirken gewissermassen als ein Gefäss für nichtlokale Zusammenhänge, die sich als kausale Zusammenhänge nicht festmachen lassen. Pseudomaschinen zeigen zwar einen realen Effekt, aber der Wirkungsmechanismus ist ein anderer, als die «Erbauer der Maschine» selber glauben. Dies ist eine durchaus wichtige Bedingung, weil sie die Anwender davon abhalten, den eigentlichen Wirkungsmechanismus zu durchschauen. Die Anwender gehen nämlich von lokalen Effekten – also Wirkungskausal-ketten – aus, die dann in Doppelblindstudien nachweisbar sein sollten. Wenn es sich jedoch beim Wirkungsmechanismus um ein nichtlokales Phänomen handelt, sind zwar Doppelblindstudien immer noch möglich, allerdings zeigt sich hierbei die oben erwähnte Elusivität und man wird eine Abnahme der gewünschten Wirkung erwarten. Wenn dagegen der vermeintliche kausale Wirkungszusammenhang der Pseudomaschine auch noch operational (also faktisch) verhindert, dass der Anwender die nichtlokalen Korrelationen zur Signalübertragung verwenden kann, ist die Wirkung der Pseudomaschine optimal, weil sie den zweiten Hauptsatz des MPI nicht ver-

letzen kann. Dies ist allerdings bei Doppelblindstudien gerade nicht der Fall – in der homöopathischen Praxis schon eher.

Man kann sich das auch so veranschaulichen, dass die Nichtlokalität etwas Fluides, etwas Flüssiges darstellt, das sich nicht «festhalten» lässt. Pseudomaschinen haben in diesem Bild die Funktion eines Schwammes, der die Flüssigkeit aufsaugt und sie zur Anwendung bringt. Ohne den Schwamm kann man mit der Flüssigkeit nicht putzen, sie läuft davon. Es wäre aber ein Fehler zu glauben, dass der Schwamm allein schon die gewünschte Putzwirkung erzielt. Für eine optimale Wirkung braucht man beides, das «flüssige Element» der Nichtlokalität und das «feste Element» der Pseudomaschine. Bei Doppelblindstudien trocknet der Schwamm allerdings zu schnell aus, weil der zweite Hauptsatz des MPI verletzt wird. Bei einer Replikationsstudie könnte man nämlich mit der Kenntnis, die man im vorherigen Experiment gewonnen hat, gegen den zweiten Hauptsatz verstossen, indem man ein Signal codiert (z. B. Besserung = Medikament = 1 und keine Besserung = Placebo = 0).

Man kann diesen Zusammenhang sehr schön in der Übersicht, die in der obigen homöopathischen Studie angegeben ist, verfolgen: In den vier vergleichbaren Studien von Reilly et al., die auch als Replikationsstudien aufgefasst werden können, weil sie untereinander sehr homogen sind, werden im wesentlichen zwei abhängige Variablen zur Messung des therapeutischen Effekts verwendet: 1. die subjektive visuelle Analogskala und 2. unterschiedliche objektive Masse, wie Histaminwerte oder Nasal inspiratory peak flow. Vom theoretischen Standpunkt aus betrachtet ist es von Bedeutung, dass eine für alle Studien vergleichbare Variable (1) und eine alternative nichtvergleichbare Variable (2) verwendet wurden.

Sollte es sich bei der homöopathischen Wirkung um eine nichtlokale Korrelation im obigen Sinne handeln, so wären vom MPI folgende Voraussagen zu machen:

1. Der therapeutische Effekt, gemessen mit der vergleichbaren Variablen (1), wird bei Replikationsstudien in dem Masse zurückgehen, wie die statistische Reliabilität dieser Variablen aufgrund der Zunahme der Stichprobengrösse (n) steigt.
2. Der therapeutische Effekt, gemessen mit der nichtvergleichbaren Variablen (2) wird bei Replikationsstudien zunehmen, also gewissermassen das Absinken der Effektstärke in Variable (1) für das Gesamtergebnis kompensieren, weil die Sicherheit, in weiteren Studien eine Voraussage über diese Variable (2) zu machen, nicht mit n zunimmt.

Wegen der Homogenität der Studien kann angenommen werden, dass die organisierte Geschlossenheit des Gesamtsystems und damit seine nichtlokale Verschränkung über die vier Studien als einigermaßen konstant angesehen werden kann.

Diese Zusammenhänge sind in der Tat in Fig. 3 der Taylor-Publikation abzulesen. Die subjektive visuelle Analogskala zeigt über den Verlauf der 4 Studien eine fast monotone Abnahme zum Erwartungswert (Spalte: Pooled analysis). Es kann sich hierbei allerdings nicht um die bekannte «Regression zur Mitte» handeln, weil die Abnahme erst in der 4. Studie

den Erwartungswert erreicht. Das MPI würde allerdings für eine weitere Studie erneut eine Zunahme des Effekts zulassen, die jedoch die Effektstärke des Gesamtergebnisses von Variable (1) (Composite) nicht überschreiten darf. Durch den «negativen» Ausgang für die Variable (1) in der 4. Studie werden gewissermassen «die Karten neu gemischt», durch die immerhin noch signifikante Gesamtabweichung sind jedoch noch einige «Erwartungen» vorhanden, die den Effekt weiter limitieren.

Auch für die Spalte: Main objective measures werden die Voraussagen des Modells erfüllt. Hier liegen allerdings nur Werte für 3 Studien vor. Es zeigt sich jedoch, dass der statistische Unterschied zwischen der Verum- und der Placebogruppe von der 2. bis zur 4. Studie zunimmt, aber erst bei der 4. Studie signifikant wird. Die Variable (2) kompensiert gewissermassen das Ausbleiben des Effekts bei Variable (1). Natürlich sind die Verhältnisse in Wirklichkeit noch etwas verwickelter, weil zur Validierung der Ergebnisse auch noch weitere Beobachtungen wie z.B. die grössere Erstverschlimmerung bei der Verumgruppe eine Rolle spielen. Insgesamt lässt sich jedoch sagen, dass die experimentellen Ergebnisse unter der Betrachtungsweise des MPI plausibel erscheinen. Unter einem lokal-kausalen Wirkungsmodell sind die Resultate dagegen nur durch Post-hoc-Annahmen über einen möglichen Mangel an Homogenität der Studien erklärbar. Es sollte vielleicht noch erwähnt werden, dass die Aussagen des MPI echte »Voraussagen« waren und nicht auf die vorliegende Studie »zugeschnitten« wurden [vgl. 3].

Literatur

- 1 Bauer E, von Lucadou W: Parapsychologie, in Asanger R, Wenninger G (Hrsg): Handwörterbuch Psychologie, pp 517–524. München, Verlags Union; 1988.
- 2 Bierman DJ: The nature of anomalous phenomena: Another reality between the world of subjective consciousness and the objective world of physics? in van Looeke (ed): The Physical Nature of Consciousness. New York, Benjamins, in press, 2000.
- 3 Von Lucadou W: Nichtlokale Korrelationen im psychologischen experiment – zum Problem der Replizierbarkeit. Positionspapier zum Expertensymposium der Carl-und-Veronika-Carstens-Stiftung: «Die Bedeutung nicht-lokaler Effekte für die klinische Forschung» am 23. Februar 1995 in München-Garching (1994).
- 4 Von Lucadou W: The Model of Pragmatic Information (MPI). Eur J Parapsychol 1995;11:58–75.
- 5 Von Lucadou W: Psyche und Chaos – Theorien der Parapsychologie. Frankfurt a. M., Insel-Verlag, 1995.
- 6 Von Lucadou W, Grösser H: Interview: »Psi-Phänomene, Homöopathie, Pseudomaskinen«. COMED 1998, Nr. 6, pp 60–63.
- 7 Varela FJ: Autonomy and autopoiesis; in Roth G, Schwengler H (eds): Self-Organizing Systems, pp 14–23. Frankfurt, Campus, 1981.

Kontaktadresse: Dr. rer. nat. Dr. phil. Walter v. Lucadou, Parapsychologische Beratungsstelle, Hildastrasse 64, D-79102 Freiburg, E-mail lucadou@freenet.de

Adressen für Sonderdrucke: David Reilly, Academic Departments, Glasgow Homoeopathic Hospital, Glasgow G12 0XQ, UK; E-mail davidreilly1@compuserve.com
Siri Aabel, Institute of General Practice and Community Medicine, Department of General Practice, PO Box 1130, Blindern, N-0317 Oslo, Norway

Fernheilung bei Warzen?

Harkness EF, Abbot NC, Ernst E: A randomized trial of distant healing for skin warts. Am J Med 2000;108:448–452.

Purpose: Distant healing, a treatment that is transmitted by a healer to a patient at another location, is widely used, although good scientific evidence of its efficacy is sparse. This trial was aimed at assessing the efficacy of one form of distant healing on common skin warts.

Subjects and methods: A total of 84 patients with warts were randomly assigned either to a group that received 6 weeks of distant healing by one of 10 experienced healers or to a control group that received a similar preliminary assessment but no distant healing. The primary outcomes were the number of warts and their mean size at the end of the treatment period. Secondary outcomes were the change in Hospital Anxiety and Depression Scale and patient's subjective experiences. Both the patients and the evaluator were blinded to group assignment.

Results: The baseline characteristics of the patients were similar in the distant healing (n=41) and control groups (n=43). The mean number and size of warts per person did not change significantly during the study. The number of warts increased by 0.2 in the healing group and decreased by 1.1 in the control group (difference [healing to control] = -1.3; 95% confidence interval) -1.0 to 3.6, p=0.25). Six patients in the distant healing group and 8 in the control group reported a subjective improvement (p=0.63). There were no significant between-group differences in the depression and anxiety scores.

Conclusions: Distant healing from experienced healers had no effect on the number or size of patient's warts.

Kommentar – M. Binder, Freiburg

Schade, dass eine weitere Untersuchung in die Kategorie derjenigen Studien eingereiht werden muss, welche die Nutzlosigkeit der Fernheilung als einer komplementären Behandlungsform nahelegen.

Die Studie besticht durch ein einfaches Design mit klarer Fragestellung und einer gut operationalisierbaren Zielvariable. Leider verbietet die spärliche Beschreibung vieler, möglicherweise entscheidender Details Interpretationen der Ergebnisse von vornherein. Bei der Darstellung einer Untersuchung wie dieser, wo die Wissenschaft noch so im dunklen tappt, wo Spekulationen und noch nicht erhärtete Vermutungen notwendiger Bestandteil von Planung, Durchführung und Auswertung sind, in einer Studie, wo viele sonst unwichtige Parameter von entscheidender Bedeutung sein können, vermissemir viele Angaben von «Nebensächlichkeiten», deren Kenntnis mir für das Verständnis der Studie und für die Interpretation ihrer Ergebnisse unabdingbar notwendig erscheint.

So wüsste ich über die Heiler(innen) und ihre Behandlung gerne folgendes: War die Art ihrer Behandlung miteinander vergleichbar? Oder betete der eine für seine Patientin, während die andere ein schamanisches Ritual durchführte? Wie lange behandelten die Heiler(innen) ihre Patienten? Es gibt Heiler, die davon überzeugt sind, dass eine Fernbehandlung nur wenige Minuten in Anspruch nimmt, während andere 1½-2 Stunden für eine Patientin aufwenden. Eine Heilerin behandelt ihre Patientin mehrmals täglich, während ein anderer dieselbe Wirkung mit einer Behandlung einmal pro Woche erzielen zu können meint. Behandelten die Heiler(innen) unterschiedlich oft oder unterschiedlich lange? War es ihnen freigestellt? Wenn die Patienten gleichmässig auf die Heiler(innen) verteilt waren, hatte jede(r) 4 Patienten zu behandeln. War sichergestellt, dass die Behandlungen gleichmässig auf die Patienten verteilt wurden? Führten die Heiler(innen) ein Behandlungsprotokoll, in das sie eintrugen, an welchen Tagen und wie lange sie jeweils Behandlungen durchführten? Wurden die Heiler(innen) besucht, die nicht bereit waren, ihre persönlichen Daten preiszugeben? Wie wurde sichergestellt, dass es sich bei ihnen nicht um «schwarze Schafe» handelt, die es ja auch in der Confederation of Healing Organizations geben dürfte?

Bekamen die Heiler(innen) die Patientendaten direkt zum Behandlungsbeginn zugesandt? Wurde die Nacherhebung direkt nach der Behandlungsphase durchgeführt? Oder wurde sichergestellt, dass die Heiler(innen) nicht (vielleicht unwillentlich) früher als geplant oder über die angegebene Behandlungsdauer hinaus behandelten? Waren auf den Skizzen nur die drei Zielwarzen angegeben oder noch weitere, auf die sich die Heiler(innen) vielleicht konzentrieren konnten? Wurde überprüft, ob nicht der/die eine oder andere Heiler(in) vielleicht Kontakt zu Patienten aufgenommen hat?

Wurden die Patienten mehrmals gesehen oder lediglich zweimal? Wie wurden sie instruiert, offen zu sein für eine Behandlungsform, die sie vielleicht gar nicht bekamen? Gab es eine schriftliche Patienteninformation oder hatten sie Gelegenheit, Fragen und Unsicherheiten, die mit einer derart exotischen Behandlungsform verbunden sind, zu klären? Schon allein die Unsicherheit, ob man nun dabei ist oder nicht, kann ja eventuell einen Behandlungseffekt hemmen.

Gerade bei Studien wie einer solchen, die (zu Recht) viele Skeptiker auf den Plan ruft, sollten solche Massnahmen nicht nur in die Projektplanung und -durchführung eingehen, sondern im Ergebnisbericht detailliert beschrieben und diskutiert werden.

Da wir über mögliche Langzeitwirkungen einer Fernheilungsbehandlung noch kaum etwas wissen, wäre es möglicherweise aufschlussreich gewesen, die Patienten nach 12 oder 18 Wochen noch einmal zu untersuchen.

Wenn auch Grösse und Anzahl der Warzen bestehend einfache Zielvariable darstellen, so dürfte doch der Leidensdruck der Patienten in aller Regel nicht mit dem von Herzpatienten oder Patienten mit chronischen Schmerzen vergleichbar sein.

Nicht wenige Heiler(innen) messen der subjektiven Bedeutung, die eine Besserung der Beschwerden oder Heilung einer Krankheit für den Patienten hat, eine entscheidende Bedeutung bei. In einer Studie wie dieser hätte die Frage nach der persönlichen Bedeutung einer Heilung oder Besserung am Gesamtergebnis nichts geändert. Die Antworten hätten jedoch möglicherweise eine Interpretationshilfe angeboten.

Dieser Kommentar ist nicht der eines Wissenden, sondern eines Fragenden. Ich bin ausschliesslich auf inhaltliche Fragen eingegangen und habe bewusst auf statistische Einwände und Vorschläge verzichtet, weil ich beim aktuellen Wissensstand die inhaltlichen Überlegungen für ungleich wichtiger halte. Ich habe viele Fragen gestellt, die mir den Vorwurf des Erbsenzählens einbringen dürften. Ich nehme dies in Kauf, denn ich bin der Überzeugung, dass die minutiöse Planung und liebevolle Durchführung einer solchen Studie unabdingbare Voraussetzungen dafür sind, dass Effekte aufscheinen und in den Blick genommen werden können.

Es ist zu hoffen, dass sich immer mehr Studien dieser interessanten und höchst relevanten Behandlungsform neugierig und unvoreingenommen annehmen und dazu beitragen, die Wahrheit von der blossen Überzeugung und die Fakten von den Mythen zu unterscheiden.

Kontaktadresse: Dipl.-Psych. Markus Binder, Institut für Grenzgebiete der Psychologie und Psychohygiene, Wilhelmstrasse 3a, D-79098 Freiburg, E-mail binder@igpp.de

Adresse für Sonderdrucke: Prof. Edzard Ernst, Department of Complementary Medicine, School of Postgraduate Medicine and Health Sciences, University of Exeter, 25 Victoria Park Road, Exeter, EX1 2ET, UK

Empirische Testung der Theorie der morphischen Resonanz – Können wir entdecken, wenn wir angeblickt werden?

Sheldrake R: The 'Sense of Being Stared At' confirmed by simple experiments. Riv Biol / Biol Forum 1999;92:53–76.

Sheldrake R: The 'Sense of Being Stared At' does not depend on known sensory clues. Riv Biol / Biol Forum 2000;93:237–252.

Hintergrund: Die Theorie der morphogenetischen Felder oder der morphischen Resonanz sagt die Existenz von Feldern voraus, die durch bewusste Intention oder häufige Wiederholung aufgebaut werden und die mit materiellen Prozessen wechselwirken. Aus dieser Theorie leitet sich die Vorhersage ab, dass Intentionen, vor allem wenn sie ein biologisch bedeutsames Muster betreffen und phylogenetisch alt sind, direkt mit materiellen Prozessen wechselwirken. Das «Angeblicktwerden», vor allem von hinten, ist eine evolutionsbiologisch äusserst bedeutsame Tatsache, die eine Intention, vor allem in der

Jäger-Beute-Beziehung, ausdrückt. Deswegen ist zu erwarten, dass Menschen, die von hinten angeblickt werden, dies spüren, auch wenn kein direkter physikalischer Kontakt besteht.

Methode: In einer ersten Versuchsserie von Experimenten wurden Schulkinder in Paaren hintereinander gesetzt im Abstand von 1 m und mehr. Die Kinder arbeiteten klassenweise. Die Betrachter hatten die Aufgabe, nach einem vorgegebenen Randomisationsplan ihre Vordermänner, die mit dem Rücken zu ihnen sassen, entweder anzublicken oder aber wegzuschauen und an etwas anderes zu denken. Der Beginn einer Versuchsreihe wurde zunächst mit einem Stift, später dann mit einem mechanischen Klicker angezeigt. Anschliessend mussten die angeblickten Kinder angeben, ob sie das Gefühl hatten, angeschaut zu werden oder nicht, was spätestens nach 20 Sekunden zu geschehen hatte. Dann folgte der nächste Durchgang. In der ersten Versuchsreihe sassen die Kinder im selben Raum und erhielten jeweils Rückmeldung über die Richtigkeit ihres Raten, in der zweiten waren sie durch eine Fensterscheibe getrennt, waren mit einer Tuchbinde verblindet und erhielten keine Rückmeldung. Anschliessend wurden die Rollen getauscht und man begann mit einer neuen Serie.

Setting: In der ersten Studie waren insgesamt 584 Personen, meist Schüler und Studenten, in verschiedenen Schulen und Colleges Englands, Deutschlands und der USA. In der zweiten Studie waren 264 Schüler aus England, Kanada, den USA und Deutschland. Die Testung erfolgte klassenweise.

Zielkriterium: Das richtige Erraten von Angeschautwerden oder Nicht-Angeschautwerden.

Ergebnisse: Im ersten Experiment waren insgesamt 54,9% der Antworten richtig, 59% beim Erraten des Angeschautwerdens, 50,7% beim Erraten des Nicht-Angeschautwerdens. In der zweiten Serie von Experimenten waren 52,6% Treffer zu verzeichnen, 54,1% beim Erraten des Angeschautwerdens, 51,1% beim Erraten des Nicht-Angeschautwerdens. Die Replikationen in den USA, Deutschland und Kanada erbrachten insgesamt 53,7% Treffer, 57,2% beim Erraten des Angeschautwerdens, 50,3% beim Erraten des Nicht-Angeschautwerdens. Diese Ergebnisse weichen jedesmal signifikant von der Zufallserwartung von 50% ab für das Gesamtergebnis und für das Erraten des Angeschautwerdens, nicht jedoch für das Erraten des Nicht-Angeschautwerdens.

Schlussfolgerung: Die Ergebnisse bestätigen die Theorie der morphischen Resonanz, weil das überzufällig richtige Erraten vor allem auf das biologisch relevante Angeschautwerden zutrifft, nicht aber auf das Nicht-Angeschautwerden.

Kommentar – S. Schmidt, Freiburg

Sheldrake beschreibt in den beiden Aufsätzen verschiedene Experimente zu «Remote Staring», einem parapsychologischen Standardexperiment, bei dem es um die Frage geht, ob Personen merken, wenn sie von jemand anderem angeschaut werden. Wenn man die Daten der berichteten Experimente

ganz grob zusammenfasst, erhält man 54,3% korrekte Rate-treffer bei 10 663 Versuchsdurchgängen für den Aufsatz von 1999 (ein Experiment ist wegen methodischer Fehler nicht berücksichtigt) und 53,0% Treffer bei 4800 Versuchsdurchgängen für den 2000er-Aufsatz, bei einer Zufallserwartung von 50%. Die Verringerung des Effekts von 1999 zu 2000 könnte von einigen methodischen Verbesserungen herrühren, die in der späteren Studie aufgrund der Kritik von Skeptikern eingeführt worden sind. In der ersten Studie waren die Angeschauten und die Anschauenden im selben Raum. Hingegen waren sie in der zweiten Studie durch ein Fenster getrennt, um sensorische Hinweisreize auszuschalten. Aus den gleichen Gründen waren die Angeschauten in der zweiten Studie ausserdem mit einem Tuch verblindet. Ferner gab es in der zweiten Studie keine unmittelbare Rückmeldung für die Angeschauten, im Gegensatz zu einigen Experimenten der ersten Serie, wo die Angeschauten nach jedem Raten eine Rückmeldung über die Korrektheit ihrer Vermutung erhielten.

Das Vermeiden sensorischer Hinweisreize und unmittelbarer Rückmeldung gehört zu den wichtigsten Grundzügen wohlkontrollierter «Remote Staring»-Experimente. Sheldrake sagt zwar, «auditory and olfactory clues made very little difference to the subjects' performance» (p. 248), und meint, die Motivation von Versuchspersonen «were generally higher when subjects are given feedback... which could have encouraged them and helped them to take more interest in the experiment» (p. 248). Dies zeigt aber, dass er sich nicht im klaren über die starken Effekte ist, die durch Informationsübertragung mittels verborgener Hinweisreize und subliminaler Wahrnehmung zustande kommen können. Besonders in diesen Experimenten, wo hochmotivierte Schüler versuchen, die bestmöglichen Resultate zu erzielen, dürfte das Suchen nach verborgenen Hinweisreizen besonders stark sein. Und selbstverständlich ist die Identifikation von hilfreichen Hinweisreizen – wie etwa eine bestimmte Art des Raschelns, oder ein Lichtreflex – um einiges erleichtert, wenn unmittelbares Feedback diese Identifikation bestätigt.

Die frühere Studie dürfte also eine leichte Überschätzung des Effekts enthalten. Jedoch erscheint dieses Artefakt in der zweiten Studie (Sheldrake, 2000) ausgeschaltet. Stellen die Ergebnisse von 53% richtiger Trefferrate einen echten Effekt dar oder gibt es hier noch andere mögliche Artefakte? Ein Problem in beiden Studien sind die Zufallssequenzen, nach denen diejenigen Personen handelten, welche die andere Person anschauen sollten oder nicht. Die Beschreibung dieser Sequenzen ist nicht besonders ausführlich, aber es sieht so aus, als ob die Mehrzahl der Experimente ein Set von 24 zufälligen Sequenzen benützt hätte, wobei jede Sequenz die Zufallszahlen für 20 Versuchsdurchgänge enthielt. Das heisst, dass verschiedene Personen, die anschauten, dieselbe Sequenz benutzten. Das wiederum könnte zu anderen Resultaten führen, verglichen mit Studien, bei denen jede Person eine eigene Sequenz erhält. Dies deshalb, weil Menschen möglicherweise natürliche Tendenzen haben, in einer bestimmten Art und

Weise zu reagieren, sogenannte Antworttendenzen. So könnten manche oder alle Kinder die Tendenz haben, immer zuerst mit «ja» zu antworten, oder immer alternierend zu raten, etc. Parapsychologen nennen diesen Effekt, der durch das multiple Benutzen von Zufallssequenzen zustande kommen kann, den «Stacking Effekt», was man lose mit «Aufstockungseffekt» übersetzen könnte, und geben eine Korrekturformel dafür an. Ob der Stacking Effekt wirklich ein Problem für Sheldrakes Daten ist, kann man nur feststellen, indem man die 24 Sequenzen miteinander vergleicht, um herauszufinden, ob sie ein bestimmtes Muster gemeinsam haben, etwa dass die meisten mit der Anweisung «anschauen» beginnen, oder häufig alternierende Anweisungen enthalten. Aber ich bin mir aufgrund unserer eigenen Erfahrung mit dem Stacking Effekt [1] ziemlich sicher, dass er kaum die 3% Abweichung von der Zufallserwartung bei einer so grossen Stichprobe erklären kann.

Zum gleichen Themenkreis gehört auch die Frage, ob die Zufallssequenzen, die Sheldrake benutzte, verschiedene Tests auf Zufälligkeit bestehen würden. Colwell et al. [2] erhielten Zugang zu Sheldrakes Sequenzen und fanden heraus, dass seine Zufallsreihen tatsächlich weniger Wiederholungen erhielten, als man zufällig erwarten würde. Das heisst, dass die Sequenzen zu viele Abwechslungen enthielten (z. B. «schauen», «nicht schauen», «schauen», «nicht schauen») und zu wenige Wiederholungen, vor allem zu wenige Triplets und Quadruplets («schauen», «schauen», «schauen», «schauen»). Es ist bekannt, dass Menschen schlechte Zufallsgeneratoren sind und die Häufigkeit, mit der Reihen gleicher Ereignisse hintereinander per Zufall auftreten können, konstant und in grossem Ausmass unterschätzen. Wenn nun Versuchspersonen in ihrem Rateverhalten ähnliche – vom echten Zufall abweichende – Muster aufweisen wie die Zufallssequenzen selber, was hier sehr wahrscheinlich ist, dann könnte eine überzufällige Übereinstimmung von Rateverhalten mit dem hier verwendeten «Zufall» möglich werden. Ob diese Art des Artefakts tatsächlich einen Teil der berichteten Effekte erklären hilft, kann man aufgrund der Publikation schwer entscheiden. Dies könnte allerdings leicht durch eine Reanalyse der Zufalls- und Ratesequenzen überprüft werden.

Ein anderer wichtiger Punkt, den man berücksichtigen muss, ist die Tatsache, dass alle Experimente Sheldrakes Feldexperimente waren, d. h. in Schulen mit Schülern unter 10 Jahren durchgeführt wurden. Dies mag von Vorteil sein für die ökologische Validität des Experiments, ist aber ein Nachteil für die interne Validität und die experimentelle Kontrolle. Ein Experiment mit Schülern und Lehrern könnte etwa mehr Flüchtigkeitsfehler enthalten als ein ähnliches Experiment, das von erfahrenen und ausgebildeten Experimentatoren in einem Labor durchgeführt wird. Ausserdem könnten Fehler, die in einem Labor vermieden werden oder auffallen würden, etwa das Schummeln eines Paares, in einer grossen Gruppentestsituation wie dieser, bei der viele Paare von Anschauenden und Angeschauten gleichzeitig arbeiten, leicht übersehen werden. Ob diese Tatsache Artefakte erzeugt, die den «Remote

Staring»-Effekt begünstigen oder nicht, kann man überprüfen, indem man das gleiche Experiment im Labor macht. Colwell et al. [2] haben das getan. Sie konnten den Effekt in einer Bedingung mit Rückmeldung replizieren (54,8% korrekte Treffer, n = 2160 Versuchsdurchgänge), nicht aber in einer Bedingung ohne Rückmeldung (49,9% Trefferrate, n = 720 Versuchsdurchgänge). Sie interpretieren das positive Ergebnis als Musterlernen durch die Versuchspersonen, das erleichtert wird durch eine verzerrte Zufallssequenz (z. B. «Nach zweimal Anschauen ist das nächste Mal Wegschauen»). Deshalb führten sie ein zweites Experiment durch, das unverzerrte Zufallssequenzen enthielt. Leider war die Anzahl der «Anschauen»- und «Wegschauen»-Durchgänge nicht balanciert, und die Reihen enthielten mehr «Anschauen»- (94) als «Wegschauen»- (86) Durchgänge. Weil die Versuchspersonen häufiger «Anschauen» als «Wegschauen» rieten, enthalten die Ergebnisse eine Verzerrung in der Antworttendenz, einen sogenannten response bias (50,1%; n = 2160). Obwohl also das perfekte Laborexperiment noch aussteht, scheinen die Ergebnisse von Colwell et al. [2] Sheldrake auf den ersten Blick zu widerlegen.

Aber ein entscheidender Teil der Ergebnisse blieb bislang unerwähnt: Sheldrake berichtet in allen seinen Experimenten mehr richtige Treffer unter der Anschauen-Bedingung als unter der Wegschauen-Bedingung. Manchmal beträgt die Anzahl der Treffer sogar 59%. Das bedeutet, dass die meisten Daten, die für das gesamthaft positive Resultat der Experimente verantwortlich sind, aus den Versuchsdurchgängen stammen, in denen die Versuchspersonen angeschaut werden. Sheldrake interpretiert diese Ergebnisse mit «a tendency for people to know when they are being looked at» (1999, p. 70). Jedoch könnte es sein, dass sie nicht in der Lage sind, das Fehlen eines solchen Effektes richtig zu beurteilen. Sheldrake führt hierfür die evolutionsbiologische Erklärung an, dass es für Lebewesen in einem Jäger-Beute-Verhältnis von entscheidender Bedeutung sein kann, dass ein potentielles Beutetier merkt, wann es angeschaut wird. Hingegen ist das Fehlen einer solchen Tatsache ohne biologischen Belang. Diese Erklärung könnte zutreffen. Jedoch gibt es noch eine zweite Erklärung für dieses Ergebnis: Es könnte sein, dass Versuchspersonen in beiden Bedingungen leicht besser sind als der Zufall. Ausserdem weisen sie noch eine Antworttendenz auf, etwa indem sie häufiger «Anschauen» als «Wegschauen» raten. Diese Antworttendenz ist die einzige alternative Erklärung, die mir einfällt, die die ungleiche Verteilung des Effekts zugunsten der positiven Urteile erklären könnte. Alle anderen möglichen Quellen für Artefakte sind immer die gleichen für beide Bedingungen und würden nicht einseitig die «Anschauen»-Bedingungen begünstigen. Deshalb ist die Kontrolle der Antworttendenz in Sheldrakes Daten zentral für die Interpretation der Ergebnisse. Wenn man eine solche Verzerrung der Antwort finden kann, könnte Sheldrakes Effekt erklärt werden durch verschiedene Artefakte, wie oben ausgeführt, bis ein wirklich gut kontrolliertes Laborexperiment das Gegenteil

beweist. Aber wenn es keine Antworttendenz gibt, bleibt die ungleiche Verteilung der Antwort zugunsten des Anschauens unerklärt und stützt Sheldrakes Hypothese.

Literatur

1 Schmidt S, Walach H: Wasser oder Gift? Ein Wünschenrutenexperiment. *Z Parapsychol Grenzgeb Psychol* 1997;39:76–91.

2 Colwell J, Schröder S, Sladen D: The ability to detect unseen staring: A literature review and empirical tests. *Br J Psychol* 2000;91:71–85.

Kontaktadresse: Stefan Schmidt, Institut für Umweltmedizin und Krankenhaushygiene, Universitätsklinikum Freiburg, Hugstetterstrasse 55, D-79106 Freiburg, E-mail Sschmidt@iuk3.ukl.uni-freiburg.de

Adresse für Sonderdrucke: Dr. Rupert Sheldrake, 20 Willow Road, London NW3 1TJ, UK

Kommentar – R. Nelson, Princeton

Before reading the articles published in *Rivista di Biologia/Biology Forum* 92 and 93 (1999 and 2000) by Rupert Sheldrake about his research on the sense of being stared at, I had a half-formed opinion. I had read summaries and heard people discuss the work, and this informal introduction left me with concerns that the experiments were not very well controlled, and the results might be misleading. Even after I began reading the first of the two papers, this skeptical feeling persisted for a while. But, although there are some questions to ask and minor points of criticism to make, a thorough reading of the formal documentation of this research leaves me in quite a different state. I remain cautious – perhaps it would be correct to say bemused – but there is little room for the ill-founded suspicion that some great unseen mistake, or that fraud of some sort might explain the remarkable findings in Sheldrake's series of experiments.

The most difficult issue to reconcile with long experience in the complex area of anomalies research is the very consistency of the results. Almost every experiment succeeds, whether directly in Sheldrake's control or done by teachers in the US or Canada or Germany, or by amateurs using the randomized protocols provided on Sheldrake's dedicated website. The results are almost 'too good' to be true. The percentages representing the anomalous communication that allows subject to guess correctly when a looker is staring at his or her back are so similar from one case to the next that this hardly looks like an 'extra-sensory' experiment. But, as Sheldrake says, if further studies confirm an ability to detect an unseen gaze, the implications for our understanding of perception and the nature of the human mind will be profound.

In fact, further studies already exist, and in the main they very clearly affirm the principle finding of an ability to react consciously or unconsciously to the stare of an unseen person. Sheldrake provides a thorough summary of past and current work by other people asking the same basic question. In one of the most interesting cases, from early in the 20th century, he notes that the negative conclusions of the skeptical researcher, J.E. Coover, were simply wrong. Coover's data, when properly analyzed, show a strong positive effect. (Incidentally, Coover's work in other paradigms investigating anomalous human capacities showed the same character. He interpreted striking positive results as non-significant by setting his acceptance criterion too high. Today we would say he made himself vulnerable to a Type 2 statistical error, falsely rejecting a hypothesis).

Current work by several investigators shows a similar rate of success in research that is rigorously controlled. Closed circuit television allows an unquestionable exclusion of any normal communication between looker and subject. Direct recording of physiological measures into computer files and automated statistical processing of the data completely remove human errors and biases from the picture in most recent research, including studies done by skeptics. In replicated experiments by Schlitz and Wiseman, for example, the two experimenters have achieved a powerfully instructive result: Schlitz is repeatedly successful in eliciting significant performance, while Wiseman consistently produces non-significant results. The difference appears to be attributable to, or at least correlated with, their differing beliefs; Schlitz expects the remote staring to produce an effect while Wiseman is a strong skeptic, albeit willing to do actual experiments.

The results of Sheldrake's series of staring experiments, in the context of independent studies with different methodologies but the same hypothesis, are persuasive. These are papers well worth reading in detail, for they cover virtually all reasonable concerns and skeptical questions in a clear and scholarly fashion. The experiments are simple, and the earlier ones were indeed vulnerable to artifactual contributions to an apparent effect. But Sheldrake has closed all of the obvious channels for ordinary communication in his later work and shown that the same extraordinary results are found. He does not describe a mechanism by which we can correctly guess that an unseen person is staring, but he provides thought-provoking evidence that studies of this kind are not only viable, but potentially of great value in understanding the nature of consciousness.

Kontaktadresse: Roger D. Nelson, Princeton Engineering Anomalies Research, C-131 E-Quad, Princeton University, Princeton, NJ 08544, USA, E-mail rdnelson@princeton.edu