

Übersichtsartikel

Sabine Windmann* und Grit Hein

Altruismus aus Sicht der Sozialen Neurowissenschaften

<https://doi.org/10.1515/nf-2017-0047>

Zusammenfassung: Altruismus ist ein verblüffendes Phänomen, vor allem aus Sicht der Biologie und der Ökonomie. Warum geben Individuen anderen etwas von ihren Ressourcen ab und verringern damit ihre eigenen Möglichkeiten? Die Antwort auf diese Frage kann auf ultimer oder proximaler Ebene gesucht werden. Die Sozialen Neurowissenschaften versuchen, die Gehirnmechanismen zu spezifizieren, die Menschen dazu antreiben, altruistisch zu handeln, denn äußerlich gleiches Verhalten kann durch unterschiedliche Motive bedingt sein. Aktivierungen und funktionelle Konnektivitäten der Anterioren Insula und der Temporoparietalen Junction spielen bestimmbare Rollen bei empathiebasiertem versus strategischem Altruismus, während der dorsolaterale präfrontale Kortex, neben anderen Regionen, bei punitiven Formen von Altruismus involviert ist. Zukünftige Forschungsarbeiten könnten sich auf die Verarbeitung von Ambiguität und Konflikt in der Verfolgung altruistischer Absichten beziehen.

Schlüsselwörter: Kognitiv; Hirnbildgebung; Motivation; Evolution; Mensch

Konzeptionen von Altruismus

Altruismus (lat. *alter*, the other) kann definiert werden als kostspieliges Verhalten, welches dem Wohlergehen anderer zu Gute kommt. Übereinstimmend mit unserem Alltagsverständnis wird es in der Wissenschaft mit prosozialem Verhalten und Kooperation in Verbindung gebracht

(Nowak, 2006). Mit Altruismus kontrastiert werden können Egoismus/Individualismus (Maximierung der eigenen Ressourcen), Konkurrenz (Maximierung der Differenz zwischen eigenen Ressourcen und denjenigen anderer) und Boshaftigkeit (Minimierung der Ressourcen anderer; siehe Murphy und Ackermann, 2014).

In den letzten ein bis zwei Jahrzehnten wurden Altruismus und Kooperation verstärkt beforscht sowohl von theoretischen als auch empirischen Disziplinen innerhalb der Ökonomie, Biologie, Psychologie und den Neurowissenschaften. Einer der Gründe für das enorme Interesse ist die schwierige Vereinbarkeit von Altruismus mit klassischen theoretischen Annahmen. Wie kann Darwins Evolutionstheorie erklären, dass Altruismus aus natürlicher Selektion hervorging, wenn es kostspielig und fremdnutzend ist? Warum sollte sich aus Sicht der klassischen Ökonomie, die an die Maximierung subjektiver Nutzen glaubt, jemals jemand dafür entscheiden, Kosten auf sich zu nehmen, um damit anderen zu nutzen?

Solche theoretischen Konflikte entstehen vor allem, wenn Altruismus in Termini von Input-Output-Relationen bzw. in Termini reproduktiver Fitness definiert wird. Aus naturwissenschaftlicher Perspektive sind solche Definitionen notwendig, denn sie erlauben die rein objektive, beobachtungsbasierte Bestimmung des zu erklärenden Phänomens. Jedwedes Verhalten, das kostspielig ist für den Akteur, aber jemand anderem nützt, wird als altruistisch angesehen, gleichgültig, ob es von absichtsvollen Menschen gezeigt wird (Becker und Eagly, 2004) oder von Fischen (Daniels, 1981), Bakterien (Lee et al., 2010) oder Pflanzen (Murphy und Dudley, 2009).

Andere Forscher, vor allem Psychologen, finden ein solches „behavioristisches“ Verständnis von Altruismus unbefriedigend und verweisen daher auf innere Zustände wie die *Intention*, jemand anderem zu nutzen. Beispielsweise definieren Batson und Shaw (1991) Altruismus als „motivationalen Zustand mit dem ultimativen Ziel, das Wohlergehen eines anderen zu erhöhen“ (S. 108, Übersetzung der Verfasserinnen). Aus dieser Perspektive müssen Forscher nicht in erster Linie das Verhalten beobachten, um über das Vorliegen von Altruismus zu entscheiden, sondern die aktiven Zielvorstellungen des Akteurs be-

*Korrespondenzautor: Sabine Windmann, Psychologisches Institut der Goethe-Universität Frankfurt am Main, Theodor-W.-Adorno Platz 6, 60323 Frankfurt am Main, E-Mail: s.windmann@psych.uni-frankfurt.de

Grit Hein, Institut für Medizinische Psychologie, Johann Wolfgang Goethe Universität, Heinrich Hoffmann Straße 10, 60528 Frankfurt/Main; Abteilung für Translationale Soziale Neurowissenschaften, Translational Social Neuroscience Unit, Julius-Maximilians-Universität Würzburg, Klinik für Psychiatrie und Psychosomatik, Margarete-Höppel-Platz 1, 97080 Würzburg, E-Mail: Hein_G@ukw.de

stimmen. Bemerkenswerterweise akzeptieren diejenigen Forscher, die solche Definitionen verwenden, oft auch proximate Variablen bei der Bewertung des Kosten-Nutzen-Verhältnisses für den Akteur. So werden Gefühle der „Herzerwärmung“, die der Akteur bei erfolgreich geleisteter Hilfe empfinden kann, für diesen als wertvoll (nützlich) eingeschätzt und daher als Evidenz gegen das Vorliegen von „echter“ Selbstlosigkeit gewertet (Batson und Shaw, 1991; Harbaugh et al., 2007).

In der vorliegenden Übersicht wollen wir nicht versuchen, metatheoretische Differenzen zwischen den Disziplinen zu diskutieren oder aufzulösen, obwohl wir betonen, dass man die Unterschiede im Hinterkopf behalten sollte beim Studium der Literatur. Stattdessen werden wir ultimate und proximate Erklärungen von Altruismus erläutern und auf den Beitrag der Sozialen Neurowissenschaften eingehen. Insgesamt finden wir, dass die Altruismus-Forschung einige der traditionellen Grenzen zwischen objektiv und subjektiv, biologisch und kulturell, proximat und ultimat in Frage stellt (Laland et al., 2011).

Ultimate Erklärungen

Schon lang ist bekannt, dass natürliche Selektion Altruismus fördern kann, wenn Geber und Rezipient des altruistischen Verhaltens miteinander verwandt sind und die Kosten für die altruistische Handlung geringer sind als der Grad der Verwandtschaft. Diese Verhältnisbildung pointierte John B. S. Haldane, als er scherzte: „Ich würde in den Fluss springen, um entweder zwei meiner Brüder zu retten oder acht meiner Cousins“ (Nowak, 2006).

Um jedoch Altruismus zwischen Nichtverwandten zu erklären, müssen evolutionäre Theorien von indirekten Nutzen für den Akteur ausgehen, bzw. für die Gruppe oder das Netzwerk, in dem der Akteur lebt (Nowak, 2006, 2012). Der Gedanke ist, kurz gesagt, dass altruistische Individuen, die durch ihre Handlungen das Wohlergehen ihrer sozialen Umgebung steigern, irgendwann „zurückprofitieren“, entweder persönlich oder indirekt durch die Stärkung ihrer Bezugsgruppe. Nowak spricht von einem „snuggle for survival“ (Nowak, 2012, p. 34).

Zwei zentrale Strategien sollen hier hervorgehoben werden aufgrund ihrer möglichen proximatsten Implikationen. Die erste ist *Reziprozität*, ein Mechanismus, durch den gesichert wird, dass das Äquivalent des Wertes einer altruistischen Handlung zum Akteur zurückkehrt (Milinski, 2016). Reziprozität gibt es in zwei Formen: Bei „direkter Reziprozität“ erhält der altruistische Akteur selbst eine gleichwertige Belohnung zurück von dem Rezipienten;

die beiden kooperieren also nach dem Prinzip „Eine Hand wäscht die andere“. Bei „indirekter Reziprozität“ hingegen sind mehr als zwei Parteien involviert. Sie gibt es in zwei Formen. Erstens, bei „abwärtsgerichteter“ indirekter Reziprozität erwirbt der altruistische Akteur durch seine Handlung einen guten Ruf, der gewährleistet, dass er in zukünftigen Situationen leichter Kooperationspartner findet (auch Sexualpartner), sodass seine reproduktive Fitness erhöht wird (Iredale und Vugt, 2009). Zweitens, bei „aufwärtsgerichteter“ indirekter Reziprozität werden die erhaltenen Belohnungen undifferenziert weitergeleitet zu anderen Gruppenmitgliedern, analog zu Kants Imperativ: „Handle nur nach derjenigen Maxime, durch die du zugleich wollen kannst, dass sie ein allgemeines Gesetz werde“.

Reziprozität als Prinzip macht Altruismus kompatibel mit Darwins Evolutionstheorie. Allerdings hat es einen schwerwiegenden Nachteil, besonders im Fall großer Gruppen, in denen der Beitrag des einzelnen zum Wohlergehen der Gruppe nicht gut verfolgt werden kann: die Gefahr des Trittbrettfahrens. Individuen, die vom Altruismus anderer Gruppenmitglieder profitieren, aber ihrerseits keinen fairen Beitrag zum Gruppenwohl leisten, weisen eine höhere Erfolgsbilanz auf als diejenigen, die kooperieren. Sie behalten von vornherein mehr von ihren Ressourcen und profitieren zusätzlich von dem Investment der anderen Gruppenmitglieder. Unterm Strich sollte daher ihre reproduktive Fitness höher sein als die der anderen Gruppenmitglieder. Und je mehr Nachkommen sie produzieren, desto mehr bringen sie kooperierende Individuen zum Aussterben, insofern Selektionsdruck herrscht. Die Bedingungen solcher Dynamiken werden mit computationalen Modellierungsstudien untersucht (Le und Boyd, 2007; Nowak und Sigmund, 1998; Riolo et al., 2001).

Trittbrettfahrer stellen also Theorien zur Erklärung von Altruismus unter Nichtverwandten vor ein großes Problem. Es gibt jedoch Hilfe, und sie kommt aus der verhaltensökonomischen Forschung. Ernst Fehr und seine Kollegen haben in ökonomischen Spielen im Labor gezeigt, dass kooperative Gruppenmitglieder bereit sind, Kosten auf sich zu nehmen, um Trittbrettfahrer zu bestrafen (Fehr und Fischbacher, 2003; Fehr und Gächter, 2002). Durch die Bestrafung wird die Bilanz der Trittbrettfahrer reduziert, was sie zukünftig von unfairem Verhalten abhalten kann. Ohne die Möglichkeit einer solchen Bestrafung bricht die Kooperation von Gruppen zusammen, aber unter der Androhung von Strafe bleibt sie aufrechterhalten (Fehr und Fischbacher, 2003; Fehr und Gächter, 2002).

Kostspielige Bestrafung ist der zweite zentrale Mechanismus von Altruismus mit hohem Potenzial für die Sozialen Neurowissenschaften. Aus objektiver Perspektive

Tabelle 1: Mögliche Erklärungen für die Entstehung und Aufrechterhaltung von Altruismus

Typ Altruismus	Hilfeverhalten (belohnend, Kooperation anregend)	Kostspielige Bestrafung (konfrontativ, gegen Unfairness und Delinquenz gerichtet)
Ultimate Erklärung	Genetische Verwandtschaft Reziprozität, direkt oder indirekt Gruppenselektion Netzwerk-Selektion	Gruppenselektion Genetisch-kulturelle Coevolution
Proximate Erklärung	Empathie Perspektivenübernahme Erwartung von Gegenseitigkeit Unfairnessaversion	Unfairnessaversion Normatives Denken Wut, Rachegefühle, Dominanzstreben Moralische Überzeugungen

ist die Bestrafung selbstlos, weil sie den Bestrafer kostet, während sie der Gemeinschaft nützt, indem sie Fairness-Normen sichert. Sie erfolgt in besonders effektiver und überzeugender Weise, wenn sie von Autoritäten oder unbeteiligten Institutionen vollzogen wird. Mindestens auf dieser hoch organisierten Ebene ist kostspielige Bestrafung wahrscheinlich ein Produkt genetisch-kultureller Coevolution (Bowles et al., 2003; Fehr und Fischbacher, 2003; Henrich et al., 2006; Nowak, 2006).

Proximate Erklärungen

Jedes evolvierte Verhalten muss einen proximalen Mechanismus einschließen, der den Organismus dazu bringt, das Verhalten zu zeigen. Wilson (1992, p. 62) fragt: „Sind Verhaltensweisen, die als altruistisch im evolutionären Sinn gelten, notwendigerweise verursacht durch proximale Mechanismen, die als altruistisch im psychologischen Sinn gelten?“ (Übersetzung der Verfasserinnen). Hier kommen die Möglichkeiten der Sozialen Neurowissenschaften ins Spiel. Mithilfe von Gehirnbildgebung und anderen Methoden können diese die proximalen Mechanismen, die altruistischem Verhalten unterliegen, spezifizieren und dissoziieren und in Beziehung setzen zu den dahintersteckenden Motiven (siehe Tabelle 1), und zwar ohne angewiesen zu sein auf Introspektion und Selbstbericht (Camerer, 2008; Hein et al., 2016). Objektive Korrelate für subjektive Motive zu identifizieren kann sehr wichtig sein bei Konzepten, die stark der sozialen Erwünschtheit unterliegen und daher anfällig sind für subjektive Verzerrungen. Natürlich muss man gegenüber der Idee eines „Mind-Reading“ kritisch bleiben aus methodologischen (Poldrack, 2011) sowie ethischen Gründen (Evers und Sigman, 2013). Aus einzelnen Studien allein (erst recht nicht aus Gehirndaten einzelner Probanden) lassen sich normalerweise keine subjektiven oder motivationalen Zustände

ablesen, hauptsächlich wegen der hohen interindividuellen Variabilität der strukturellen und funktionellen Neuroanatomie sowie der hohen Vernetztheit und Multifunktionalität kortikaler Areale. Daher können die Sozialen Neurowissenschaften nicht einfach Gehirndaten erheben und direkt interpretieren; vielmehr müssen sie spezielle experimentelle Bedingungen realisieren und vergleichen bzw. die Motive der Probanden variieren oder bestimmen. Anschließend setzen sie diese in Beziehung zu den gemessenen neuronalen Korrelaten und Dynamiken.

Üblicherweise treffen die Probanden in solchen Experimenten Entscheidungen, die einer oder mehreren anderen Person zu Gute kommen, aber mit Kosten für sie selbst verbunden sind. In vielen Studien sind diese Kosten monetär (Hein et al., 2016; Hutcherson et al., 2015; Morishima et al., 2012); eine Übersicht experimenteller Paradigmen bieten Fehr und Camerer, 2007 und Sanfey, 2007. In einigen Studien ist eine altruistische Entscheidung aber auch an physische Kosten, z. B. an einen schmerzhaften Schock am Handrücken, gekoppelt (Hein et al., 2011; Hein et al., 2010). Während die Probanden ihre Entscheidungen treffen oder Einschätzungen abgeben, werden Korrelate ihrer Hirnaktivierung, Konnektivitäten zwischen verschiedenen Hirnarealen, hormoneller Status oder genetische Parameter erhoben, bzw. untersucht, ob diese Maße mit den Entscheidungen oder Präferenzen der Probanden in Verbindung stehen. Das Gesamtmuster der Daten kann über die neuronale Implementation der Motive informieren, die den altruistischen Entscheidungen zugrunde liegen.

Mittlerweile zeigen zahlreiche Befunde, dass altruistische Entscheidungen auf Veränderungen von neuronalen Aktivierungen bzw. neuronalen Konnektivitäten in Netzwerken von Hirnregionen basieren, die allgemein mit sozialen Prozessen in Verbindung gebracht werden (Adolphs, 2009; Rilling und Sanfey, 2011; Van Overwalle, 2009). Manchmal werden diese Hirnregionen auch als das „soziale Gehirn“ bezeichnet. Im Kontext von altruistischen

Entscheidungen spielen unter anderem die Insel, das Striatum, der anteriore zinguläre Kortex (ACC), die sogenannte Temporoparietale Junction (TPJ), und der dorsolaterale präfrontale Kortex (DLPFC; Abbildung 1) eine Rolle. Diese Regionen sind jedoch auch bei nicht-sozialen Anforderungen aktiviert. Daher wird angenommen, dass sie mit übergeordneten Funktionen korrelieren, die dann auch in sozialen Kontexten benutzt werden. Die Insula wird beispielsweise mit der Verarbeitung interozeptiver (Craig, 2009) und viszeral-sensorischer Reize (z. B. Schmerz; (Critchley und Harrison, 2013)) in Verbindung gebracht und korreliert im sozialen Kontext mit der Beobachtung und Simulation solcher emotionaler und körperlicher Zustände bei anderen Personen (z. B. Empathie für den Schmerz eines anderen (Zaki et al., 2016)). Das Striatum ist reich an Dopamin, d. h. einer der wichtigsten Neurotransmitter für die Übermittlung von Belohnungssignalen. Darauf basierend spielt das Striatum eine zentrale Rolle bei der neuronalen Verarbeitung von Belohnungen in sozialen wie auch nicht-sozialen Kontexten (Schultz, 2017). Der anteriore zinguläre Kortex steht mit der Verarbeitung von Fehlern, Belohnungen und Konflikten (z. B. zwischen Stimuluskategorien) in Verbindung (Kolling et al., 2016), anhand derer flexibles, adaptives Verhalten in sozialen und nicht-sozialen Situationen initiiert wird (Shackman et al., 2011). Im Folgenden werden zwei Beispielstudien besprochen, bei denen Hirnaktivierungen und neuronale Konnektivitäten zur Untersuchung von altruistischen Verhalten und dessen Motiven eingesetzt wurden.

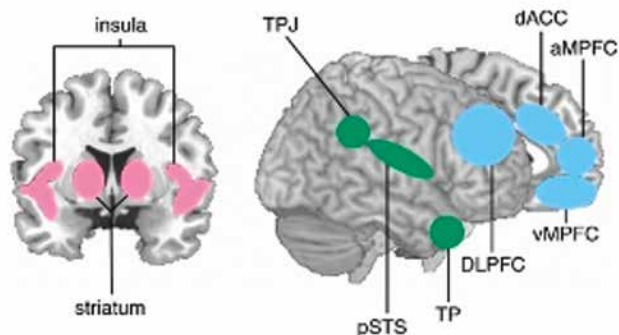


Abb. 1: Übersicht von Hirnregionen, die mit der Verarbeitung sozialer Prozesse in Verbindung stehen. TPJ = Temporal Parietal Junction; dACC = dorsaler Teil des anterioren zingulären Kortex, DLPFC = dorsolateral präfrontaler Kortex, pSTS = posteriorer Teil des superioren temporären Sulcus, TP = temporaler Pol; vMPFC = ventraler Teil des medialen präfrontalen Kortex, aMPFC = anteriorer Teil des medialen präfrontalen Kortex.

In einer Beispielstudie (Hein et al., 2010) wurde das Motiv untersucht, welches dem Mangel an altruistischem Ver-

halten gegenüber Mitgliedern von Fremdgruppen zugrunde liegt. Die Probanden beobachteten, wie ein Mitglied ihrer eigenen Gruppe (ingroup member) oder ein Mitglied einer Fremdgruppe (outgroup member) Schmerz am Handrücken erhielt. Sie konnten entscheiden, die Hälfte dieses Schmerzes auf ihre eigene Hand umzuleiten und damit den Schmerz der anderen Person zu halbieren, d. h. der anderen Person zu helfen, indem sie physische Kosten (Schmerz) auf sich nahmen. Im Durchschnitt halfen die Probanden dem Mitglied der eigenen Gruppe signifikant häufiger als dem Mitglied der Fremdgruppe. Interessanterweise wurde das Ausmaß, in dem die eigene Gruppe bevorzugt wurde, von Unterschieden in der Aktivität der Anterioren Insula (AI) vorhergesagt, welche die Probanden vorher beim Beobachten des Schmerzes des Eigen- und Fremdgruppenmitglieds gezeigt hatten, und welche mit Empathie korrelierte. Je stärker ein Proband die AI beim Beobachten des Schmerzes in der Eigengruppe, aber nicht in der Fremdgruppe aktivierte, desto wahrscheinlicher war es, dass er dem Eigengruppenmitglied, aber nicht dem Fremdgruppenmitglied später half (Abbildung 2). Probanden, die besonders negativ gegenüber dem Fremdgruppenmitglied eingestellt waren, aktivierten außerdem das ventrale Striatum, also einen Teil des Belohnungsverarbeitungssystems, wenn sie Schmerz in der Fremdgruppe beobachteten. Das Ausmaß dieser Aktivierung sagte fehlendes Helfen gegenüber dem Fremdgruppenmitglied vorher.

In einer zusätzlichen Analyse wurde verglichen, wie gut Hirnmaße (Aktivierung in der AI und im ventralen Striatum) und Fragebogenmaße (Skalen, die die Empathie und die Einstellung des Probanden bezüglich des Eigen- und Fremdgruppenmitglieds erfassten) späteres helfendes Verhalten gegenüber dem Eigengruppen- und Fremdgruppenmitglied vorhersagen. Es zeigte sich, dass Verhalten gegenüber dem Eigengruppenmitglied von Hirnmaßen und Fragebogenmaßen gleich gut vorhergesagt wurde, dass jedoch Hirnmaße deutlich besser zur Vorhersage des Verhaltens gegenüber der Fremdgruppe geeignet waren. Insgesamt weisen die Befunde darauf hin, dass Unterschiede in altruistischen Verhalten gegenüber Eigen- und Fremdgruppenmitgliedern durch Unterschiede in Empathie motiviert sein können, und dass Hirnmaße Verhalten in sozial sensiblen Situationen vorhersagen können, z. B. gegenüber Mitgliedern einer ungeliebten Fremdgruppe.

Eine weitere aktuelle Beispielstudie benutzte monetäre Kosten, um zwei Motive voneinander zu trennen, die zum gleichen altruistischen Verhalten führten (Hein et al., 2016). Im ersten Teil der Studie wurde entweder ein Empathie- oder ein Reziprozitätsmotiv aktiviert. Um das Empathiemotiv zu aktivieren, beobachteten die Probanden, wie

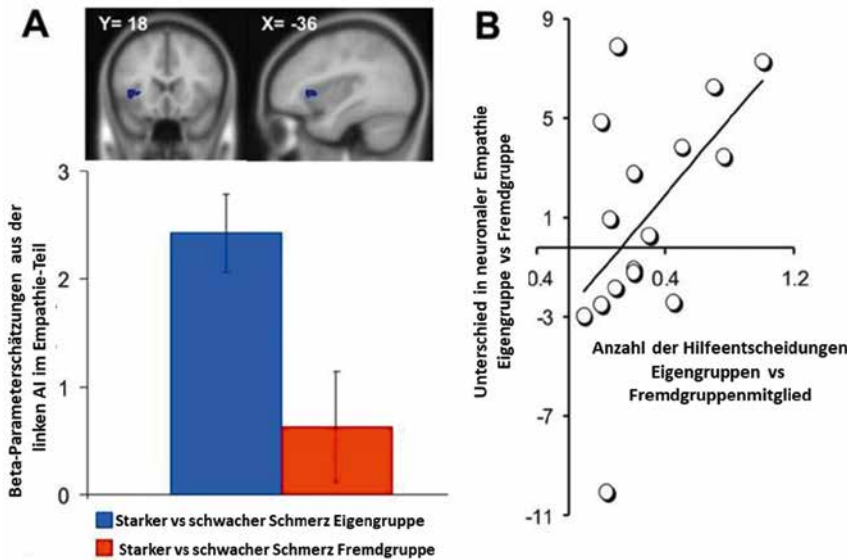


Abb. 2: Befunde einer Beispielstudie.

A) Die Ergebnisse zeigten eine stärkere durchschnittliche Aktivierung der linken anterioren Insel (AI) beim Beobachten des Schmerzes eines Mitglieds der Gruppe des Probanden (ingroup – IG) im Vergleich zum Beobachten des Schmerzes eines Mitglieds einer fremden Gruppe (outgroup – OG). B) Die Unterschiede in empathiebezogener Aktivierung in der AI zwischen der ingroup und der outgroup Bedingung sagten Unterschiede im späteren helfenden Verhalten voraus. Je stärker ein Proband die AI in der ingroup und nicht der outgroup Bedingung aktivierte, desto wahrscheinlicher war es, dass der Proband später dem Mitglied seiner Gruppe und nicht dem Mitglied der fremden Gruppe half. Originalabbildung von Hein et al., 2010, aus *Neuron* 68, S. 149–160.

eine andere Person eine Schmerzstimulation am Handrücken erhielt. Das Reziprozitätsmotiv wurde aktiviert, indem eine andere Person dem Proband einen Gefallen tat. Basierend auf diesen beiden Motiven trafen die Probanden dann ökonomische Entscheidungen, in denen sie Punkte zwischen sich und einer anderen Person aufteilten, die später in Geld umgewandelt wurden. Die Aktivierung des Empathie- und des Reziprozitätsmotivs führten zu einer vergleichbaren Anzahl von altruistischen Entscheidungen (d. h. Entscheidungen, in denen der Proband die Punkte des anderen maximierte und dabei gleichzeitig seine Punkte minimierte). Das heißt, dass die beiden Motive nicht anhand des Verhaltens der Probanden unterschieden werden konnten. Auch bezüglich funktionaler Hirnaktivierungen zeigten sich keine Unterschiede. Empathie-basierte und reziprozitätsbasierte Entscheidungen aktivierten ein vergleichbares neuronales Netzwerk, welches aus der AI, dem ventralen Striatum und dem ACC bestand. In einem nächsten Schritt analysierten die Autoren die neuronalen Konnektivitäten zwischen diesen Regionen mithilfe von Dynamic Causal Modeling (DCM), ein Verfahren, durch das aufgrund von Annahmen über das Zustandekommen des gemessenen Signals effektive Verbindungen zwischen aktivierten Regionen geschätzt werden sollen. Es zeigte sich, dass empathiebasierte und reziprozitätsbasierte altruistische Entscheidungen zwar das gleiche Netzwerk aktivierten, aber verschiedene neuronale Konnektivitätsmuster innerhalb dieses Netzwerks hervorriefen (Abbildung 3). Anhand dieser Unterschiede in den neuronalen Konnektivitätsmustern konnte statistisch vorhergesagt werden, ob die altruistische Entscheidung einer Person von Empathie oder Reziprozität getrieben war.

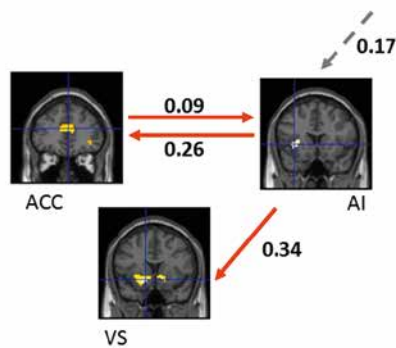
Als ein weiterer interessanter Befund wurde gezeigt, dass vorwiegend egoistische und vorwiegend prosoziale Personen unterschiedlich auf die Aktivierung des Empathie- und des Reziprozitätsmotivs reagierten. Überwiegend egoistische Personen zeigten mehr altruistische Entscheidungen nach der Aktivierung des Empathiemotivs. Im Gegensatz erhöhten Personen, die schon „von Hause“ aus prosozial eingestellt waren, ihre altruistischen Entscheidungen nur nach Aktivierung des Reziprozitätsmotivs, und nicht nach der Aktivierung von Empathie.

Die beiden Studien von Hein et al. (2010) und Hein et al. (2016) zeigen exemplarisch die Methoden und Paradigmen des noch jungen Feldes der Sozialen Neurowissenschaften auf, die in weiteren Studien konsolidiert, optimiert und ständig weiterentwickelt werden müssen.

Belohnende versus punitive Formen von Altruismus

Die bislang beschriebene Forschung hat diverse Motive spezifiziert, die hinter kostspieligem Helfen stecken (Helfen im Sinne vom Teilen materieller oder physischer Ressourcen), also einem belohnenden Verhalten, das Kooperation fördert und so das Wohlergehen aller maximiert. Wie verhalten sich diese Befunde jedoch zum altruistischen Bestrafen, dem zweiten Schlüsselmechanismus von Kooperation? Interessanterweise haben Verhaltensexperimente mit ökonomischen Spielen gezeigt, dass diese beiden Typen von Verhaltensweisen, kostspieliges Belohnen einerseits und kostspieliges Bestrafen andererseits,

▪ Reziprozität



▪ Empathie

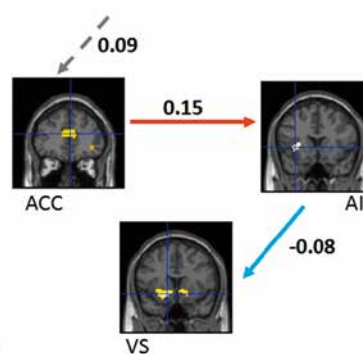


Abb. 3: Befunde einer Beispielstudie, welche unterschiedliche Motive anhand von neuronalen Konnektivitäten klassifizierte. Durchgestrichene Pfeile zeigen an, wie stark das Ausmaß der Aktivierung einer Region die Aktivierung einer anderen Region verändert (Effektive Konnektivität). Gestrichelte Pfeile zeigen Veränderung der Aktivierung innerhalb einer Region in Abhängigkeit von der experimentellen Manipulation (Input) an. Die Zahlen sind Modell-Parameter, die die Stärke der effektiven Konnektivitäten oder Inputs reflektieren. ACC = anteriorer zingulärer Kortex, AI = anteriore Insel, VS = ventrales Striatum.

im Grunde unkorreliert sind (Peysakhovich et al., 2014; Yamagishi et al., 2012). Erneut können die Sozialen Neurowissenschaften dabei helfen, die Unterschiede in Termini der proximalen Mechanismen zu spezifizieren, speziell der neuronalen Basis der zugrunde liegenden Motive.

Im Hinblick auf zugrunde liegende emotionale Prozesse hat eine Reihe von Studien gezeigt, dass kostspieliges Helfen/Belohnen von „Herzerwärmung“ und ähnlichen positiven Gefühlen begleitet ist (Harbaugh et al., 2007; Hu et al., 2016a; Rand et al., 2015). Demgegenüber ist kostspieliges Bestrafen von Wut und Racheglusten geprägt (Crockett et al., 2014; Fehr und Gächter, 2002; Seip et al., 2009; Singer et al., 2006). Insofern wirkt letzteres eher kompetitiv und konfrontativ, ganz entgegen den fürsorglichen prosozialen Absichten, die bei belohnendem Hilfeverhalten eine Rolle spielen.

Bildgebende Verfahren bestätigen, dass die beiden Verhaltensweisen mit unterschiedlichen neuronalen Mechanismen verbunden sind. Obwohl beide ventrale und dorsale Teile des Striatums aktivieren als anreizvermittelnde Struktur (für Hilfeverhalten, (Genevsky et al., 2013; Harbaugh et al., 2007; Hein et al., 2010; Kuss et al., 2013), für Bestrafung (Buckholz et al., 2008; de Quervain et al., 2004; Hu et al., 2015; Strobel et al., 2011), involviert speziell Bestrafung recht konsistent den (typischerweise rechten) DLPFC. Dies wurde nicht nur mit funktionellem MRI gezeigt (Buckholz et al., 2008; Buckholz et al., 2015; Sanfey et al., 2003), sondern auch mit repetitiver transkranieller Gehirnstimulation (Buckholz et al., 2015; Knoch et al., 2008; Strang et al., 2015) und Enzephalografie im Ruhezustand (Knoch et al., 2010). Der DLPFC ist bekannt für seine Rolle in der Modulation automatischer Entscheidungsfindung und Handlungsauswahl; seine Rolle im Zusammenhang mit kostspieligem selbstlosen Verhalten könnte also in der Vermittlung von Information über soziale Normen,

moralische Werte und ähnlichen abstrakten Gedanken liegen, wodurch dominante reflexive Impulse daran gehindert werden, das Verhalten zu bestimmen (Feng et al., 2015). Demgegenüber sollen fürsorgliche Tendenzen, die dem Helfen und Teilen zugrunde liegen, entweder durch affektive Empathie gegenüber der/n bedürftigen Person/en angetrieben werden, dabei die AI und den medialen PFC involvierend, oder durch kognitive Empathie und Perspektivenübernahme, basierend auf der TPJ, neben anderen Arealen (Haas et al., 2015; Hein et al., 2010; Morelli et al., 2014; Strombach et al., 2015; Tusche et al., 2016).

Schlussfolgernd scheinen Hilfeverhalten und kostspielige Bestrafung, zwei Verhaltensweisen, die ultimat beide der evolutionären Definition von Altruismus genügen, proximat auf unterschiedlichen neuronalen Netzwerken zu basieren. Zwar sind konkrete Interpretationen über die funktionale Rolle der beteiligten Strukturen zunächst vorläufig und spekulativ, teilweise auch simplifizierend, sie orientieren sich jedoch an der Gesamtschau der Befundlage, indem sie die Ergebnisse im Lichte von etablierten Befunden und Theorien aus den Kognitiven Neurowissenschaften betrachten. Über die Zeit soll auf diese Weise ein Puzzle zusammengesetzt werden, das die verschiedenen Facetten von Altruismus sowie die zugrunde liegenden Prinzipien und Prozesse auf neuronaler Ebene aufzeigt und miteinander in Beziehung setzt.

Ausblick

Zusammenfassend können die Sozialen Neurowissenschaften dabei helfen, durch die Untersuchung neuronaler Korrelate die unterschiedlichen proximalen Motive zu spezifizieren, die Verhaltensweisen zugrunde liegen, welche die evolutionäre Definition von Altruismus erfüllen.

Einige dieser Mechanismen basieren auf empathischer Fürsorge, andere rekrutieren strategisches Denken, um über den Weg der Kooperation eigene Vorteile zu erlangen. Wiederum andere Verhaltensweisen scheinen in der Motivation zu wurzeln, anderen Menschen Normen aufzuerlegen, um Fairness und andere abstrakte Ziele zu erreichen anstelle direkter Belohnungen. So können die Sozialen Neurowissenschaften Interpretationshinweise liefern, um die proximalen Motive für Altruismus mit grundlegenden kognitiven, emotionalen und motivationalen Prozessen in Beziehung zu setzen. Um Kooperation im großen gesellschaftlichen Maßstab zu erreichen unter Menschen, die einander meist nicht näher kennen, ist anscheinend eine Bandbreite von proximalen Motiven erforderlich, von denen nur einige altruistisch sind im psychologischen Sinne.

Zukünftige Arbeiten könnten die proximalen Ursachen von belohnendem und bestrafendem Verhalten direkt kontrastieren (experimentelle Ansätze hierfür bieten (Hu et al., 2016b; Hu et al., 2015)). Sind es tatsächlich prosoziale moralische Werte und Kognitionen, die beim altruistischen Bestrafen aktiviert werden, unter Mitwirkung des DLPFC, oder spielen möglicherweise soziale Dominanzbestrebungen eine Rolle (z. B. das Bedürfnis nach Homogenisierung der Gruppe)? Wie sind zivilcouragierte Ambitionen einzuordnen (z. B. Whistleblowing), altruistische Akte, die über die Orientierung an sozialen Normen hinausgehen und hohe Risiken bergen, im Unterschied zum altruistischen Bestrafen? Welche Rolle spielen Empathie und Emotionskontrolle, wenn fürsorgliche Intentionen nur durch soziale Konfrontation oder durch die Induktion aversiver Zustände im Empfänger der Hilfe verfolgt werden können (e.g. Lopez-Perez et al., 2017)? Einige dieser Fragen zielen in die Richtung von Motivkonflikten, die das soziale Gehirn auflösen muss. Sie zeigen die Vielschichtigkeit des Phänomens *Altruismus* und betonen die Notwendigkeit interdisziplinärer Ansätze in der Forschung unter Mitwirkung der Sozialen Neurowissenschaften.

Funding: Diese Arbeit wurde von der Deutschen Forschungsgemeinschaft unterstützt (DFG, HE 4566/5-1 an GH).

Literatur

Adolphs, R. (2009). [The social brain: neural basis of social knowledge](#). *Annu. Rev. Psychol.* 60, 693–716.

Batson, C. D. and Shaw, L. L. (1991). [Evidence for altruism: toward a pluralism of prosocial motives](#). *Psychol. Inq.* 2(2), 107–122.

Becker, S. W. and Eagly, A. H. (2004). [The heroism of women and men](#). *Am. Psychol.* 59, 163–78.

Bowles, S., Choi, J. K. and Hopfensitz, A. (2003). The co-evolution of individual behaviors and social institutions. *J. Theor. Biol.* 223, 135–47.

Buckholz, J. W., Asplund, C. L., Dux, P. E., Zald, D. H., Gore, J. C., Jones, O. D. and Marois, R. (2008). [The neural correlates of third-party punishment](#). *Neuron* 60, 930–40.

Buckholz, J. W., Martin, J. W., Treadway, M. T., Jan, K., Zald, D. H., Jones, O. and Marois, R. (2015). From Blame to Punishment: Disrupting Prefrontal Cortex Activity Reveals Norm Enforcement Mechanisms. *Neuron* 87, 1369–80.

Camerer, C. F. (2008). [Neuroeconomics: opening the gray box](#). *Neuron* 60, 416–9.

Craig, A. D. (2009). How do you feel – now? The anterior insula and human awareness. *Nat. Rev. Neurosci.* 10, 59–70.

Critchley, H. D. and Harrison, N. A. (2013). Visceral influences on brain and behavior. *Neuron* 77, 624–38.

Crockett, M. J., Ozdemir, Y. and Fehr, E. (2014). The value of vengeance and the demand for deterrence. *J. Exp. Psychol. Gen.* 143, 2279–86.

Daniels, R. A. (1981). Altruism in an antarctic fish. *Science* 213, 1281.

de Quervain, D. J., Fischbacher, U., Treyer, V., Schellhammer, M., Schnyder, U., Buck, A. and Fehr, E. (2004). The neural basis of altruistic punishment. *Science* 305, 1254–8.

Evers, K. and Sigman, M. (2013). [Possibilities and limits of mind-reading: a neurophilosophical perspective](#). *Conscious. Cogn.* 22, 887–97.

Fehr, E. and Camerer, C. F. (2007). Social neuroeconomics: the neural circuitry of social preferences. *Trends Cognit. Sci.* 11, 419–27.

Fehr, E. and Fischbacher, U. (2003). The nature of human altruism. *Nature* 425, 785–91.

Fehr, E. and Gächter, S. (2002). Altruistic punishment in humans. *Nature* 415, 137–40.

Feng, C., Luo, Y. J. and Krueger, F. (2015). [Neural signatures of fairness-related normative decision making in the ultimatum game: a coordinate-based meta-analysis](#). *Hum. Brain Mapp.* 36, 591–602.

Genevsky, A., Vastfjall, D., Slovic, P. and Knutson, B. (2013). [Neural underpinnings of the identifiable victim effect: affect shifts preferences for giving](#). *J. Neurosci.* 33, 17188–96.

Haas, B. W., Brook, M., Remillard, L., Ishak, A., Anderson, I. W. and Filkowski, M. M. (2015). I know how you feel: the warm-altruistic personality profile and the empathic brain. *PLoS One* 10, e0120639.

Harbaugh, W. T., Mayr, U. and Burghart, D. R. (2007). Neural responses to taxation and voluntary giving reveal motives for charitable donations. *Science* 316, 1622–5.

Hein, G., Lamm, C., Brodbeck, C. and Singer, T. (2011). Skin conductance response to the pain of others predicts later costly helping. *PLoS One* 6, e22759.

Hein, G., Morishima, Y., Leiberg, S., Sul, S. and Fehr, E. (2016). The brain's functional network architecture reveals human motives. *Science* 351, 1074–8.

Hein, G., Silani, G., Preuschoff, K., Batson, C. D. and Singer, T. (2010). Neural responses to ingroup and outgroup members' suffering predict individual differences in costly helping. *Neuron* 68, 149–60.

- Henrich, J., McElreath, R., Barr, A., Ensminger, J., Barrett, C., Bolyanatz, A., Cardenas, J. C., Gurven, M., Gwako, E., Henrich, N., Lesorogol, C., Marlowe, F., Tracer, D. and Ziker, J. (2006). Costly punishment across human societies. *Science* 312, 1767–70.
- Hu, T. Y., Li, J., Jia, H. and Xie, X. (2016a). Helping Others, Warming Yourself: Altruistic Behaviors Increase Warmth Feelings of the Ambient Environment. *Front. Psychol.* 7, 1349.
- Hu, Y., Scheele, D., Becker, B., Voos, G., David, B., Hurlmann, R. and Weber, B. (2016b). The Effect of Oxytocin on Third-Party Altruistic Decisions in Unfair Situations: An fMRI Study. *Sci. Rep.* 6, 20236.
- Hu, Y., Strang, S. and Weber, B. (2015). Helping or punishing strangers: neural correlates of altruistic decisions as third-party and of its relation to empathic concern. *Front. Behav. Neurosci.* 9, 24.
- Hutcherson, C. A., Bushong, B. and Rangel, A. (2015). A Neurocomputational Model of Altruistic Choice and Its Implications. *Neuron* 87, 451–62.
- Iredale, W. and Vugt, M. V. (2009). The peacock's tail of altruism. *The Psychologist* 22, 4.
- Knoch, D., Gianotti, L. R., Baumgartner, T. and Fehr, E. (2010). A neural marker of costly punishment behavior. *Psychol. Sci.* 21, 337–42.
- Knoch, D., Nitsche, M. A., Fischbacher, U., Eisenegger, C., Pascual-Leone, A. and Fehr, E. (2008). Studying the neurobiology of social interaction with transcranial direct current stimulation – the example of punishing unfairness. *Cereb. Cortex* 18, 1987–90.
- Kolling, N., Wittmann, M. K., Behrens, T. E., Boorman, E. D., Mars, R. B. and Rushworth, M. F. (2016). Value, search, persistence and model updating in anterior cingulate cortex. *Nat. Neurosci.* 19, 1280–5.
- Kuss, K., Falk, A., Trautner, P., Elger, C. E., Weber, B. and Fliessbach, K. (2013). A reward prediction error for charitable donations reveals outcome orientation of donors. *Soc. Cogn. Affect. Neurosci.* 8, 216–23.
- Laland, K. N., Sterelny, K., Odling-Smee, J., Hoppitt, W. and Uller, T. (2011). Cause and effect in biology revisited: is Mayr's proximate-ultimate dichotomy still useful? *Science* 334, 1512–6.
- Le, S. and Boyd, R. (2007). Evolutionary dynamics of the continuous iterated prisoner's dilemma. *J. Theor. Biol.* 245, 258–67.
- Lee, H. H., Molla, M. N., Cantor, C. R. and Collins, J. J. (2010). Bacterial charity work leads to population-wide resistance. *Nature* 467, 82–5.
- Lopez-Perez, B., Howells, L. and Gummerum, M. (2017). Cruel to Be Kind: Factors Underlying Altruistic Efforts to Worsen Another Person's Mood. *Psychol. Sci.* 28, 862–871.
- Milinski, M. (2016). Reputation, a universal currency for human social interactions. *Philos. Trans. R. Soc., B.* 371, 20150100.
- Morelli, S. A., Rameson, L. T. and Lieberman, M. D. (2014). The neural components of empathy: predicting daily prosocial behavior. *Soc. Cogn. Affect. Neurosci.* 9, 39–47.
- Morishima, Y., Schunk, D., Bruhin, A., Ruff, C. C. and Fehr, E. (2012). Linking brain structure and activation in temporoparietal junction to explain the neurobiology of human altruism. *Neuron* 75, 73–9.
- Murphy, G. P. and Dudley, S. A. (2009). Kin recognition: Competition and cooperation in *Impatiens* (Balsaminaceae). *Am. J. Bot.* 96, 1990–6.
- Murphy, R. O. and Ackermann, K. A. (2014). Social value orientation: theoretical and measurement issues in the study of social preferences. *Pers. Soc. Psychol. Rev.* 18, 13–41.
- Nowak, M. A. (2006). Five rules for the evolution of cooperation. *Science* 314, 1560–3.
- Nowak, M. A. (2012). Why we help: Far from being a nagging exception to the rule of evolution, cooperation has been one of its primary architects. *Sci. Am.* 307, 34–9.
- Nowak, M. A. and Sigmund, K. (1998). Evolution of indirect reciprocity by image scoring. *Nature* 393, 573–7.
- Peysakhovich, A., Nowak, M. A. and Rand, D. G. (2014). Humans display a 'cooperative phenotype' that is domain general and temporally stable. *Nat. Commun.* 5, 4939.
- Poldrack, R. A. (2011). Inferring mental states from neuroimaging data: from reverse inference to large-scale decoding. *Neuron* 72, 692–7.
- Rand, D. G., Kraft-Todd, G. and Gruber, J. (2015). The collective benefits of feeling good and letting go: positive emotion and (dis)inhibition interact to predict cooperative behavior. *PLoS One* 10, e0117426.
- Rilling, J. K. and Sanfey, A. G. (2011). The neuroscience of social decision-making. *Annu. Rev. Psychol.* 62, 23–48.
- Riolo, R. L., Cohen, M. D. and Axelrod, R. (2001). Evolution of cooperation without reciprocity. *Nature* 414, 441–3.
- Sanfey, A. G. (2007). Social decision-making: insights from game theory and neuroscience. *Science* 318, 598–602.
- Sanfey, A. G., Rilling, J. K., Aronson, J. A., Nystrom, L. E. and Cohen, J. D. (2003). The neural basis of economic decision-making in the Ultimatum Game. *Science* 300, 1755–8.
- Schultz, W. (2017). Reward prediction error. *Curr. Biol.* 27, R369–R371.
- Seip, E. C., van Dijk, W. W. and Rotteveel, M. (2009). On hotheads and Dirty Harries: the primacy of anger in altruistic punishment. *Ann. N. Y. Acad. Sci.* 1167, 190–6.
- Shackman, A. J., Salomons, T. V., Slagter, H. A., Fox, A. S., Winter, J. J. and Davidson, R. J. (2011). The integration of negative affect, pain and cognitive control in the cingulate cortex. *Nat. Rev. Neurosci.* 12, 154–67.
- Singer, T., Seymour, B., O'Doherty, J. P., Stephan, K. E., Dolan, R. J. and Frith, C. D. (2006). Empathic neural responses are modulated by the perceived fairness of others. *Nature* 439, 466–9.
- Strang, S., Gross, J., Schuhmann, T., Riedl, A., Weber, B. and Sack, A. T. (2015). Be nice if you have to – the neurobiological roots of strategic fairness. *Soc. Cogn. Affect. Neurosci.* 10, 790–6.
- Strobel, A., Zimmermann, J., Schmitz, A., Reuter, M., Lis, S., Windmann, S. and Kirsch, P. (2011). Beyond revenge: neural and genetic bases of altruistic punishment. *Neuroimage* 54, 671–80.
- Strombach, T., Weber, B., Hangebrauk, Z., Kenning, P., Karipidis, I., Tobler, P. N. and Kalenscher, T. (2015). Social discounting involves modulation of neural value signals by temporoparietal junction. *Proc. Natl. Acad. Sci. U.S.A.* 112, 1619–24.
- Tusche, A., Bockler, A., Kanske, P., Trautwein, F. M. and Singer, T. (2016). Decoding the Charitable Brain: Empathy, Perspective Taking, and Attention Shifts Differentially Predict Altruistic Giving. *J. Neurosci.* 36, 4719–32.

- Van Overwalle, F. (2009). Social cognition and the brain: a meta-analysis. *Hum. Brain Mapp.* 30, 829–58.
- Wilson, D. S. (1992). On the relationship between evolutionary and psychological definitions of altruism and selfishness. *Biol. Philos.* 7, 61–68.
- Yamagishi, T., Horita, Y., Mifune, N., Hashimoto, H., Li, Y., Shinada, M., Miura, A., Inukai, K., Takagishi, H. and Simunovic, D. (2012). Rejection of unfair offers in the ultimatum game is no evidence of strong reciprocity. *Proc. Natl. Acad. Sci. U.S.A.* 109, 20364–8.
- Zaki, J., Wager, T. D., Singer, T., Keysers, C. and Gazzola, V. (2016). The Anatomy of Suffering: Understanding the Relationship between Nociceptive and Empathic Pain. *Trends Cognit. Sci.* 20, 249–59.

Anmerkung: Übersetzung der englischen Version des Artikels online verfügbar unter <https://doi.org/10.1515/nf-2017-A047>

Autoreninformationen



Prof. Dr. Sabine Windmann
Psychologisches Institut der Goethe-Universität Frankfurt am Main, Theodor-W.-Adorno Platz 6, 60323 Frankfurt am Main
Tel.: +49 69 79835313
E-Mail: s.windmann@psych.uni-frankfurt.de

Sabine Windmann ist seit 2006 Professorin für Allgemeine Psychologie II an der Johann Wolfgang Goethe-Universität Frankfurt am Main. Sie studierte Psychologie in Gießen und Bonn, promovierte an der Universität Trier und habilitierte sich an der Ruhr-Universität Bochum. Zwei Jahre ihrer Post-Doc Forschungstätigkeit absolvierte sie an der University of California San Diego und zwei weitere an der University of Plymouth (UK). Ihre Forschungsinteressen beziehen sich auf die Auflösung von Unsicherheit, Ambiguität und Konflikt, zuletzt vor allem in sozialen Kontexten. Sie lebt mit ihrem Mann und ihren beiden Kindern in Weinheim an der Bergstraße.



Prof. Dr. Grit Hein
Institut für Medizinische Psychologie,
Johann Wolfgang Goethe Universität, Heinrich Hoffmann Straße 10, 60528 Frankfurt/Main;
Abteilung für Translationale Soziale Neurowissenschaften, Translational Social Neuroscience Unit, Julius-Maximilians-Universität Würzburg, Klinik für Psychiatrie und Psychosomatik, Margarete-Höppel-Platz 1, 97080 Würzburg
E-Mail: Hein_G@ukw.de

Grit Hein ist Psychologin, Neurowissenschaftlerin, und Professorin für Translationale Soziale Neurowissenschaften an der Universität Würzburg. Nach ihrem Psychologiestudium an der Humboldt-Universität zu Berlin und dem City College of New York promovierte sie am Max-Planck-Institut für Kognitions- und Neurowissenschaften in Leipzig. Nach Forschungsaufenthalten in Cambridge, England, Frankfurt am Main und Berkeley, USA, war sie als Wissenschaftlerin an der Universität Zürich und als Dozentin an der Universität Bern tätig. Schwerpunkte ihrer Forschung sind neuronale Grundlagen sozialer Motive, und ihre Auswirkung auf Verhalten, das Zusammenspiel zwischen sozialer Motivation und Lernen, und motivationale Defizite bei Patienten mit psychischen Störungen. Grit Hein ist verheiratet und Mutter von zwei Söhnen.

