

Wissen

Sind Tausende Hirnscan-Studien falsch?

Schwedische Forscher haben eines der wichtigsten bildgebenden Verfahren überprüft und entdeckt, dass die Auswertungen der Messungen in vielen Fällen falsch sind. Die Experten streiten sich über die Folgen für die Hirnforschung.

Matthias Meili

Mönche waren schon drin, auch Verliebte und natürlich zahlreiche Patienten: Die modernen Scanner, die ein Aktivitätsmuster des Gehirns aufzeichnen, insbesondere die Magnetresonanztomografie (MRI), sind aus der Neurologie und der Hirnforschung nicht mehr wegzudenken. Bei Mönchen untersuchten die Forscher den Einfluss der Meditation auf die Schmerzregulierung. Und vor einem Jahr sorgte eine Studie für Aufsehen, wonach im Gehirn von Frischverliebten selbst im Ruhezustand die Zentren für Belohnung, Motivation und - nicht verwunderlich - auch die Regionen, die für die Regulierung der Gefühle zuständig sind, besonders aktiv sind.

Doch jetzt kommen schwedische Forscher zum Schluss, dass die Software hinter den bunten Bildern fehlerhaft ist. «Unsere Ergebnisse stellen 40 000 Studien infrage», teilte Anders Eklund von der Universität Linköping mit, als er die Studie Ende Juni im Fachjournal «PNAS» publizierte. Die Arbeit sorgte für gehörigen Wirbel unter den Hirnforschern. Inzwischen haben die Autoren die Zahl 40 000 freilich zurückgenommen und sprechen noch von 3500 möglicherweise falschen Studien.

25 Jahre Forschung für nichts?

«Der Aufruhr ist gross», bestätigt Philipp Stämpfli. Er leitet das MRI-Zentrum an der Psychiatrischen Universitätsklinik Zürich und hat tagtäglich mit Messungen von Hirnbildern zu tun. Auch Christoph Boesch, MRI-Forscher an der Universität Bern und Forschungsrat des Schweizerischen Nationalfonds, hält Eklunds Arbeit für wichtig. Zugleich warnen beide vor einer Überinterpretation der Resultate.

Die funktionelle Magnetresonanztomografie (fMRI) hat vor rund 25 Jahren in der Hirnforschung Einzug gehalten. Seither wurde das bildgebende Verfahren stetig weiterentwickelt - vom einfachen Bildschnitt bis zum zeitnahen Verfolgen des Gehirns bei der Arbeit. Auch der bekannte Zürcher Neuroökonom Ernst Fehr benutzt die Methode, um herauszufinden, was genau den Menschen antreibt. «Diese nicht invasiven bildgebenden Verfahren sind in der Forschung mit Menschen sehr wichtig», sagt Fehr.

Bei der Untersuchung werden die Probanden in eine Röhre geschoben, in der ein starkes Magnetfeld angelegt ist. Mittels regelmässiger Pulse wird dann die Veränderung des Blutflusses gemessen. In einem typischen Versuch, der 5 bis 15 Minuten dauert, wird zunächst der Ruhezustand aufgenommen, in dem der Proband entspannt daliegen soll. Dann folgt ein Block mit einer Aufgabe, einer Frage oder Bildern, die der Versuchsperson gezeigt werden. Effektiv gemessen werden die Sauerstoffsättigung und deren Veränderung. Je höher die Sättigung, umso aktiver ist das Gehirn in diesem Bereich, so die Annahme. Die Signale werden für jeden Bereich des Gehirns zwischen dem Ruhezustand und dem Experimentalzustand verglichen und mit komplizierten statistischen Formeln in einen Aktivitätswert umgerechnet. Daraus entstehen Aktivitätskarten, die auf anatomische Karten übertragen und koloriert werden.

Doch im Prinzip gaukeln die bunten Hirnbilder etwas vor, was in der Realität der Hirnwindungen nicht so passiert. Wenn eine Hirnregion beim Betrachten der Angebeteten rot gekennzeichnet dargestellt wird, bedeutet dies nicht, dass diese Region auch tatsächlich rot leuchtet. Auch könnten viele Areale, die nicht aufleuchten, trotzdem aktiv sein. Denn Unterschiede zwischen den Aktivitätslevels von Ruhe- und Experimentalzustand sind oft nur minimal. Erst die statistische Auswertung filtert aus dem Rauschen der Signale die Aktivitätsspitzen heraus - das Resultat ist von den Annahmen und Werten abhängig, die in der Software eingebaut sind.

Ein toter Lachs, der nachdenkt

Die Forscher um Anders Eklund haben nun diese statistischen Methoden hinter den Hirnbildern überprüft. Dabei liessen sie die Rohdaten von bereits durchgeführten Experimenten aus internati-



Wissenschaftler bereiten den buddhistischen Mönch Matthieu Ricard im Juni 2008 auf ein Experiment im Magnetresonanztomografen vor. Foto: Jeff Miller

onal bekannten Forschungsprojekten noch einmal durch acht gängige Software-Pakete laufen. Alle Daten stammten aus Sessionen, bei denen nur der Ruhezustand gemessen wurde. Weil die Probanden ihre Gedanken dabei freifliessen lassen, sollte keine spezielle Aktivität erkennbar sein.

Doch dem war nicht so. Drei der acht Software-Pakete produzierten bis zu 70 Prozent falsch-positive Signale. Standardwert bei solchen wissenschaftlichen Auswertungen ist eine falsch-positive Rate von 5 Prozent. Anders gesagt: In mehr als der Hälfte der Fälle könnten die Methoden Aktivitäten in einer Hirnregion zeigen, wo gar keine ist. Betroffen waren ausgerechnet die drei Clusterbasierten Software-Produkte, die bei Forschern sehr beliebt sind, weil sie eher zu statistisch signifikanten Resultaten führen können.

Dass die statistische Auswertung ein wunder Punkt des fMRI-Verfahrens ist, weiss man schon lange. 2009 legte der amerikanische Postdoktorand Craig

Bennett einen toten Lachs in den Scanner und präsentierte ihm getreu dem üblichen Vorgehen ein paar Bilder von traurigen und fröhlichen Menschen. In der Auswertung fand er prompt in gewissen Hirnregionen des Lachses Aktivitäten. Bennett überschrieb seine Arbeit mit dem bewusst ironischen Titel «Toter Lachs kann emotionale Zustände von Menschen erkennen». Dann begann sich Anders Eklund mit der Frage zu beschäftigen. Bereits 2012 hatte der schwedische Forscher fMRI-Daten auf ihre Gültigkeit untersucht und ähnliche Resultate zeigen können, damals allerdings noch mit simulierten Daten.

Hoher Publikationsdruck

Dank immer mehr Computerpower ist es nun möglich geworden, den Test mit den Daten aus realen Versuchen durchzurechnen. «Man muss bedenken, dass die Software mit Wahrscheinlichkeiten rechnet», sagt der Schmerzforscher Michael Meier, der fMRI-Untersuchungen für die chiropraktische Medizin der

Universitätsklinik Balgrist durchführt. Dabei seien die Grenzwerte bei der Auswertungssoftware entscheidend. «Diese werden oft grosszügig gesetzt, weil man dann eher Resultate sieht.» Diese seien dennoch mit statistischen Argumentationen vertretbar, so Meier. Deshalb sei es wichtig, dass sich der Forscher mit der statistischen Modellierung genauso auskenne, um die Resultate nicht überzuinterpretieren. Manchmal spielt aber auch der hohe Publikationsdruck auf die Forscher eine Rolle, denn ohne schnelle und zahlreiche Publikationen bleibt einem eine wissenschaftliche Karriere oft verbaut.

Dass Eklund mit seiner Arbeit in ein Wespennest gestochen hat, zeigt ein Blick in die Foren der Hirnforscher. Inzwischen werfen ihm verschiedene Neuroforscher vor, seine Aussagen seien übertrieben. Zudem bedeute der Fakt, dass so viele Studien falsch sein könnten, noch nicht, dass sie auch falsch seien. Tatsächlich stützen sich viele Studien nicht alleine auf fMRI-Bilder. Der ameri-

kanische Neuroblogger Mark Reimers betont, dass der Wert von 70 Prozent falsch-positiven Ergebnissen nur ein Spitzenwert ist, der bei weitem nicht auf alle bisherigen Studien zutrafte.

Sicher sein könnte man nur, wenn man von allen Studien die Rohdaten noch einmal über eine korrekte Software laufen liesse oder die Experimente wiederholen würde. Dies ist jedoch illusorisch, eine Stunde fMRI-Versuch kostet laut Michael Meier in der Schweiz mindestens 200 bis 300 Franken. Die Datenmengen, die bei typischen Versuchen anfallen, sind zudem exorbitant gross. «Auf einem zentralen Server würden schnell einmal mehrere Petabytes zusammenkommen», sagt der Berner MRI-Forscher Christoph Boesch.

Neue Wissenschaftskultur

Gemäss den Experten wird Eklunds Arbeit in der Szene einiges bewegen. «Zuerst müssen die Einstellungen in der Software korrigiert werden», sagt Boesch. Bei einer der getesteten Softwares wurde dies laut dem Hersteller bereits gemacht. Zudem wäre es sinnvoll, wenn solche Experimente von Experten begleitet würden, die sich mit den mathematischen Methoden auskennen. «So könnten Artefakte viel rascher erkannt werden», sagt Boesch, der für den Nationalfonds auch Gesuche in diesem Bereich beurteilt.

Eine andere Möglichkeit wäre die strikte Offenlegung aller Rohdaten und verwendeten Codes. Dadurch könnten die Reviewer oder andere interessierte Forscher die Resultate besser überprüfen. Einige wichtige Journals fordern dies bereits als Bedingung für eine Publikation. Laut Boesch sind entsprechende Anstrengungen auch bei der Forschungsförderung im Gange. Eine völlige Offenlegung sei jedoch gerade im medizinischen Bereich nicht einfach umsetzbar, weil die Daten unter anderem anonymisiert werden müssten, um den gesetzlichen Vorgaben des Datenschutzes zu genügen. Der Magnetresonanztomografie-Experte stellt noch eine weitergehende Forderung: «Publizieren von negativen Resultaten, reproduzieren von gemachten Experimenten, selbst nur mit den Rohdaten - dazu brauchte es eine neue Wissenschaftskultur, die solche Publikationen auch belohnt.»

Folgen für die Hirnforschung

«Diese Verfahren sind und bleiben extrem wichtig»

Mit Ernst Fehr sprach Matthias Meili

Schwedische Forscher haben gezeigt, dass 3500 Hirnstudien falsch sein könnten. Ein Skandal?

Die Ergebnisse wurden sehr reisserisch dargestellt. Die bildgebenden Verfahren sind noch ein junges Gebiet - da ist es normal, dass manchmal noch Fehler passieren. Für mich ist diese Arbeit aber ein wichtiger Teil im Selbstreinigungsprozess der Wissenschaft.

Stellt die Tatsache, dass in vielen Studien Hirnaktivitäten gesehen werden, wo gar keine sind, nicht den Nutzen der Methode für die Hirnforschung infrage?

Im Gegenteil. Es ändert nichts daran, dass die funktionelle Magnetresonanztomografie (fMRI) in der Hirnforschung extrem wichtig ist und bleiben wird. Wenn wir das menschliche Gehirn studieren wollen, müssen wir solche nicht invasiven Verfahren anwenden.

Wie wichtig ist diese Methode in Ihrer Forschung?

Sie spielt eine grosse Rolle. Aber ich habe immer auch andere Methoden verwendet. Denn die fMRI-Methode erlaubt nur die Identifikation von Korrelationen zwischen Verhalten und Hirnaktivität und kann keine kausalen Beziehungen ermitteln. Deshalb habe ich früh auch auf die nicht invasive Hirnstimulation gesetzt, bei der wir neuronale Netzwerke gezielt beeinflussen können. Hinzu kommen viele Labor- und Verhaltensstudien. Diese Methoden kombinieren wir. So sind wir nicht allein von den fMRI-Versuchen abhängig.

Hat das aktuelle Paper konkrete Folgen in Ihrer Arbeit?

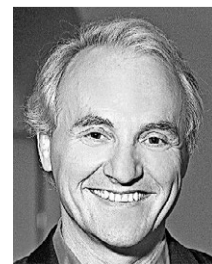
Die beanstandete Software haben wir eher selten verwendet. Eine Folge wird sein, dass die Experten, die Forschungsarbeiten begutachten, die Methode jetzt kritischer anschauen. Das beste Rezept gegen Fehler ist aber die unabhängige Replikation von Ergebnissen.

Wird das gemacht?

Ja sicher. Ein gutes Beispiel ist die berühmte Studie von Tanja Singer, in welcher sie 2004 Empathienetzwerke im Hirn entdeckte. Dieses Resultat wurde seither immer wieder bestätigt.

Spielt nicht der Publikationsdruck in der Forschung eine Rolle?

Das hat möglicherweise einen Einfluss, aber dieses Problem gibt es in allen Wissenschaftsdisziplinen. Die Anreizsysteme müssen deshalb so optimiert werden, dass bei Publikationen mehr auf Qualität statt Quantität geschaut wird.



Ernst Fehr
Der Neuroökonom von der Universität Zürich gilt als aussichtsreicher Nobelpreiskandidat. In seiner Forschung arbeitet Fehr auch intensiv mit der fMRI-Methode.