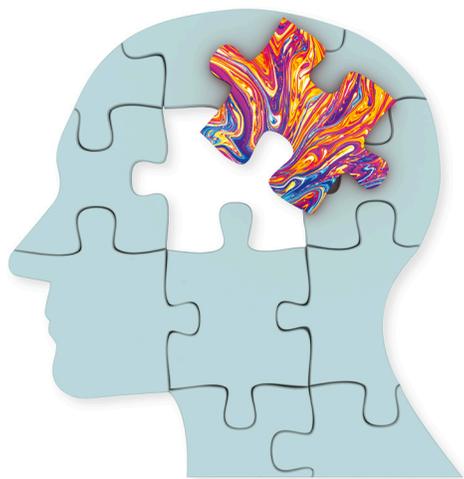


A THOUSAND BRAINS



EINE NEUE THEORIE
DER INTELLIGENZ

Mit einem Vorwort von Richard Dawkins

riva

JEFF HAWKINS

JEFF HAWKINS

A

**THOUSAND
BRAINS**

© 2022 der Titel »A Thousand Brains« von Jeff Hawkins (ISBN 978-3-7432-2112-8) by riva Verlag,
Münchner Verlagsgruppe GmbH, München. Nähere Informationen unter: www.rv-vg.de

JEFF HAWKINS

**A
THOUSAND
BRAINS**

**INE NEUE THEORIE
DER INTELLIGENZ**

© 2022 der Titel »A Thousand Brains« von Jeff Hawkins (ISBN 978-3-7433-2112-8) by riva Verlag,
Münchner Verlagsgruppe GmbH, München. Nähere Informationen unter: www.rv-vg.de

riva

Bibliografische Information der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie. Detaillierte bibliografische Daten sind im Internet über <http://dnb.d-nb.de> abrufbar.

Für Fragen und Anregungen

info@rivaverlag.de

Wichtiger Hinweis

Ausschließlich zum Zweck der besseren Lesbarkeit wurde auf eine genderspezifische Schreibweise sowie eine Mehrfachbezeichnung verzichtet. Alle personenbezogenen Bezeichnungen sind somit geschlechtsneutral zu verstehen.

1. Auflage 2022

© 2022 by riva Verlag, ein Imprint der Münchner Verlagsgruppe GmbH

Türkenstraße 89

80799 München

Tel.: 089 651285-0

Fax: 089 652096

Die amerikanische Originalausgabe erschien 2021 bei Basic Books unter dem Titel *A Thousand Brains: A New Theory of Intelligence*. © 2021 by Jeffrey C. Hawkins. All rights reserved.

Alle Rechte, insbesondere das Recht der Vervielfältigung und Verbreitung sowie der Übersetzung, vorbehalten. Kein Teil des Werkes darf in irgendeiner Form (durch Fotokopie, Mikrofilm oder ein anderes Verfahren) ohne schriftliche Genehmigung des Verlages reproduziert oder unter Verwendung elektronischer Systeme gespeichert, verarbeitet, vervielfältigt oder verbreitet werden.

Übersetzung: Simone Fischer

Redaktion: Matthias Michel

Umschlaggestaltung: Karina Braun

Umschlagabbildung: ©iStockphoto.com/alexsl; Shutterstock.com/Dmitri Gomon

Abbildungen im Innenteil: S.12: [stock.adobe.com/Bill Fehr](https://stock.adobe.com/Bill_Fehr); S. 14: adaptiert von »Distributed Hierarchical Processing in the Primate Cerebral Cortex« von D. J. Felleman und D. C. Van Essen, 1991, *Cerebral Cortex*, 1(1):1.; S. 16: S. R. y Cajal; S. 109: E. H. Adelson; S. 177: B. Derksen; S. 177: nachgedruckt unter den Bedingungen der GNU-Lizenz für freie Dokumentation

Satz: Satzwerk Huber, Germering

Druck: GGP Media GmbH, Pößneck

Printed in Printed in Germany

ISBN Print 978-3-7423-2212-8

ISBN E-Book (PDF) 978-3-7453-1974-3

ISBN E-Book (EPUB, Mobi) 978-3-7453-1975-0



Weitere Informationen zum Verlag finden Sie unter

www.rivaverlag.de

Beachten Sie auch unsere weiteren Verlage unter www.m-vg.de

INHALT

Vorwort von Richard Dawkins	7
TEIL 1: EIN NEUES VERSTÄNDNIS DES GEHIRNS	15
Kapitel 1: Altes Gehirn – neues Gehirn	29
Kapitel 2: Vernon Mountcastles große Idee	41
Kapitel 3: Ein Modell der Welt im Kopf	51
Kapitel 4: Das Gehirn offenbart seine Geheimnisse	63
Kapitel 5: Karten im Gehirn	83
Kapitel 6: Konzepte, Sprache und komplexes Denken	97
Kapitel 7: Die Thousand-Brains-Theorie der Intelligenz	121
TEIL 2: KÜNSTLICHE INTELLIGENZ	145
Kapitel 8: Warum es kein »Ich« in der KI gibt	153
Kapitel 9: Maschinen mit Bewusstsein	171
Kapitel 10: Die Zukunft der künstlichen Intelligenz	183
Kapitel 11: Die existenziellen Risiken der künstlichen Intelligenz	201
TEIL 3: MENSCHLICHE INTELLIGENZ	213
Kapitel 12: Falsche Überzeugungen	217
Kapitel 13: Die existenziellen Risiken der menschlichen Intelligenz	231
Kapitel 14: Die Verschmelzung von Gehirnen und Maschinen	247
Kapitel 15: Nachlassplanung für die Menschheit	259
Kapitel 16: Gene vs. Wissen	275

Abschließende Gedanken	297
Literaturempfehlungen	303
Danksagung	3II
Über den Autor	3I5
Register	3I6

VORWORT

VON RICHARD DAWKINS

Lesen Sie dieses Buch nicht vor dem Schlafengehen. Nicht, dass es beängstigend wäre. Es wird Ihnen keine Alpträume bereiten. Aber es ist so erfrischend, so anregend, dass es Ihren Geist in einen wirbelnden Strudel von aufregend provokativen Ideen verwandelt – Sie werden lieber losstürmen und jemandem davon erzählen wollen, als schlafen zu gehen. Der Verfasser des Vorworts ist ein Opfer dieses Strudels, und ich denke, das wird sich erweisen.

Charles Darwin war unter den Wissenschaftlern insofern eine Ausnahmeerscheinung, als er über die Mittel verfügte, außerhalb der Universitäten und ohne staatliche Forschungsgelder zu arbeiten. Jeff Hawkins wird es vielleicht nicht gefallen, als das Silicon-Valley-Äquivalent eines Gentleman-Wissenschaftlers bezeichnet zu werden, aber – Sie verstehen sicher die Parallele. Darwins mächtige Idee war zu revolutionär, um in Form eines kurzen Artikels verstanden zu werden, und die ge-

meinsamen Veröffentlichungen von Darwin und Wallace aus dem Jahr 1858 wurden so gut wie ignoriert. Wie Darwin selbst sagte, musste die Idee in Buchlänge ausgebreitet werden. Und tatsächlich war es sein großes Buch, das ein Jahr später die viktorianischen Grundüberzeugungen erschütterte. Auch Jeff Hawkins' Thousand-Brains-Theorie bedarf einer buchfüllenden Abhandlung. Und sein Konzept der Bezugsrahmen – »Der eigentliche Akt des Denkens ist eine Form der Bewegung« – ist eine Glanzleistung! Diese beiden Ideen sind jeweils tiefgründig genug, um ein ganzes Buch zu füllen. Aber das ist noch nicht alles.

Thomas Henry Huxley sagte nach der Lektüre von Darwins *Über die Entstehung der Arten* bekanntlich: »Wie ausgesprochen dumm von mir, nicht daran gedacht zu haben.« Ich behaupte natürlich nicht, dass Hirnforscher dasselbe sagen werden, wenn sie dieses Buch zuklappen. Es ist ein Buch mit vielen spannenden Ideen statt einer einzigen großen Idee wie der von Darwin.

Ich glaube allerdings, dass nicht nur T. H. Huxley, sondern auch seine drei brillanten Enkel es geliebt hätten: Andrew, weil er entdeckte, wie der Nervenimpuls funktioniert (Hodgkin und Huxley sind der Watson und der Crick des Nervensystems); Aldous wegen seiner visionären und poetischen Reisen in die entlegensten Winkel des Geistes; und Julian, weil er dieses Gedicht schrieb, in dem er die Fähigkeit des Gehirns preist, ein Modell der Realität, einen Mikrokosmos des Universums zu konstruieren:

*Die Welt der Dinge kam in deinen kindlichen Verstand,
um das Kristallkabinett zu bevölkern.
Innerhalb seiner Wände trafen sich die seltsamsten Partner,
und aus Dingen wurden Gedanken, die ihre Art fortpflanzten.*

*Denn, einmal drinnen, fand die körperliche Realität einen Geist.
Die Realität und du in gegenseitiger Schuld
habt dort euren kleinen Mikrokosmos erschaffen,
der doch seinem kleinen Selbst die größten Aufgaben zugewiesen hatte.*

*Tote können dort leben und sich mit den Sternen unterhalten:
Der Äquator spricht mit dem Pol, die Nacht mit dem Tag;*

*Der Geist löst die materiellen Schranken der Welt auf.
Eine Million Abgrenzungen verglühen.
Das Universum kann leben und arbeiten und planen.
Endlich wurde Gott im Geist des Menschen geschaffen.*

Das Gehirn sitzt im Dunkeln und nimmt die Außenwelt nur durch einen Hagelsturm von Andrew Huxleys Nervenimpulsen wahr. Ein Nervenimpuls aus dem Auge unterscheidet sich nicht von einem aus dem Ohr oder dem großen Zeh. Es kommt darauf an, wo sie im Gehirn landen. Jeff Hawkins ist nicht der erste Wissenschaftler oder Philosoph, der behauptet, dass die von uns wahrgenommene Realität eine konstruierte Realität ist, ein Modell, das durch die von den Sinnen einströmenden Informationen aktualisiert und informiert wird. Aber Hawkins ist meines Erachtens der Erste, der der Idee wortgewaltig Raum bietet, dass es nicht nur ein einziges solches Modell gibt, sondern Tausende, und zwar eines in jeder der vielen fein säuberlich gestapelten Säulen, die den Kortex des Gehirns bilden. Es gibt etwa 150.000 dieser Säulen, und sie sind die Stars des ersten Teils des Buches, zusammen mit dem, was er »Bezugsrahmen« nennt. Hawkins' Thesen zu diesen beiden Themen sind provokant, und es wird interessant sein zu sehen, wie sie von anderen Hirnforschern aufgenommen werden; gut, vermute ich. Nicht zuletzt fasziniert seine Idee, dass die kortikalen Säulen bei ihrer Weltmodellierung halbautonom arbeiten. Was »wir« wahrnehmen, ist eine Art demokratischer Konsens zwischen den Säulen.

Demokratie im Gehirn? Konsens und sogar Streit? Was für eine erstaunliche Idee. Sie ist ein Hauptthema des Buches. Wir menschlichen Säugetiere sind die Opfer eines immer wiederkehrenden Streits: einem Gerangel zwischen dem alten Reptiliengehirn, das unbewusst die Überlebensmaschine steuert, und dem Neokortex der Säugetiere, der sozusagen den Fahrersitz besetzt. Dieses neue Säugetiergehirn – die Großhirnrinde – denkt. Die Großhirnrinde ist der Sitz des Bewusstseins. Sie ist sich der Vergangenheit, der Gegenwart und der Zukunft bewusst, und sie sendet Anweisungen an das alte Gehirn, das diese ausführt.

Das alte Gehirn, das durch natürliche Selektion über Millionen von Jahren trainiert wurde, als Zucker knapp und wertvoll für das Überleben war, sagt: »Kuchen. Ich will Kuchen. Mmmm, Kuchen. Gib her.« Das neue Gehirn, das von Büchern und Ärzten in nur zehn Jahren trainiert wurde, als Zucker im Überfluss vorhanden war, sagt: »Nein, nein. Kein Kuchen. Das darf nicht sein. Bitte, iss diesen Kuchen nicht.« Das alte Gehirn entgegnet: »Schmerz, Schmerz, schrecklicher Schmerz, lass den Schmerz *sofort* aufhören«. Das neue Gehirn fordert: »Nein, nein, ertrage die Folter, verrate dein Land nicht, indem du dich ihm hingibst. Loyalität gegenüber dem Land und den Kameraden kommt sogar vor deinem eigenen Leben.«

Der Konflikt zwischen dem alten Reptilien- und dem neuen Säugertiergehirn liefert die Antwort auf solche Rätsel wie »Warum muss Schmerz so verdammt schmerzhaft sein?«. Wozu ist Schmerz überhaupt gut? Schmerz ist ein Stellvertreter für den Tod. Er ist eine Warnung an das Gehirn: »Tu das nicht noch einmal: ärgere keine Schlange, hebe keine heiße Glut auf, springe nicht aus großer Höhe. Diesmal hat es nur wehgetan; beim nächsten Mal könnte es dich umbringen.« Nun könnte ein Konstrukteur sagen, was wir hier brauchen, ist das Äquivalent einer schmerzfreien Flagge im Gehirn. Wenn die Flagge in die Höhe schießt, darfst du nicht wiederholen, was du gerade getan hast. Doch statt der einfachen und schmerzfreien Flagge des Ingenieurs bekommen wir in Wirklichkeit Schmerzen – oft quälende, unerträgliche Schmerzen. Und warum? Was ist falsch an dieser so sinnvollen Flagge?

Die Antwort liegt wahrscheinlich in der umstrittenen Natur der Entscheidungsprozesse des Gehirns: dem Ringen zwischen dem alten und dem neuen Gehirn. Da es für das neue Gehirn zu einfach ist, das Votum des alten Gehirns zu überstimmen, würde das schmerzfreie Flaggensystem nicht funktionieren. Genauso wenig wie Folter.

Das neue Gehirn würde sich dazu berechtigt fühlen, meine hypothetische Flagge zu ignorieren und jede Menge Bienenstiche, verstauchte Knöchel oder Daumenschrauben der Folterknechte zu ertragen, wenn es das aus irgendeinem Grund »wollte«. Das alte Gehirn, dem es wirklich wichtig ist zu überleben, um die Gene weiterzugeben, könnte vergeblich »protestieren«. Vielleicht hat die natürliche Auslese im Inter-

esse des Überlebens den »Sieg« des alten Gehirns sichergestellt, indem sie den Schmerz so verdammt schmerzhaft gemacht hat, dass das neue Gehirn ihn nicht überwinden kann. Ein anderes Beispiel: Wäre sich das alte Gehirn des Verrats am darwinistischen Ziel des Geschlechtsverkehrs »bewusst«, wäre das Überstreifen eines Kondoms unerträglich schmerzhaft.

Hawkins gehört zur Mehrheit der informierten Wissenschaftler und Philosophen, die mit dem Dualismus nichts anfangen können: Es gibt keinen Geist in der Maschine, keine gespenstische Seele, die so weit von der Hardware entfernt ist, dass sie den Tod der Hardware überlebt, kein »kartesisches Theater« (Dan Dennetts Begriff), in dem ein Farbbildschirm einem beobachtenden Ich einen Film von der Welt zeigt. Stattdessen schlägt Hawkins mehrere Modelle der Welt vor, konstruierte Mikrokosmen, die durch den Regen von Nervenimpulsen, die von den Sinnen einströmen, informiert und angepasst werden. Übrigens schließt Hawkins nicht völlig aus, dass man langfristig dem Tod entgehen kann, indem man sein Gehirn in einen Computer hochlädt, aber er glaubt nicht, dass das viel Spaß machen würde.

Zu den wichtigsten Modellen des Gehirns gehören die Modelle des Körpers selbst, die sich damit auseinandersetzen müssen, wie die Eigenbewegung des Körpers unsere Perspektive auf die Welt außerhalb der Gefängnismauer des Schädels verändert. Und dies ist relevant für das Hauptthema des mittleren Teils des Buches, nämlich der Intelligenz von Maschinen. Jeff Hawkins hat, wie ich auch, großen Respekt vor jenen klugen Menschen, Freunden von ihm und mir, die das Herannahen superintelligenter Maschinen fürchten, die uns verdrängen, unterjochen oder sogar ganz beseitigen könnten. Aber Hawkins fürchtet sie nicht, auch weil die Fähigkeiten, die zur Beherrschung von Schach oder Go führen, nicht die sind, die mit der Komplexität der realen Welt fertig werden können. Kinder, die kein Schach spielen können, wissen, »wie Flüssigkeiten verschüttet werden, Bälle rollen und Hunde bellen. Sie wissen, wie man Stifte, Marker, Papier und Klebstoff benutzt. Sie wissen, wie man Bücher aufschlägt und dass Papier reißen kann«. Und sie haben ein Selbstbild, ein Körperbild, das sie in der Welt der physischen Realität verortet und es ihnen ermöglicht, sich mühelos in ihr zurechtzufinden.

Es ist nicht so, dass Hawkins die Macht der künstlichen Intelligenz und der Roboter der Zukunft unterschätzt. Ganz im Gegenteil. Aber er ist der Meinung, dass die meisten aktuellen Forschungen in die falsche Richtung gehen. Der richtige Weg besteht seiner Meinung nach darin zu verstehen, wie das Gehirn funktioniert, und seine Methoden zu übernehmen, sie aber enorm zu beschleunigen.

Und es gibt keinen Grund (wirklich nicht), die Verhaltensweisen des alten Gehirns zu übernehmen, seine Begierden und seinen Hunger, sein Verlangen und seine Wut, seine Gefühle und Ängste, die uns auf Wege treiben können, die vom neuen Gehirn als schädlich angesehen werden. Schädlich zumindest aus der Perspektive, die Hawkins und ich, und wahrscheinlich auch Sie, schätzen. Denn er ist sich darüber im Klaren, dass unsere aufgeklärten Werte stark von dem primären und primitiven Wert unserer egoistischen Gene abweichen müssen und dies auch tun – dem rohen Imperativ, sich um jeden Preis fortzupflanzen. Ohne ein altes Gehirn gibt es nach seiner Ansicht (die, wie ich vermute, umstritten ist) keinen Grund zu erwarten, dass eine KI uns gegenüber böswillige Gefühle hegt. Aus demselben Grund, und vielleicht ist auch dies umstritten, hält er das Ausschalten einer bewussten KI nicht für einen Mord: Warum sollte sie ohne ein altes Gehirn Angst oder Traurigkeit empfinden? Warum sollte sie überleben wollen?

In Kapitel 16, »Gene vs. Wissen«, werden wir über die Diskrepanz zwischen den Zielen des alten Gehirns (egoistischen Genen zu dienen) und des neuen Gehirns (Wissen) ausführlich aufgeklärt. Es ist der Ruhm der menschlichen Großhirnrinde, dass sie – einzigartig unter allen Tieren und beispiellos in allen geologischen Zeiten – die Macht hat, sich dem Diktat der egoistischen Gene zu widersetzen. Wir können Sex genießen, ohne uns fortzupflanzen. Wir können unser Leben der Philosophie, der Mathematik, der Poesie, der Astrophysik, der Musik, der Geologie oder der Wärme der menschlichen Liebe widmen, entgegen dem genetischen Drängen des alten Gehirns, dass dies Zeitverschwendung sei – Zeit, die stattdessen mit dem Kampf gegen Rivalen und der Jagd nach mehreren Sexualpartnern verbracht werden »sollte«: »Meiner Meinung nach müssen wir eine tiefgreifende Entscheidung treffen. Wir müssen uns entscheiden, ob wir das alte Gehirn oder das neue Gehirn bevorzugen.

Genauer gesagt: Wollen wir, dass unsere Zukunft von den Prozessen bestimmt wird, die uns hierhergebracht haben, nämlich von natürlicher Auslese, Wettbewerb und dem Antrieb egoistischer Gene? Oder wollen wir, dass unsere Zukunft von Intelligenz und dem Wunsch, die Welt zu verstehen, bestimmt wird?»

Ich begann mit einem Zitat von T. H. Huxley, einer liebenswert bescheidenen Bemerkung zur Lektüre von Darwins *Über die Entstehung der Arten*. Ich schließe mit einer der vielen faszinierenden Ideen von Jeff Hawkins – er fasst sie auf nur wenigen Seiten zusammen –, die mich an Huxley denken ließen. Hawkins verspürt das Bedürfnis nach einem kosmischen Grabstein, etwas, das die Galaxie wissen lässt, dass wir einmal hier waren. Und dieser kosmische Grabstein muss in der Lage sein, diese Tatsache zu verkünden. Hawkins stellt fest, dass alle Zivilisationen vergänglich sind. Auf der Skala der universellen Zeit ist das Intervall zwischen der Erfindung der elektromagnetischen Kommunikation durch eine Zivilisation und ihrem Aussterben wie das Aufleuchten eines Glühwürmchens. Die Chance, dass ein solches Aufleuchten mit einem anderen zusammenfällt, ist unglücklicherweise gering. Was wir also brauchen – weshalb ich ihn Grabstein genannt habe –, ist eine Botschaft, die nicht sagt: »Wir sind hier«, sondern »Wir waren einmal hier«. Und der Grabstein muss von kosmischer Dauer sein: Er muss nicht nur aus Parallaxensekunden sichtbar sein, sondern auch Millionen, wenn nicht Milliarden von Jahren überdauern, sodass er seine Botschaft noch verkündet, wenn andere Geistesblitze ihn lange nach unserem Aussterben abfangen. Es reicht nicht aus, Primzahlen oder die Ziffern der Kreiszahl π (Pi) zu senden. Jedenfalls nicht als Funksignal oder als gepulster Laserstrahl. Sie künden zwar von biologischer Intelligenz, weshalb sie das Hauptthema von SETI (Search for Extraterrestrial Intelligence, deutsch: Suche nach außerirdischer Intelligenz) und Science-Fiction sind, aber sie sind zu kurz und zu gegenwärtig. Welches Signal würde also lange genug anhalten und aus einer sehr großen Entfernung in jeder Richtung wahrnehmbar sein? An dieser Stelle hat Hawkins meinen inneren Huxley geweckt.

Heute können wir das nicht, aber in der Zukunft, bevor unser Glühwürmchen-Aufleuchten verbraucht ist, könnten wir eine Reihe von Sa-

telliten in eine Umlaufbahn um die Sonne bringen, »die einen Teil des Sonnenlichts in einem Muster blockieren, das in der Natur nicht vorkommt. Diese kreisenden Sonnenblocker würden die Sonne noch Millionen von Jahren umkreisen, lange nachdem wir nicht mehr da sind, und sie könnten aus großer Entfernung entdeckt werden«. Selbst wenn die Abstände dieser schattenwerfenden Satelliten nicht buchstäblich eine Reihe von Primzahlen sind, könnte die Botschaft unmissverständlich gemacht werden: »Hier gab es intelligentes Leben.«

Was ich sehr erfreulich finde – und ich biete Jeff Hawkins diese Zeilen an, um ihm für das Vergnügen zu danken, das mir sein brillantes Buch bereitet hat –, ist, dass eine kosmische Botschaft, die in Form eines Musters von Intervallen zwischen Spikes (oder in seinem Fall Anti-Spikes, da seine Satelliten die Sonne abdunkeln) kodiert ist, dieselbe Art von Code wie ein Neuron verwenden würde.

Dies ist ein Buch darüber, wie das Gehirn funktioniert. Das Gehirn funktioniert auf eine Art und Weise, die geradezu berauschend ist.

TEIL 1

Ein neues Verständnis des Gehirns

© 2022 der Titel: «A Thousand Brains» von Jeff Hawkins (ISBN 978-3-7433-2112-8) by riva Verlag,
Münchner Verlagsgruppe GmbH, München. Nähere Informationen unter: www.rv-vg.de

Die Zellen in Ihrem Kopf lesen diese Wörter. Stellen Sie sich bitte einmal vor, wie bemerkenswert das ist. Zellen sind einfach gestrickt. Eine einzelne Zelle kann weder lesen noch denken noch irgendetwas anderes tun. Wenn wir jedoch genügend Zellen zusammensetzen, um daraus ein Gehirn zu bilden, lesen diese Zellen nicht nur Bücher, sie schreiben sie auch. Sie entwerfen Gebäude, erfinden Technologien und entschlüsseln die Geheimnisse des Universums. Doch wie ein Gehirn, das aus einfachen Zellen besteht, Intelligenz hervorbringt, ist eine hochinteressante Frage und nach wie vor ein Rätsel.

Die Erforschung der Funktionsweise des Gehirns ist eine der großen Herausforderungen der Menschheit. Diese Aufgabe hat eine Vielzahl nationaler und internationaler Initiativen ins Leben gerufen, wie beispielsweise das Human Brain Project der Europäischen Kommission und die International Brain Initiative. Zehntausende von Neurowissenschaftlern aus Dutzenden von Fachgebieten arbeiten in praktisch jedem Land der Welt daran, das Gehirn zu verstehen. Obwohl Neurowissenschaftler die Gehirne unterschiedlicher Tierarten untersuchen und sich mit den verschiedensten Fragestellungen beschäftigen, besteht das ultimative Ziel der Neurowissenschaft darin herauszufinden, wie das menschliche Gehirn die menschliche Intelligenz hervorbringt.

Meine Behauptung, dass das menschliche Gehirn immer noch ein Rätsel ist, überrascht Sie vielleicht. Denn Jahr für Jahr werden neue Entdeckungen im Zusammenhang mit der Hirnforschung bekanntgegeben und unzählige neue Bücher über das Gehirn veröffentlicht. Zudem be-

haupten Forscher auf verwandten Gebieten wie der künstlichen Intelligenz, dass ihre Schöpfungen bereits der Intelligenz einer Maus oder einer Katze nahekommen. Daraus könnte man leicht schließen, dass die Wissenschaftler eine ziemlich gute Vorstellung davon haben, wie das Gehirn funktioniert. Fragt man jedoch Neurowissenschaftler, ob das tatsächlich so ist, so werden fast alle zugeben, dass wir noch im Dunkeln tappen. Wir haben eine enorme Menge an Wissen und Fakten über das Gehirn zusammengetragen, aber wir wissen nur wenig darüber, wie das Ganze funktioniert.

1979 schrieb Francis Crick, berühmt für seine Arbeit an der DNA, einen Artikel über den Stand der Gehirnforschung mit dem Titel »Thinking about the Brain« (»Gedanken über das Gehirn«). Er beschrieb die umfangreiche Menge an Fakten, die Wissenschaftler über das Gehirn gesammelt hatten, und kam zu dem Schluss, dass »trotz der kontinuierlichen Anhäufung von Detailwissen die Funktionsweise des menschlichen Gehirns immer noch zutiefst rätselhaft ist«. Er fuhr fort: »Was ganz offensichtlich fehlt, ist ein breiter gedanklicher Rahmen, in dem diese Ergebnisse interpretiert werden können.«

Crick stellte fest, dass Wissenschaftler zwar seit Jahrzehnten Daten über das Gehirn gesammelt und viele Erkenntnisse gewonnen hatten, dass jedoch niemand in der Lage war, diese Fakten zu etwas Sinnvollem zusammzusetzen. Das Gehirn war wie ein riesiges Puzzlespiel mit Tausenden von Teilen. Die Puzzlestücke lagen vor uns, doch wir konnten sie nicht sinnvoll zusammensetzen, weil niemand wusste, wie die Lösung aussehen sollte. Crick zufolge war das Gehirn nicht deswegen ein Rätsel, weil wir nicht genug Daten gesammelt hatten, sondern weil wir nicht wussten, wie wir die vorhandenen Puzzleteile anordnen sollten. In den mehr als vierzig Jahren, die seit Cricks Artikel vergangen sind, hat es viele bedeutende Entdeckungen über das Gehirn gegeben, auf die ich später noch eingehen werde, aber im Großen und Ganzen ist seine Aussage immer noch richtig. Wie aus den Zellen in unserem Kopf Intelligenz entsteht, ist immer noch ein großes Rätsel. Da jedes Jahr weitere Puzzleteile gesammelt werden, wirkt es manchmal so, als kämen wir dem Verständnis des Gehirns nicht näher, sondern entfernten uns immer weiter davon.

Als ich Cricks Artikel als junger Mann las, inspirierte mich dies. Ich hatte das Gefühl, dass wir das Rätsel des Gehirns noch zu meinen Lebzeiten lösen könnten, und seitdem verfolge ich dieses Ziel. In den letzten fünfzehn Jahren habe ich ein Forschungsteam im Silicon Valley geleitet, das einen Teil des Gehirns, den Neokortex, untersucht. Der Neokortex nimmt etwa 70 Prozent des Volumens des menschlichen Gehirns ein und ist für alles verantwortlich, was wir mit Intelligenz in Verbindung bringen, von unseren Seh-, Tast- und Hörsinnen über Sprache in all ihren Formen bis hin zu abstraktem Denken, wie Mathematik und Philosophie. Ziel unserer Forschung ist es, die Funktionsweise des Neokortex so genau zu verstehen, dass wir die Biologie des Gehirns erklären und intelligente Maschinen bauen können, die nach denselben Prinzipien arbeiten.

Anfang 2016 änderte sich der Fortschritt unserer Forschung dramatisch, da wir einen Durchbruch in unserem Verständnis erzielten. Wir erkannten, dass sowohl wir als auch andere Wissenschaftler einen wichtigen Bestandteil übersehen hatten. Dank dieser neuen Erkenntnis verstanden wir endlich, wie die Puzzleteile zusammenpassen. Mit anderen Worten: Ich glaube, dass wir den Rahmen entdeckt haben, über den Crick geschrieben hat, einen Rahmen, der nicht nur die Grundlagen der Funktionsweise des Neokortex erklärt, sondern auch eine neue Art, über Intelligenz nachzudenken, hervorbringt. Wir haben noch keine vollständige Theorie des menschlichen Gehirns entwickelt – davon sind wir weit entfernt. Auf dem Feld der Wissenschaft beginnt man in der Regel mit einem theoretischen Rahmen, und erst später werden die Details ausgearbeitet. Das vielleicht berühmteste Beispiel ist Darwins Evolutionstheorie. Darwin schlug eine kühne neue Denkweise über den Ursprung der Arten vor, aber die Details, wie zum Beispiel die Funktionsweise der Gene und der DNA, wurden erst viele Jahre später bekannt.

Um intelligent zu werden, muss das Gehirn sehr viele Dinge über die Welt lernen. Dies beinhaltet nicht nur das, was wir in der Schule lernen, sondern auch grundlegende Dinge, beispielsweise wie Alltagsgegenstände aussehen, sich anhören und anfühlen. Wir müssen begreifen, wie sich Gegenstände verhalten, wie sich also Türen öffnen und schließen oder was die Apps in unseren Smartphones tun, wenn wir den

Bildschirm berühren. Außerdem müssen wir lernen, wo sich alles auf der Welt befindet, vom Aufbewahrungsort unserer persönlichen Gegenstände in der Wohnung bis hin zur Lage der Bibliothek und des Postamts in unserer Stadt. Und natürlich erlernen wir auch übergeordnete Konzepte, wie die Bedeutung von »Mitgefühl« oder von »Regierung«. Darüber hinaus lernt jeder von uns die Bedeutung von Zehntausenden von Wörtern. Jeder einzelne Mensch verfügt über eine enorme Menge an Wissen über die Welt. Dabei sind einige unserer grundlegenden Fähigkeiten durch unsere Gene festgelegt, etwa wie wir essen oder wie wir vor Schmerzen zurückschrecken. Aber das meiste, was wir über die Welt wissen, haben wir erlernt.

Nach Ansicht einiger Wissenschaftler erlernt das Gehirn ein Modell der Welt. Das Wort »Modell« impliziert, dass das, was wir wissen, nicht nur als eine Ansammlung von Fakten gespeichert wird, sondern auf eine Weise organisiert ist, die die Struktur der Welt und alles, was diese enthält, widerspiegelt. Um zum Beispiel zu wissen, was ein Fahrrad ist, erinnern wir uns nicht an eine Reihe von Fakten über Fahrräder. Stattdessen erstellt unser Gehirn ein Modell von Fahrrädern, das die verschiedenen Bestandteile, die Anordnung dieser Teile zueinander und die Bewegungsabläufe und das Zusammenspiel der verschiedenen Teile umfasst. Um etwas zu erkennen, müssen wir zunächst lernen, wie es aussieht und sich anfühlt. Und um Ziele zu erreichen, müssen wir lernen, wie sich die Dinge in der Welt normalerweise verhalten, wenn wir mit ihnen interagieren. Intelligenz ist eng mit dem Weltmodell des Gehirns verbunden. Wenn wir verstehen wollen, wie das Gehirn Intelligenz hervorbringt, müssen wir herausfinden, wie das Gehirn, das aus einfachen Zellen besteht, ein Modell der Welt und aller Dinge darin erlernt.

Unsere Entdeckung von 2016 erklärt, wie das Gehirn dieses Modell erlernt. Wir kamen zu dem Schluss, dass der Neokortex alles, was wir wissen, also unser gesamtes Wissen, in sogenannten Bezugsrahmen speichert. Ich werde dies später ausführlicher erklären, möchte es aber für den Moment mit dem Beispiel einer Landkarte veranschaulichen. Eine Landkarte oder auch ein Stadtplan ist eine Art Modell. Stellen Sie sich dies so vor: Ein Stadtplan ist ein Modell der Stadt, und die Gitternetzlinien, wie Breiten- und Längengrade, sind eine Art Bezugsrahmen. Die