# (20) Neuromythologie – Was passiert, wenn man statistische Voraussetzungen verletzt

Soeben ist vermutlich die größte publikatorische Bombe explodiert, die ich seit langer Zeit gesehen habe: Eine Gruppe von schwedischen Autoren haben zusammen mit einem englischen Statistiker eine riesige Simulationsstudie publiziert. Sie zeigt, dass möglicherweise bis zu 70% oder mehr der insgesamt mehr als 40.000 publizierten neurowissenschaftlichen Studien, die mit funktioneller magnetischer Resonanzspektroskopie (fMRI) gearbeitet haben, unbrauchbare Ergebnisse geliefert haben und daher eigentlich ausgemistet oder repliziert gehören. [1]

Dies scheint mir einer der größten wissenschaftlichen Kollektivskandale der letzten Zeit zu sein. Und man kann an ihm sehr viel über Statistik lernen. Aber der Reihe nach.

Bevor wir uns dieser Studie widmen: Das heißt nicht, dass die MRI-Methodik falsch ist und die sog. strukturellen Imaging-Methoden unbrauchbar sind. Es geht einzig und allein um Aussagen über räumliche Ausbreitung von Aktivität bei funktioneller magnetischer Bildgebung. Aber auch das ist schon ein gewaltiger Brocken. Folgen Sie mir.

**Was ist passiert?**

**Funktionelle magnetische Resonanzspektroskopie** oder –Imaging (fMRI) ist als Forschungsmethode sehr populär. Das Verfahren basiert darauf, dass Wasserstoffatome – die überall vorkommen – über starke äußere Magnetfelder ausgerichtet werden können. Durch das gleichzeitige Anlegen und Abtasten von elektromagnetischen hochfrequenten Wellen kann man die Atome lokalisieren. Je nachdem, welche Frequenz dabei gewählt wird, kann man auch unterschiedliche Typen von Strukturen oder Molekülen sichtbar machen. Dies kann man sich zunutze machen, um zum Beispiel den Unterschied festzustellen zwischen Blut, dessen rote Blutkörperchen mit Sauerstoff gesättigt sind und solchem, das seinen Sauerstoff abgegeben hat.

Dieses sogenannte BOLD-Signal, kurz für „blood oxygenation level dependent signal“, also ein Signal, das abhängig ist von der Sauerstoffsättigung des Blutes, kann man verwenden, um abzuleiten, wie hoch die metabolische Aktivität in einem bestimmten Areal des Körpers ist, z.B. in einem Gehirnareal. Eine Zunahme deutet auf erhöhten Sauerstoffverbrauch, erhöhte Blutzufuhr, erhöhten Metabolismus und damit erhöhte Aktivität in einem Areal des Gehirns hin. Eine Abnahme auf das Gegenteil.

Damit man nun in einer funktionellen magnetischen Resonanz Imaging (also Bildgebungs-) Studie überhaupt etwas sieht, muss man natürlich Unterschiede zwischen experimentellen und Kontrollbedingungen erzeugen. Dies geschieht in der Regel dadurch, dass Menschen in der MRI-Röhre unterschiedliche Aufgaben in bestimmter Folge erledigen müssen, sog. Blöcke. Sie müssen z.B. einen Text auf einem Bildschirm lesen, oder sollen an etwas Bestimmtes denken, oder im Geiste ein auswendig gelerntes Gedicht aufsagen, oder eben nicht und sich stattdessen entspannt hinlegen. Dies geschieht in festgelegten Sequenzen. Somit kann man die Sequenzen, in denen etwas Definiertes im Geist passiert mit denen vergleichen, in denen Ruhe herrscht.

Aus der Differenz der Signale wird dann der Unterschied in den Aktivierungsniveaus der beiden Bedingungen in bestimmten Gehirnarealen errechnet und daraus Ableitungen darüber getroffen, welche Bereiche des Gehirns für welche Funktionen zuständig sind. Zusätzlich werden solche Bedingungen häufig mit Situationen verglichen, bei denen Kontrollpersonen nur gemessen („gescannt“ sagt man im Neuro-Jargon) werden, ohne dass etwas passiert.

Sicherheitshalber sei noch hinzugesetzt: Man kann die Methode auch verwenden, um anatomische Strukturen darzustellen oder die Funktionalität von Verbindungen innerhalb des Gehirns zu erfassen. Diese beiden Einsatzbereiche sind von der hier besprochenen Studie nicht erfasst, sondern nur die Aktivierung von Gehirnarealen als Folge von Aktivitätsveränderung aufgrund experimenteller Anweisung.

**Nun müssen die Signale, die aus der Messung entstehen**, das kann man sich auch als Laie leicht vorstellen, durch eine Reihe komplexer mathematischer und statistischer Prozeduren laufen, bevor am Ende die hübschen bunten Bilder entstehen, die wir in den Publikationen und Hochglanzbroschüren bewundern.

Bei denen erklären dann Experten, das Gehirn würde „aufleuchten“, wenn ein Mensch dies oder jenes täte. Dieses „Aufleuchten“ bezieht sich auf die Falschfarbendarstellung der Zu- oder Abnahme des BOLD-Signals in bestimmten Arealen, die man als signifikanten Effekt aus dem Hintergrundrauschen statistisch isoliert hat. Es ist diese statistische Filterungsprozedur, die dann zur Farbgebung führt – die ja nichts anderes als die bildliche Umsetzung statistisch signifikanter Signalentdeckung ist – die in dieser Publikation untersucht und in der überwiegenden Mehrzahl der Fälle als unzuverlässig gefunden wurde. Warum?

**Diese statistische Filterungsprozedur ist unzuverlässig – warum?**

Die Signaldetektion in einer fMRI-Studie erfolgt im Wesentlichen in zwei Schritten. Der erste Schritt ist das Aufgreifen der Rohsignale aus dem gepulsten Einsatz der Magnetfelder und ihrer Abschaltung und deren Abtasten mit einem hochfrequenten elektromagnetischen Feld. Dieses liefert die Rohdaten über Aktivitätsveränderungen der Blutversorgung im Gehirn, also über die Sauerstoffsättigung des Blutes und der Veränderung der Verteilung des Blutes im Gehirn. Das ergibt natürlich, das sieht man sofort, Millionen von Datenpunkten, die in rascher Folge ermittelt werden und die als solche nicht roh verwendbar sind.

**Der zweite und entscheidende Schritt** ist nun die statistische Entdeckungs- und Zusammenfassungsprozedur. Dazu werden die Rohdaten mit speziellen Programmen analysiert. Die hier besprochene Studie hat die drei populärsten Programme untersucht. Um zu verstehen, wie komplex das Ganze ist, muss man sich vorstellen, dass ja die fMRI-Signale zunächst an unterschiedlichen Stellen an der Oberfläche des Kopfes aufgegriffen werden und zudem aus unterschiedlichen Tiefenbereichen des Gehirns entstammen. Es handelt sich also um dreidimensionale Datenpunkte, die analog zu den zweidimensionalen Datenpunkten eines Bildschirms, wo sie mittlerweile für alle bekannt „Pixel“ heißen, in Analogie dazu als Voxel bezeichnet werden. Voxel sind also dreidimensionale Pixel, die von einem definierten Ort herstammen und in der Intensität schwanken. Da Voxel gerade mal 1 Kubikmillimeter abdecken, wäre das Bild, das entstehen würde, extrem wirr, wenn man sie alle einzeln analysieren müsste.

Aus diesem Grunde fasst man die Voxel in der Regel zu größeren Arealen

zusammen. Dies geschieht, indem man Annahmen darüber trifft, wie die Aktivität von benachbarten Punkten miteinander zusammenhängen, wenn ein größeres funktionales Gehirnareal, sagen wir mal das Sprachzentrum bei der Generierung von mentalem Monolog, aktiviert wird. Dies geschieht über sog. Autokorrelationsfunktionen räumlicher Natur. Wir alle kennen Autokorrelationsfunktionen zeitlicher Natur: Wenn das Wetter heute sehr schön ist, ist die Wahrscheinlichkeit, dass es morgen auch sehr schön ist höher, als wenn es schon 2 Wochen lang schön war. Denn dann ist die Wahrscheinlichkeit, dass es morgen schlechter wird allmählich höher, und umgekehrt.

Analog zu einer solchen zeitlichen Autokorrelation kann man sich auch eine räumliche vorstellen: Je nachdem wie hoch die Aktivität an einem Punkt des Voxel-Universums ist, wird die Wahrscheinlichkeit, dass ein benachbartes Voxel zu einer funktionellen Einheit gehört höher oder geringer sein. Zu Anfangszeiten der Programm-Entwicklung zur Analyse solcher Daten lagen dazu noch relativ wenige Informationen vor. Also hat man eine vernünftige, aber wie sich nun herausstellt falsche, Annahme getroffen: dass sich nämlich die räumliche Autokorrelationsfunktion als eine sich räumlich ausbreitende Gausskurve oder Normalverteilung verhält.

**Kontrolle der Kontrolldaten**

Nun liegen mittlerweile Tausende von Datensätzen von Menschen vor, die sozusagen zu Kontrollzwecken, ohne irgendwelche Aufgaben, mit MRI-Scannern gemessen wurden, und dank der Möglichkeit offener Plattformen werden diese Daten Wissenschaftlern offen zur Verfügung gestellt. Jeder kann sie herunterladen und damit Analysen anstellen. Diese Möglichkeit haben die Wissenschaftler genutzt und Daten von knapp 500 gesunden Menschen aus unterschiedlichen Regionen der Welt, die ohne irgend eine Aufgabe in einem Scanner vermessen wurden, mit simulierten Analysemethoden nachgerechnet, indem sie die drei gängigsten Analysesoftwarepakete darauf anwandten.

Insgesamt haben sie 192 Kombinationen von möglichen Einstellungen in mehr als 3 Millionen Simulationsrechnungen überprüft. Etwas vereinfacht gesagt haben die Wissenschaftler also so getan, als wären die Daten dieser 500 Leute aus echten fMRI-Experimenten mit Ein- und Ausschalt-Blocks mit bestimmten Aufgaben oder Fragestellungen entstanden. Es steht aber fest, dass dies nicht der Fall war, weil es sich um Kontrolldaten handelte.

**Man würde bei einer solchen Prozedur erwarten**, dass sich immer eine bestimmte Anzahl falsch positiver Ergebnisse findet, also Ergebnisse, wo die Statistik sagt: „Hurrah, wir haben einen signifikanten Effekt gefunden“, wo aber in Tat und Wahrheit kein Effekt vorliegt. Dieser sogenannte Fehler erster Art oder alpha-Fehler wird durch das nominelle Signifikanz-Niveau kontrolliert, das man per Konvention festlegen kann und das häufig bei 5% liegt (p = 0.05), im Fall der fMRI-Studien aber häufig schon von vorneherein niedriger, nämlich auf 1% (p = 0.01) oder 0.1% (p = 0.001) gesetzt wird. Denn dieser Alpha-Fehler gibt an, wie häufig wir einen Fehler machen, wenn wir einen Effekt behaupten, obwohl keiner da ist. Bei 5%-igem Alpha-Fehler Niveau machen wir einen solchen Fehler in 5 von 100 Fällen. Bei einem 1%-Niveau des Alpha Fehlers in einem von 100 Fällen. Und bei einem 1 Promille Niveau in einem von 1.000 Fällen.

Wenn wir nun viele statistische Tests parallel auf den gleichen Datensatz anwenden, dann multipliziert sich dieser Fehler natürlich, weil wir bei jedem Test wieder die gleiche Wahrscheinlichkeit erhalten, einen Fehler zu machen, wenn wir eine Tatsachenbehauptung aufstellen, die in Wirklichkeit nicht zutrifft. Aus der nominellen Fehler-Wahrscheinlichkeit von p = 0.05, also 5%, wird dann bei zwei gleichzeitigen Tests die Fehlerwahrscheinlichkeit von p = 0.1 oder 10% Wir machen also doppelt so viele Fehler. Um die nominelle Wahrscheinlichkeit von 5% einhalten zu können, müssen also bei zwei gleichzeitigen Tests am gleichen Datensatz die individuellen Wahrscheinlichkeiten auf p = 0.025 gesetzt werden, damit die gemeinsame Fehlerwahrscheinlichkeit p = 0.05 bestehen bleibt. Dies nennt man „Korrektur für multiples Testen“.

Weil bei den fMRI-Auswertungspaketen gleich sehr viele Tests gemacht werden, setzt man dort die Entdeckungsschwelle für das, was man als signifikantes Signal bereit ist zu akzeptieren gleich von vorneherein auf p = 0.01 (also eine Fehlerwahrscheinlichkeit, die für 5 gleichzeitige Tests bereinigt ist) oder gar auf p = 0.001. Dies ist eine Fehlerwahrscheinlichkeit, die für 50 gleichzeitige Tests bereinigt und damit das nominelle Fehlerniveau von 5% bei 50 Tests einhält. Diese Korrektur ist bei den untersuchten Softwarepaketen bereits eingebaut; das gefundene Problem hängt also nicht damit zusammen.

Alle diese Parametereinstellungen wurden bei der hier durchgeführten Studie verwendet. Gleichzeitig wurden Szenarien durchgespielt, die in der Realität der fMRI-Forschung gängige Praxis sind, also dass man z.B. 8 mm große Cluster nimmt und die benachbarten Voxel mit einer Entdeckungsschwelle von p = 0.001

zusammenfasst, was als ganz vernünftig erscheint.

Dann wurden in komplexen Simulationsrechnungen alle möglichen vermeintlichen experimentellen Vergleiche über diese Kontrolldaten gelegt und dokumentiert, wie häufig die verschiedenen Softwarepakete signifikante „Entdeckungen“ machen, obwohl bekannterweise keinerlei Signale in den Daten versteckt sind.

Wenn Cluster gebildet werden, also Voxel zusammengefasst werden zu größeren Arealen, dann finden sich falsch positive Werte, also Signale, wo es keine gibt, in bis zu 50% der Analysen. Oder anders gesagt: manche Softwarepakete entdecken Signale mit einer 50%igen Fehlerwahrscheinlichkeit, wo gar keine Signale sind. Nochmals anders gesagt: in 50 von 100 Studien sagt die Analyse: „Hier ist ein signifikanter Effekt vorhanden“ wo gar kein Effekt vorliegt.

Wenn die Clustergröße kleiner ist und die Zusammenfassungsschwelle von Voxeln zu Clustern höher ist, dann nähert sich die Fehlerwahrscheinlichkeit der nominellen Signifikanzgrenze von 5%. Für die voxel-basierte Analyse, also wenn man keine Annahmen über den Zusammenhang von Voxelaktivitäten trifft und dafür in Kauf nimmt, dass man ein wirres Bild von vielen Voxeln interpretieren muss, bleibt die Analyse bei fast allen Software-Paketen nahe der Fehlerwahrscheinlichkeit von 5%.

Und für die sog. nichtparametrische Methode, also eine statistische Auswertung die auf einer Simulationsrechnung basiert, bei der die Wahrscheinlichkeit nicht von einer zugrundegelegten und vermuteten Verteilung, sondern von einer aktuellen Simulationsrechnung aufgrund der vorliegenden Daten abgeleitet wird, bleiben die nominellen Signifikanzwerte immer erhalten.

**Das Problem ist allerdings: Die Software-Pakete werden ja eingesetzt, weil** man eine mühsame Interpretation einer voxel-basierten Auswertung nicht selber machen, sondern an den Computer delegieren will, und weil man eben nicht wochenlange Simulationsrechnungen zur Bestimmung der wahren Wahrscheinlichkeit vornehmen will. Außerdem würden bei voxelbasierter Auswertung Signalrauschem oder Artefakte, wie sie z.B. von Bewegungen herrühren, zu stark ins Gewicht fallen. Also versucht man vermeintlich robustere Größen zu finden, eben jene Cluster, die man dann testet.

Für ein sehr häufig angewendetes Szenario, die oben beschriebenen 8 mm

großen Cluster mit einer anscheinend konservativen Entdeckungsschwelle von p = 0.001 von Voxel zu Voxel, bevor man geneigt ist ein Cluster als „signifikant aktiviert“ oder „signifikant inaktiviert“ anzusehen, sehen die Werte düster aus: die Fehlerhäufigkeit steigt je nach Programm auf bis zu 90% und **eine 70%ige Fehlerwahrscheinlichkeit quer durch die Literatur ist eine robuste Schätzung**.

Nur eine nichtparametrische Simulationsstatistik würde auch hier keine überzogenen Fehler machen. Allerdings kommt diese so gut wie gar nicht vor. Übrigens wurde das gleiche Problem auch für aktive Daten aus wirklichen Studien gefunden. Auch hier ist eine sog. Inflation des alpha-Fehlers oder eine viel zu häufige Entdeckung von Effekten wo gar keine vorliegen nachgewiesen.

**Woher kommt das Problem?**

Man kann an diesem Beispiel die Bedeutung von Voraussetzungen für die Gültigkeit von Statistik studieren. Zum einen machen die Softwarepakete und die Anwender Annahmen über den Zusammenhang der Voxel über räumliche Autokorrelationsfunktionen, wie ich oben beschrieben habe. Die Anwender wählen außerdem die Größe der zu untersuchenden Areale und die dabei eingesetzten Glättungen. Diese ursprünglichen Annahmen waren zunächst vernünftig, wurden aber aufgestellt zu einer Zeit, als man noch relativ wenige Daten hatte. Keiner hat sie überprüft. Bis jetzt. Und siehe da, genau diese zentrale Voraussetzung, die den mathematischen Zusammenhang benachbarter Voxel beschreibt, war falsch. Also: zurück zu den Büchern; Softwareprogramme modifizieren, neue, der empirischen Wirklichkeit näher liegende Autokorrelationsfunktionen implementieren. Und neu rechnen.

Andere Voraussetzungen haben damit zu tun, dass man statistische Verteilungen für die Daten annimmt. Das ist etwas, was man häufig macht. Man nennt die damit verbundenen Schlußfolgerungsverfahren daher „parametrische Statistik“, weil man eine bekannte Verteilung für die Daten annimmt. Die bekannte Verteilung kann man normieren. Man interpretiert dann die Fläche unter der Kurve als „1“. Wenn man dann einen Wert irgendwo auf der Achse abträgt und die dahinter liegende Fläche berechnet, kann man diesen Flächenanteil von 1 als Wahrscheinlichkeit interpretieren.

So liegen etwa hinter dem Achsenwert „2“ (oder „-2“) der

Standardnormalverteilung mehr als 95% oder weniger als 5% der Fläche. Weil die Fläche auf „1“ normiert ist, kann man dies dann als Wahrscheinlichkeit interpretieren. Also kann man aus einer bekannten Verteilung Fehlerwahrscheinlichkeiten berechnen. Eine häufig gemachte Verteilungsannahme ist die auf Normalverteilung, aber es gibt auch eine Fülle von anderen statistischen Verteilungskurven, bei denen man dann auf die gleiche Weise den Flächenanteil einer genormten Kurve ausrechnen und damit die Wahrscheinlichkeit bestimmten kann.

**Nur: wir wissen selten, ob diese Annahmen auch stimmen**. Daher, das zeigt diese Analyse, ist eigentlich ein nicht-parametrisches Verfahren, also eines, das keine Verteilungsannahme über die Daten macht, klüger. Die Diskussion darüber ist schon sehr alt und bekannt, ebenso die Verfahren [2]. Wir haben sie verschiedentlich, vor allem in kritischen Situationen eingesetzt [3,4]. Wenn man solche Simulations- oder nonparametrische Statistik richtig einsetzt, dann muss man eigentlich die empirisch gefundenen Daten nehmen. Man lässt den Computer, neue Datensätze generieren, sagen wir 10.000, die ähnliche Charakteristika haben, z.B. gleich viele Punkte mit bestimmten Merkmalen, und lässt dann auszählen, wie häufig die in den empirisch gefundenen Daten auftretenden Merkmale auch in den simulierten Daten vorkommen. Teilt man dann die Anzahl der empirisch aufgetretenen Merkmale durch die Anzahl der durch Zufall gefundenen Merkmale, dann hat man die wahre Wahrscheinlichkeit dafür, dass der empirische Befund durch Zufall hätte zustande kommen können.

Klarerweise sind solche Simulationsrechnungen, oft auch Monte-Carlo Analysen genannt (weil in Monte-Carlo das große Spielcasino steht) – oder eben nicht-parametrischen Analysen – sehr aufwendig. Selbst moderne, schnelle Rechner brauchen bei komplexen Analysen oft Wochen, um sie durchzuführen.

Man sieht jedenfalls an diesem Beispiel, was passiert, wenn man statistische Annahmen verletzt: Man kann die auf Annahmen basierenden Wahrscheinlichkeitswerte nicht mehr interpretieren und die Ergebnisse an die Hasen verfüttern. In diesem konkreten Fall ist eine riesige Literatur von Neuromythologie entstanden. Mehr als die Hälfte, vielleicht sogar bis zu 70% der etwa 40.000 Studien zu fMRI Methodik, müssten eigentlich wiederholt oder zumindest neu ausgewertet werden, beklagen die Autoren. Man müsste an sich dazu die Daten öffentlich zugänglich haben, dann wäre das machbar. Sind sie dummerweise in den meisten Fällen nicht. Hier trifft sich die Klage der

neurowissenschaftlichen Gemeinde mit der soeben von Psychologen erhobenen Aufforderung alles, aber auch wirklich alles, öffentlich zugänglich zu machen, Protokolle, Ergebnisse, Daten [5]. Die Autoren rufen ein Moratorium aus: erst an die Hausaufgaben, erst Altlasten aufarbeiten, dann erst wieder neue Studien machen. Das wird auch nicht überall gehen. Denn in vielen Fällen wurden Studiendaten aufgrund geltender Gesetze nach 5 Jahren gelöscht.

**Sagenhaft**

Das ist jetzt schön dumm, finde ich. Man muss überlegen: Die meisten größeren klinischen Einheiten in Krankenhäusern und die meisten größeren Universitäten in Deutschland und der Welt unterhalten MRI-Scanner; die englische Wikipedia schätzt 25.000 Scanner seien weltweit im Einsatz. Das Problem mit diesen Geräten ist, dass sie, wenn sie einmal in Betrieb gegangen sind, immer am Stromnetz hängen und damit hohe Betriebskosten erzeugen. Man kann sie auch nicht einfach abschalten wie einen Computer, denn das könnte das Gerät schädigen, bzw. das Abschalten und Hochfahren ist selbst ein sehr komplexer und aufwändiger Prozess. Daher müssen diese Geräte im Dauereinsatz gehalten werden, damit sich ihre Anschaffung von inzwischen mehreren Millionen Euro lohnt. Daher werden auch so viele Studien damit gemacht. Denn wer Studien macht, bezahlt Scannerzeit. Kaum hat einer eine einigermaßen klug scheinende Idee – „lasst uns doch mal schauen, welche Gehirnareale aktiv sind, wenn man Leuten Musik vorspielt oder Bilder zeigt, die sie nicht mögen“ – findet er auch beim heutigen Klima das Geld, um eine solche Studie finanziert zu bekommen.

> *Dass die Hirnforschung noch eine Reihe anderer Probleme hat, ist schon anderen aufgefallen, wie Hirnforscher Hasler in einem leicht verständlichen Artikel ausführt.*

Und so kommt es, dass wir einen riesigen Bestand, mittlerweile müssen wir sagen, von Märchenbüchern darüber haben, was im Gehirn so alles passieren kann, wenn Tante Emma strickt und klein Mäxchen Kinderreime auswendig lernt. Wunderschöne Bilder, hübsche Narrative, die uns alle suggerieren, das wichtigste in der Welt der Wissenschaft gegenwärtig sei das Wissen darüber, was das Gehirn treibt. Nur, dass alle diese Geschichten in der Mehrheit der Fälle kaum mehr Wert haben als die Sagen des klassischen Altertums. Die Sagen des klassischen Altertums enthalten manchmal einen Kern Wahrheit und sind mindestens

spannend. Ob der Wahrheitskern der publizierten fMRI-Studien größer ist als der der Sagen? In der Tat: Die bunten Bilder der fMRI-Studien sind die Barock-Kirchen der Postmoderne: schöne, bildhafte Narrative einer fragwürdigen Theologie.

Quellen und Literatur:

[1] Eklund, A., Nichols, T. e., & Knutsson, H. (2016). Cluster failure: Why fMRI inferences for spatial extent have inflated false-positive rates. Proceedings of the National Academy of Science, early edition. Doi: 10.1073/pnas.1602413113.
[2] Edgington, E. S. (1995, orig. 1987). Randomization Tests. 3rd Edition. New York: Dekker.
[3] Wackermann, J., Seiter, C., Keibel, H., & Walach, H. (2003). Correlations between brain electrical activities of two spatially separated human subjects. Neuroscience Letters, 336, 60-64.
[4] Schulte, D., & Walach, H. (2006). F.M. Alexander technique in the treatment of stuttering – A randomized single-case intervention study with ambulatory monitoring. Psychotherapy and Psychosomatics, 75, 190-191.
[5] Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. Science, 349(6251), aac4716.

Zurück zu Kapitel 19 →
Weiter zu Kapitel 21 →