

# Gebrochene Versprechen

Ein Lehrstück darüber, wie leicht Hirnforscher bei der Interpretation ihrer Ergebnisse über das Ziel hinausschießen

VON STEPHAN SCHLEIM

In einer Studie, die Ende 2009 im Fachjournal »Neuron« erschien, untersuchten Neuroökonominnen die Hirnaktivität unehrlicher Personen. Die Ergebnisse weisen laut der Autoren Thomas Baumgartner und Ernst Fehr von der Universität Zürich sowie Urs Fischbacher von der Universität Konstanz darauf hin, dass Menschen, die mit großer Wahrscheinlichkeit ein Versprechen brechen, einen emotionalen Konflikt erleben. Die Forscher zogen daraus weit reichende Schlüsse: Ein einzigartiges, »betrügerisches« Muster von neuronaler Aktivität erlaube es nicht nur, ehrliche und unehrliche Menschen voneinander zu unterscheiden. Man könne allein durch den Blick ins Gehirn sogar vorhersagen, ob sich eine Person zukünftig fair verhalten werde oder nicht.

Bei dem Experiment, das sich die Studienautoren ausgedacht hatten, ging es um ein Vertrauensspiel. Ein Proband sollte, während er im Hirnscanner lag, die Wahrscheinlichkeit angeben, mit der er einem anonymen Mitspieler Geld zurückzahlen würde, das er von diesem erhielt (»immer«, »meistens«, »manchmal« oder »nie«). Daraufhin konnte der andere bestimmen, ob er dem Ersten 50 Schweizer

Rappen anvertrauen wollte. Wenn ja, verfünffachte der Versuchsleiter die Summe auf 2,50 Franken. Nun entschied wiederum der erste Proband, ob er diesen Betrag mit dem vertrauensseligen Mitspieler teilen und die Hälfte der Summe an ihn zurückgeben – oder alles selbst einstreichen wollte.

Diese ganze Prozedur wiederholte jeder Teilnehmer 24-mal, jeweils mit einem anderen Spielpartner. Bei der Hälfte der Durchgänge musste zuvor kein Versprechen abgegeben werden – sie dienten als Kontrollbedingung.

## Ein Muster für Unehrllichkeit?

Anhand ihres Verhaltens ließen sich die Versuchspersonen anschließend in zwei Gruppen einteilen: die der ehrlichen und die der unehrlichen. Allen war gemeinsam, dass sie sich gegenüber ihrem Mitspieler als sehr zuverlässig beschrieben und angaben, das Geld immer oder doch meistens teilen zu wollen. Doch während die ehrlichen Probanden in rund 90 Prozent ihre Zusage wahr machten, hielten die anderen nur in rund jedem dritten Fall Wort. Auch in den Runden ohne Versprechen agierten die Teilnehmer äh-

lich: Die ehrlichen teilten das Geld meistens, die Unehrllichen dagegen nur selten.

Die Neuroökonominnen registrierten währenddessen per funktioneller Magnetresonanztomografie (fMRT) die Hirnaktivität der Probanden – und zwar in drei verschiedenen Phasen: beim Versprechen, beim Warten auf die Entscheidung des Mitspielers und bei der eigenen Entscheidung, ob das Geld geteilt wird. Während der ersten Phase fanden die Forscher bei den unehrlichen Menschen stärkere Aktivität im vorderen Zingulum sowie in der so genannten Insula im Frontalhirn. Das vordere Zingulum wurde in früheren Studien bereits mit emotionalen Konflikten und deren Kontrolle in Zusammenhang gebracht, die Insula mit unangenehmem Gefühlserleben. Daher vermuten die Forscher, die unehrlichen Probanden hegten schon zu diesem Zeitpunkt die Absicht, ihr Versprechen später zu brechen. Das Muster der Hirnaktivierungen sei somit womöglich Ausdruck eines schlechten Gewissens.

Zu dieser Interpretation passen die parallel gemessenen Reaktionszeiten aber nicht so richtig. Weder bei der Abgabe des Versprechens noch bei den Entscheidun-

gen, das Geld zu teilen, gab es Unterschiede. Hätten die unehrlichen Versuchspersonen einen Konflikt verspürt und diese Gefühle kontrollieren müssen, so hätten ihre Entscheidungen eigentlich mehr Zeit in Anspruch nehmen müssen. Das zeigt etwa die einschlägigen Experimente des Psychologen Joshua Greene von der Harvard University in Boston (damals noch an der Princeton University) zum moralischen Urteilen: Das Überwinden eines mit Aktivierung des Zingulums einhergehenden Konflikts dauert im Mittel zwei

diese Art von Umkehrschlüssen. Ihm zufolge verfähre ein Teil der Neurowissenschaften offenbar nach dem Prinzip, die Hirnaktivität einiger Versuchspersonen zu messen und anschließend »eine Geschichte darum herumzustricken«.

Das gleiche Problem begegnet uns in den weiteren Spielphasen des Experiments von Baumgartner und Kollegen. Ihre Versuchsteilnehmer mussten sechs Sekunden warten, während der Mitspieler sich dafür oder dagegen entschied, ihnen Geld anzuvertrauen. Die Neuroöko-

lieren die Forscher. Da sich bei den ehrlichen Probanden umgekehrt keine erhöhten Aktivierungen fanden, folgern sie, Versprechen zu halten entspreche der Norm. Dass sich die Abweichung von der Norm wiederum nicht an den Reaktionszeiten nachvollziehen lässt, ist unplausibel.

Schon Daniel Langleben von der University of Pennsylvania in Philadelphia, ein Pionier der neurowissenschaftlichen Lügenforschung, erklärte auf Grund eines Experiments von 2002 Wahrheit zum

---

## *Da die meisten Hirnregionen an zahlreichen Prozessen beteiligt sind, gibt es keine einfache **Eins-zu-eins-Entsprechung** zwischen Gehirn und Geist*

---

Sekunden. Berücksichtigt man ferner, dass die Probanden schon für das bloße Drücken der Knöpfe etwa 2,3 Sekunden benötigten (wie sich aus den Spielrunden ohne Versprechen ergab), so erscheint die »Konflikttheorie« noch unplausibler. Sollte der Entschluss zum Lügen in ganzen 0,8 Sekunden gefallen sein, ohne dass sich die Reaktionszeit im Vergleich zu ehrlichen Menschen verlängerte? Das widerspräche zentralen Annahmen der experimentellen Entscheidungsforschung.

Thomas Baumgartner und seine Kollegen machten aus der Not eine Tugend und folgerten, die Kernspintomografie offenbare etwas, was Verhaltensmaße allein nicht verriet. Damit handeln sich die Forscher aber ein größeres Problem ein, als ihnen womöglich bewusst ist. Denn sie schließen aus nichts anderem als der neuronalen Aktivität auf Konflikt, Kontrolle und schlechtes Gewissen. Da die meisten Hirnregionen aber an zahlreichen Prozessen beteiligt sind, gibt es keine einfache Eins-zu-eins-Entsprechung zwischen Gehirn und Geist.

Russell Poldrack, kognitiver Neurowissenschaftler an der University of California in Los Angeles, kritisierte bereits 2006

nomen glauben, dass Lügner in dieser Zeit besonders viel Unsicherheit und Stress erlebten, denn deren Insula sowie der frontale Gyrus der rechten Hirnhälfte wiesen höhere Aktivität auf als bei ehrlichen Probanden. Dies gelte ausgerechnet für jene Fälle, in denen die Versuchspersonen gar kein Versprechen abgeben konnten! Die Erklärung der Forscher, die Lügner würden sich hier mit Blick auf zu brechende Versprechen besonders unsicher fühlen, ist daher nicht nur unplausibel, sondern sogar logisch unmöglich.

Für den dritten und letzten Abschnitt, die eigene Entscheidung über das Teilen des Profits, fanden die Forscher bei den Wortbrechern stärkere Aktivität im vorderen Zingulum sowie im dorsolateralen (am oberen und seitlichen Rand gelegenen) Teil des Frontalhirns. Das Feuern des Zingulums zeige den emotionalen Konflikt an, das des Frontalhirnareals dessen mentale Kontrolle. Das Resultat entspreche früher gefundenen Hirnaktivierungen bei Experimenten, in denen Versuchspersonen lügen mussten. So wie diese ihr wahres Wissen unterdrücken müssten, hielten auch die Wortbrecher ihre ehrliche Reaktion offenbar im Zaum, speku-

»Standardmodus« des Gehirns. Doch 2005 fand er in einer leicht modifizierten fMRT-Studie größere Überlappungen der Hirnaktivitäten bei wahren und gelogenen Antworten.

### **Problematische Statistik**

Zudem registrierte er auch für wahre verglichen mit falschen Aussagen stärkere Aktivierungen – in sage und schreibe 35 abgrenzbaren Arealen. Lügen gingen dagegen nur in vier Regionen mit stärkerem Feuern einher. Das zeigt, wie schnell solche weit reichenden Schlüsse fragwürdig erscheinen können.

Die Züricher und Konstanzer Forscher künden schon im Titel ihrer Arbeit mutig von den »neuronalen Schaltkreisen eines gebrochenen Versprechens«. Es ist zwar durchaus verbreitet, von den Schaltkreisen, Korrelaten oder neuronalen Grundlagen bestimmter kognitiver Funktionen zu sprechen – ganz korrekt ist es aber nicht. So sind die gefundenen Unterschiede statistischer Natur und hängen von vielen Messparametern, Korrektur- und Auswertungsschritten ab. Die Forscher zeigen in den acht Hirnabbildungen ihrer Publikation die aktiven Regionen

bei einer relativ zu den Ergebnissen doppelt so niedrigen statistischen Schwelle – zur besseren Darstellung, wie es heißt.

Die Aktivierungen auf dem ursprünglichen Niveau wiederum würden mindestens zehn Voxel umfassen. Ein Voxel ist die kleinste Volumeneinheit des mit dem Kernspintomografen untersuchten Gewebes und kennzeichnet wie in einem Koordinatensystem einen bestimmten Ort im Gehirn. Bei der Standardeinstellung entspricht das einem Würfel mit einer Kantenlänge von drei Millimetern. Wir sprechen also von 27 Kubikmillimeter Hirngewebe. Je nach Region und Schätzung enthält solch ein Kubus 0,5 bis 3 Millionen Neurone mit 11 bis 27 Milliarden Synapsen und mehr als 100 Kilometern Axonen.

Die kleinste in der Arbeit von Baumgartner und seinen Kollegen beschriebene Einheit ist zehnmal so groß und bei der geringeren statistischen Schwelle noch einmal deutlich umfangreicher. Welche der unzähligen darin enthaltenen Schaltkreise sind nun für das gebrochene Versprechen verantwortlich? Obwohl die

sind keinesfalls in jeder Hirnregion und bei jedem Menschen gleich. Alter, Krankheit, Medikamente und andere Faktoren können sie stark verändern.

Aus pragmatischen Gründen arbeiten Forscher bei solchen Versuchen mit einer festgelegten Annäherung. Alles, was diesem Modell nicht entspricht, wird als Störfaktor aufgefasst und fällt aus den Ergebnissen heraus. Inwieweit die gewonnenen Erregungsmuster wirklich die neuronale Aktivität widerspiegeln, muss die neurobiologische Grundlagenforschung noch zeigen.

### Zweifel an der Bildgebung

Valentine Marcar, Neuropsychologin an der Universität Zürich, und ihre Mitarbeiter fanden beispielsweise 2004 heraus, dass auch ein Rückgang neuronaler Aktivität zu einem Anwachsen des Kernspinsignals führen kann. Ein Team um James Schummers vom Massachusetts Institute of Technology in Cambridge (USA) berichtete 2008 vom Beitrag anderer Zellen, der Astrozyten, zum Durchblutungsanstieg im Hirngewebe. Yevgeniy Sirotin und Ani-

sprechen untersuchen konnten, sondern nur deren Begleitumstände. Es wird nicht darauf eingegangen, wie die Entscheidungen selbst im Gehirn realisiert sein könnten.

Zudem behaupten Baumgartner und seine Kollegen, dass Hirnaktivierungen die unehrlichen Absichten schon entlarven könnten, bevor die Entscheidungen überhaupt getroffen werden. Laut Pressemitteilung der Universität Zürich ist es gar das »wichtigste Ergebnis der Studie«, dass »verräterische Gehirnaktivierungsmuster uns erlauben, eine Prognose über zukünftiges Verhalten abzugeben«. Passend dazu spekuliert Baumgartner darüber, Hirnmessungen könnten eines Tages vielleicht helfen, »betrügerische und kriminelle Machenschaften zu verhindern«.

Das Problem dabei ist: Die Neuroökonom haben überhaupt keine Prognose berechnet! Die beiden Gruppen – ehrlich und unehrlich – wurden anhand des Kriteriums gebildet, ob die Probanden ihre Versprechen hielten oder brachen. Alle nachfolgenden Berechnungen basierten auf dieser Unterscheidung. Das heißt, die

---

## *Forscher müssen einen ungeheuren Aufwand treiben, wenn sie aus der Hirnaktivierung tatsächlich einen **gedanklichen Prozess** oder eine Verhaltensweise prognostizieren wollen*

---

Autoren von Schaltkreisen sprechen, geben sie hierauf keine Antwort. Die Kernspintomografie ist dafür auch nicht das geeignete Verfahren.

Eine Reihe neuerer Forschungsergebnisse lässt fraglich erscheinen, wie gut sich Bildgebungsdaten als Indikator neuronaler Aktivität eignen. fMRT-Scanner messen ein sehr komplexes Signal, das von Stoffwechselprozessen und Veränderungen an den Blutgefäßen abhängt. Diese Vorgänge laufen in einem viel größeren zeitlichen Rahmen ab als dem Millisekudentakt, in dem Neurone feuern. Auch greifen verschiedene physiologische Prozesse unterschiedlich ineinander und

ruddha Das, Neurowissenschaftler an der Columbia University in New York, fanden 2009 heraus, warum fMRT-Signale auch ganz ohne neuronale Aktivitätsunterschiede auftreten können. Kurzum, neuronale Aktivierung ist weder notwendige noch hinreichende Bedingung für die bunten Flecken aus dem Tomografen.

Ein weiteres Problem ergibt sich daraus, dass die Neuroökonom gar nicht die Entscheidungen selbst modelliert haben. Das Versprechen zu geben und es später zu brechen, sind lediglich zwei Momente im gesamten Zeitfenster ihrer Auswertung. Daher spricht vieles dafür, dass die Forscher gar keine gebrochenen Ver-

Forscher haben nicht etwa mit Hilfe der Hirnaktivierung zwischen ehrlichen und unehrlichen Menschen differenziert; sie haben lediglich untersucht, ob die Hirnaktivierung ehrlicher und unehrlicher Menschen in ihrem Experiment unterschiedlich ausfiel.

Für ihre Auswertung war es eine notwendige Voraussetzung, dass sie bereits wussten, wer sein Versprechen (meistens) hält und wer nicht. Das ist keine Vorhersage, sondern eine »Nachhersage«.

Forscher müssen einen ungeheuren Aufwand treiben, wollen sie aus der Hirnaktivierung tatsächlich einen gedanklichen Prozess oder eine Verhaltensweise

---

## Über was die NeuroökonomInnen berichten, ist der Versuch, Gehirnaktivität einem *bestimmten kognitiven Prozess* zuzuordnen – dem der Täuschung

---

prognostizieren. John-Dylan Haynes vom Bernstein Center for Computational Neuroscience in Berlin kann davon ein Lied singen. Mit modernen Verfahren der Mustererkennung trainierte er Computer zunächst in einer Lernphase darauf, bestimmte Hirnaktivierungen mit bestimmten experimentellen Bedingungen zu verknüpfen (siehe G&G 9/2006, S. 60). Erst danach ließ sich anhand neuer Messdaten, die nicht bereits für das Training verwendet worden waren, die Prognose testen: Wie oft kann das Verfahren die Hirnaktivierungen richtig zuordnen und wie oft irrt es?

Jeder Versuch der Vorhersage führt automatisch zu der Frage nach der Trefferquote. Baumgartner und seine Kollegen geben diese Quote nicht an – sie haben sie gar nicht berechnet. Die Forscher unternahmen gleichwohl einen ersten Schritt in die Richtung einer Vorhersage. Man könnte etwa ein Programm zur Mustererkennung auf jene Regionen ansetzen, in denen sich die ehrlichen und unehrlichen Probanden unterscheiden. In der entscheidenden Testphase müsste der Algorithmus dann allerdings selbst die Zuordnung zu einer der beiden Gruppen vornehmen.

Über was die NeuroökonomInnen berichten, ist also der Versuch, eine Gehirnregion einem kognitiven Prozess zuzuordnen – dem der Täuschung. Nichts anderes machen Hirnforscher seit Jahrzehnten und nennen das Hirnkartierung. 1974 fanden beispielsweise David Ingvar und Göran Franzén von der Universität in Lund (Schweden) bei Schizophrenen ungewöhnliche Frontalhirnaktivitäten.

Gemäß der Logik von Baumgartner und Kollegen hätten die Skandinavier damit behaupten können, eine schizophrene Person an ihrer Hirnaktivierung identifizieren zu können. Tatsächlich ist nichts dergleichen der Fall.

Stattdessen stellten andere Forscher danach mal den gleichen, mal den gegenteiligen und mal auch gar keinen Effekt im Frontalhirn von Schizophrenen fest. Mit anderen Worten: Das Kriterium der Abnormalität von Frontalhirnaktivität ist bis heute – 35 Jahre später – nicht gut genug, um eine zuverlässige Diagnose zu ermöglichen.

### Test auf Vorhersagekraft

Im Fall der neuroökonomischen Versuchsanordnung müsste daher konkret überprüft werden, wie gut man anhand der beobachteten Aktivität im Zingulum und in der Insula den Bruch eines Versprechens vorhersagen kann, ohne die Antwort bereits zu kennen. Dabei müssten die Experimente statt mit einer Hand voll Studenten mit einer möglichst repräsentativen Gruppe von Versuchspersonen durchgeführt werden.

Tatsächlich können sich Baumgartner und Kollegen noch nicht einmal sicher sein, dass jede ihrer als unehrlich eingestuften Versuchspersonen tatsächlich das »trügerische« Muster aufwies und umgekehrt keine der ehrlichen. Denn die verwendeten statistischen Verfahren berechnen allein Gemeinsamkeiten, die über viele Probanden hinweg relativ stabil sind. Worin sich die Individuen unterscheiden, erscheint dabei als Störfaktor – obwohl es für die individuelle Entscheidung ausschlaggebend sein kann.

Noch deutlicher wird dies, wenn man die Verhaltensdaten näher betrachtet: Auch die unehrlichen Spieler teilten in 30 Prozent der Fälle das Geld, und immerhin zehn Prozent der ehrlichen behielten es. Diese Ausnahmen haben die Forscher nicht gesondert betrachtet, sondern mit den gegenteiligen Entscheidungen der jeweiligen Gruppenmehrheit in einen Topf geworfen.

Und schließlich: Selbst wenn die Lügner in einer emotionalen Zwickmühle gesteckt haben mögen, es wären durchaus Gegenbeispiele denkbar, für die das nicht zutrifft. Vielleicht gibt es auch abgebrühte Menschen, die beim Bruch eines Versprechens weder einen Konflikt noch ein schlechtes Gewissen verspüren?

Selbst wenn manche der Aussagen von Baumgartner und seinen Kollegen hochspekulativ oder sogar falsch sind, ist ihre Studie doch nicht einfach bloß Unsinn. Sie wirft allerdings mehr Fragen auf, als sie klare Antworten liefert. Von der Aufklärung des »neuronalen Schaltkreises eines gebrochenen Versprechens« und der Vorhersage individuellen Verhaltens kann jedenfalls keine Rede sein. ☞

*Stephan Schleim* ist Forscher und Dozent in der Abteilung für Grundlagen und Geschichte der Psychologie der Universität Groningen.

#### QUELLEN

**Baumgartner, T. et al.:** The Neural Circuitry of a Broken Promise. In: *Neuron* 64, S. 756–770, 2009.

**Greene, J.D. et al.:** An fMRI Investigation of Emotional Engagement in Moral Judgment. In: *Science* 293, S. 2105–2108, 2001.

**Poldrack, R. A.:** Can Cognitive Processes be Inferred from Neuroimaging Data? In: *Trends in Cognitive Science* 10(2), S. 59–63, 2006.

**Schleim, S., Rosier, J.P.:** fMRI in Translation: the Challenges Facing Real-World Applications. In: *Frontiers in Human Neuroscience* 3(63), S. 1–7, 2009.

#### LITERATURTIPP

**Schleim, S.:** Gedankenlesen: Pionierarbeit der Hirnforschung. Heise Verlag, Hannover 2008.

# Psychologie und Bildgebung

Eine Entgegnung auf Stephan Schleims Kritik

VON THOMAS BAUMGARTNER, URS FISCHBACHER UND ERNST FEHR

Der vorangehende Beitrag wirft grundsätzliche Fragen hinsichtlich der Interpretation von Daten aus bildgebenden Studien sowie unseres eigenen experimentellen Vorgehens auf. Wir möchten die Kritik im Folgenden Punkt für Punkt entkräften. Wegen des begrenzten Platzes können wir hier nur kurz auf die wichtigsten Aspekte eingehen. Eine ausführliche Fassung unserer Replik findet der interessierte Leser auf unserer Homepage (Link am Artikelende).

## 1. Misst die Kernspintomografie tatsächlich neuronale Aktivität?

In seinen Ausführungen bezweifelt Stephan Schleim dies grundsätzlich. Damit stellt er Tausende von bildgebenden Studien in Frage – darunter auch die von ihm positiv erwähnten Arbeiten von Daniel Langleben (2005) und Joshua Greene (2004).

Das mittels fMRT gemessene Signal basiert vereinfacht ausgedrückt auf der Tatsache, dass aktive Hirnregionen mehr Sauerstoff in ihren Blutgefäßen aufweisen als weniger aktive Hirnregionen. Dieser Unterschied kann durch fMRT und daran anschließende komplexe Analyseverfahren sichtbar gemacht werden.

In der Tat wurden in den letzten Jahren einige Studien veröffentlicht, welche darauf hinwiesen, dass die Korrelation zwischen dem Kernspinsignal und der tatsächlichen neuronalen Aktivität nicht perfekt ist. Die große Mehrheit der Studien zeigt aber: In den meisten Situationen und Hirnregionen gibt das Kernspinsignal einen recht guten Hinweis auf die entsprechenden neuronalen Aktivitäten.

Hierzu gibt es sehr viele eindrucksvolle Beispiele. Damit wir Handbewegungen ausführen können, müssen Neurone im primären motorischen Kortex feuern, und zwar in jenem Areal, das für die Steuerung der Hand zuständig ist. Wenn man zu diesem Zeitpunkt mittels fMRT Messungen in der Handregion vornimmt, findet man immer in diesem Areal ein erhöhtes Kernspinsignal.

Oder: Damit wir Geräusche wahrnehmen können, müssen Neurone im primären auditorischen Kortex feuern. Wenn man gleichzeitig eine fMRT-Messung durchführt, sieht man immer einen Anstieg des Kernspinsignals in dieser Hirnregion. Solche Beispiele ließen sich fast endlos fortführen. Alle diese methodenübergreifenden Befunde in sensorischen, aber auch vielen für Emotionen zuständigen Arealen zeigen, wie Ekstrom in einem aktuellen Übersichtsartikel zusammenfasst, »dass das Kernspinsignal in vielen Situationen einen guten Indikator für neuronale Aktivität darstellt«.

## 2. Kann man von Hirnscans auf psychologische Vorgänge schließen?

Den bedenkenlosen Umkehrschluss von neuronalen Daten auf psychologische Vorgänge kritisiert Stephan Schleim zu Recht. Die Gefahren des Umkehrschlusses sind insbesondere bei neuen experimentellen Paradigmen – wie demjenigen von Joshua Greene – beträchtlich, weil es dafür noch nicht genügend überprüfte und bestätigte Daten gibt. Wenn jedoch innerhalb einer bestimmten experimentellen Aufgabe immer wieder dieselben psychologischen Befunde und Hirnaktivierungen

miteinander korrelieren, dann erhöht sich die Zuverlässigkeit des Umkehrschlusses im Rahmen dieser Aufgabe erheblich. So ist auch der von Schleim zitierte Aufsatz von Russell Poldrack (2006) zum Großteil gerade der Entwicklung einer Prozedur gewidmet, die solche Schlüsse ermöglichen soll.

Bei unserer Arbeit können wir uns auf sehr viele Studien mit kognitiven oder emotionalen Konfliktparadigmen stützen, die Aktivierungen des vorderen Zinngulums (ACC) und des frontoinsularen Kortex zeigen. Damit ist eine der wichtigsten Voraussetzungen für einen empirisch validen Umkehrschluss gegeben – eine hohe A-priori-Wahrscheinlichkeit, dass die experimentelle Aufgabe tatsächlich die angeführten Hirnaktivierungen hervorruft.

Die zweite Voraussetzung lautet, dass es keine oder kaum andere plausible kognitive Prozesse während der Aufgabe gibt, welche die betreffenden Hirnaktivierungen hervorrufen. Dies ist unseres Erachtens ebenfalls gegeben, da während des Versprechens mit hoher Wahrscheinlichkeit die Überlegung im Vordergrund stand, ob man den Mitspieler täuschen soll – und der damit verbundene Konflikt mit vorherrschenden Moralvorstellungen. Es macht deshalb Sinn – im Rahmen unseres Paradigmas –, von diesen Hirnaktivitäten auf derartige Konflikte zu schließen.

Stephan Schleim irrt, wenn er behauptet, dass wir von »nichts anderem als der neuronalen Aktivität« auf kognitive oder emotionale Konflikte geschlossen hätten. Tatsächlich äußern wir vor dem

Hintergrund zahlreicher Studien, die eine starke Korrelation zwischen Konflikt und Hirnaktivierung in ähnlichen Experimenten belegen, die Vermutung, dass auch in unserem Experiment solch eine Korrelation vorliegt. Diese Vermutung wird nicht zuletzt dadurch gestützt, dass in unseren Daten die besagten Hirnregionen umso stärker aktiviert sind, je häufiger der jeweilige Proband sein Versprechen bricht und daher einem Konflikt ausgesetzt ist.

banden unterscheidet, die später ihr Versprechen hielten. Vor unserer Studie war das neuronale Korrelat während der Abgabe eines Versprechens komplett unerforscht, da es in Lügenparadigmen keine Versprechensphase gibt.

Interessanterweise unterschieden sich unsere Probanden nicht in ihrem Versprechen selbst – alle gaben an, Wort zu halten. Somit wusste zum Zeitpunkt der Versprechensabgabe nicht einmal der Experimentator, ob der jeweilige Proband später sei-

In unserer Studie fanden wir keine neuronalen Korrelate von ehrlichem Verhalten – das heißt, es gibt keine Hirnregion, welche im Vergleich zu unehrlichem Verhalten eine erhöhte Aktivität aufweist. Eine plausible Interpretation dieser Befunde ist, dass Ehrlichkeit keine zusätzlichen hirnphysiologischen Ressourcen beansprucht.

Schleim kritisiert diese Vermutung mit dem Hinweis auf eine Lügenstudie von Daniel Langleben. Wie wir in der ausführlichen Fassung der Replik aufzeigen, unterläuft dem Kritiker hier allerdings ein Interpretationsfehler, weil er die Eigenheiten der Versuchsanordnung von Langleben nicht genügend berücksichtigt. Diese erklären nämlich letztlich, warum die Untersuchung von Langleben eine der ganz wenigen Lügenstudien war, die von einem solchen »Wahrheitskorrelat« berichteten.

---

## Unsere Studie ist ein *wichtiger Schritt* hin zur Vorhersage von falschen Versprechen

---

### 3. Was besagen fehlende Unterschiede in den Reaktionszeiten?

Stephan Schleim kritisiert, dass unsere Annahme eines emotionalen Konflikts bei Wortbrüchen schon deswegen unplausibel sei, weil wir keine unterschiedlichen Reaktionszeiten zwischen unehrlichen und ehrlichen Personen gefunden haben. Diese Kritik lässt völlig außer Acht (wie wir in unserer ausführlichen Fassung der Replik detailliert erörtern), dass Reaktionszeitunterschiede nicht generell bei emotionalen Konflikten auftreten, sondern in erster Linie bei Konflikten im Rahmen von komplexen Entscheidungsproblemen.

So stellte etwa Daniel Langleben 2005 fest, dass längst nicht alle einschlägigen Studien verlängerte Antwortzeiten bei Lügen gefunden haben. Letztlich ist es sogar ein Vorteil, dass wir keine Reaktionszeitunterschiede zwischen gehaltenen und gebrochenen Versprechen fanden, weil die neuronalen Unterschiede zum Zeitpunkt der Versprechensabgabe davon beeinflusst werden könnten.

### 4. Können wir tatsächlich eine Vorhersage machen?

Eine zentrale Frage unserer Studie war, ob wir schon bei der Abgabe des Versprechens einen neuronalen Hinweis darauf finden, was Wortbrecher von jenen Pro-

banden Versprechen treu bleiben würde. Folglich sehen wir es als spannenden Befund an, dass die Hirnaktivierung (aber nicht die Wahl des Versprechens) schon zu diesem Zeitpunkt einen Hinweis darauf gibt, ob das Versprechen später eher gehalten oder gebrochen wird. Mehr noch: Wir finden eine starke Korrelation ( $r = 0,7$ ) zwischen der Anzahl der Versprechensbrüche und der Hirnaktivierung im vorderen Zingulum und im frontoinsularen Kortex. Dies bedeutet, dass insbesondere diejenigen Probanden, welche immer oder sehr häufig ihr Versprechen brachen, mit einer starken Aktivität in diesen Regionen reagierten. Probanden, die selten bis nie ein Versprechen brachen, wiesen dort die geringste Aktivität auf.

Es gibt nun methodische Ansätze, welche es rein theoretisch erlauben würden, unsere Analyse um einiges differenzierter zu gestalten. Allerdings konnten diese Ansätze bisher nicht auf soziale Interaktionsexperimente angewendet werden und scheitern daher an der praktischen Umsetzung. Langer Rede kurzer Sinn: Wir stimmen mit Stephan Schleim überein, dass unsere Studie »einen ersten Schritt in die Richtung einer Vorhersage macht«. Nicht mehr – aber auch nicht weniger.

### 5. Fanden wir neuronale Korrelate von ehrlichem Verhalten?

Fazit: Schleims grundsätzliche Zweifel an der Aussagekraft von fMRT-Daten sind ebenso unbegründet wie die Grundsatzkritik an neuronalen Umkehrschlüssen. Die fehlenden Reaktionszeitunterschiede zwischen falschen und wahren Versprechen sind eher ein Vorteil für unsere Interpretation der Hirnvorgänge, als dass sie diese in Frage stellen. Unsere Studie ist sicher »nur« ein erster Schritt hin zur Vorhersage von falschen Versprechen – ein wichtiger jedoch, der noch nie gemacht wurde. Und die behaupteten neuronalen Korrelate ehrlichen Verhaltens entpuppen sich bei genauerem Hinsehen als methodisch bedingt. Angesichts dieser mageren Ausbeute muss sich der Kritiker schon die Frage gefallen lassen, ob er nicht selbst über sein Ziel hinausgeschossen ist. ~

*Thomas Baumgartner* arbeitet als Neurowissenschaftler am Institut für Empirische Wirtschaftsforschung der Universität Zürich, das von *Ernst Fehr* geleitet wird.

*Urs Fischbacher* ist Wirtschaftswissenschaftler an der Universität Konstanz.

Eine ausführliche Fassung dieser Replik finden Sie im Internet unter:

[www.iew.uzh.ch/institute/people/baumgartner/publications1.html](http://www.iew.uzh.ch/institute/people/baumgartner/publications1.html)