

Patrick Krauss

# Künstliche Intelligenz und Hirnforschung

Neuronale Netze, Deep Learning  
und die Zukunft  
der Kognition



SACHBUCH



Springer

# Künstliche Intelligenz und Hirnforschung

Patrick Krauss

# Künstliche Intelligenz und Hirnforschung

Neuronale Netze, Deep Learning  
und die Zukunft der Kognition

 Springer

Patrick Krauss  
Universitätsklinikum Erlangen  
Friedrich-Alexander-Universität  
Erlangen-Nürnberg  
Erlangen, Deutschland

ISBN 978-3-662-67178-8      ISBN 978-3-662-67179-5 (eBook)  
<https://doi.org/10.1007/978-3-662-67179-5>

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

© Der/die Herausgeber bzw. der/die Autor(en), exklusiv lizenziert an Springer-Verlag GmbH, DE, ein Teil von Springer Nature 2023

Das Werk einschließlich aller seiner Teile ist urheberrechtlich geschützt. Jede Verwertung, die nicht ausdrücklich vom Urheberrechtsgesetz zugelassen ist, bedarf der vorherigen Zustimmung des Verlags. Das gilt insbesondere für Vervielfältigungen, Bearbeitungen, Übersetzungen, Mikroverfilmungen und die Einspeicherung und Verarbeitung in elektronischen Systemen.

Die Wiedergabe von allgemein beschreibenden Bezeichnungen, Marken, Unternehmensnamen etc. in diesem Werk bedeutet nicht, dass diese frei durch jedermann benutzt werden dürfen. Die Berechtigung zur Benutzung unterliegt, auch ohne gesonderten Hinweis hierzu, den Regeln des Markenrechts. Die Rechte des jeweiligen Zeicheninhabers sind zu beachten.

Der Verlag, die Autoren und die Herausgeber gehen davon aus, dass die Angaben und Informationen in diesem Werk zum Zeitpunkt der Veröffentlichung vollständig und korrekt sind. Weder der Verlag noch die Autoren oder die Herausgeber übernehmen, ausdrücklich oder implizit, Gewähr für den Inhalt des Werkes, etwaige Fehler oder Äußerungen. Der Verlag bleibt im Hinblick auf geografische Zuordnungen und Gebietsbezeichnungen in veröffentlichten Karten und Institutionsadressen neutral.

Planung/Lektorat: Ken Kissinger

Springer ist ein Imprint der eingetragenen Gesellschaft Springer-Verlag GmbH, DE und ist ein Teil von Springer Nature.

Die Anschrift der Gesellschaft ist: Heidelberger Platz 3, 14197 Berlin, Germany

*Für Sofie und Hannes*

# Vorwort

Wie funktioniert Künstliche Intelligenz? Wie funktioniert das Gehirn? Was sind die Gemeinsamkeiten von natürlicher und künstlicher Intelligenz und was die Unterschiede? Ist das Gehirn ein Computer? Was sind neuronale Netze? Was ist Deep Learning? Sollten wir versuchen, das Gehirn nachzubauen, um echte allgemeine Künstliche Intelligenz zu erschaffen, und wenn ja, wie gehen wir dabei am sinnvollsten vor?

Wir befinden uns in einer äußerst spannenden Phase der kulturellen und technologischen Entwicklung der Menschheit. In jüngster Zeit haben Künstliche Intelligenz (KI) und Maschinelles Lernen in immer mehr Bereichen, etwa in Medizin, Wissenschaft, Bildung, Finanzen, Technik, Unterhaltung und sogar Kunst und Musik, Einzug gehalten und sind dabei, im Leben des 21. Jahrhunderts allgegenwärtig zu werden. Insbesondere auf dem Gebiet des sogenannten Deep Learning sind die Fortschritte in jeder Hinsicht außergewöhnlich, und tiefe künstliche neuronale Netze zeigen in einer Vielzahl von Anwendungen wie der Verarbeitung, der Erkennung und Erzeugung von Bildern oder natürlicher Sprache beeindruckende Leistungen. Insbesondere in Kombination mit einem Verfahren namens Verstärkungslernen werden die Netze immer leistungsfähiger, wenn es beispielsweise darum geht, Videospiele zu spielen, oder sie erreichen sogar übermenschliche Fähigkeiten bei komplexen Brettspielen wie Go, wenn sie trainiert werden, indem sie Millionen Partien gegen sich selbst spielen.

Viele der heute in der KI genutzten Algorithmen, Designprinzipien und Konzepte wie neuronale Netze oder das bereits erwähnte Verstärkungslernen haben ihren Ursprung in Biologie und Psychologie. An immer mehr Universitäten sind daher Vorlesungen über Neurowissenschaften fester

Bestandteil in Studiengängen wie Informatik oder Künstlicher Intelligenz. Doch auch für Hirnforscher lohnt sich die Beschäftigung mit Künstlicher Intelligenz, stellt sie doch nicht nur wichtige Werkzeuge für die Auswertung von Daten zur Verfügung, sondern dient auch als Modell für natürliche Intelligenz und hat das Potential, unser Verständnis des Gehirns zu revolutionieren.

Führt man sich die Ziele von KI und Neurowissenschaften vor Augen, so fällt auf, dass diese komplementär zueinander sind. Ziel der KI ist es, Kognition und Verhalten auf menschlichem Niveau zu erreichen, und das Ziel der Neurowissenschaften ist, menschliche Kognition und Verhalten zu verstehen. Man könnte also sagen, Künstliche Intelligenz und Hirnforschung sind zwei Seiten einer Medaille. Die Konvergenz beider Forschungsfelder verspricht tiefgreifende Synergien, und schon heute steht fest, dass die sich daraus ergebenden Erkenntnisse unsere Zukunft nachhaltig prägen werden.

In den letzten Jahren habe ich viele Vorträge zu diesen und angrenzenden Themen gehalten. Aus den anschließenden Diskussionen und zahlreichen Rückfragen lernte ich, dass die tiefe Verbindung zwischen KI und Hirnforschung den meisten einerseits sofort einleuchtet, andererseits aber vorher nicht wirklich bewusst war. Obwohl sich dies allmählich zu ändern beginnt, assoziieren die meisten Menschen mit KI ausschließlich Studiengänge wie Informatik oder Datenwissenschaften und weniger z. B. Kognitionswissenschaften oder Computational Neuroscience, obwohl gerade diese Zweige der Wissenschaft viel zur Grundlagenforschung in der KI beitragen können. Umgekehrt ist Künstliche Intelligenz aus der modernen Hirnforschung inzwischen nicht mehr wegzudenken. Um zu verstehen, wie das menschliche Gehirn funktioniert, nutzen Forschungsteams in immer stärkerem Maß Modelle, die auf Verfahren der Künstlichen Intelligenz basieren, und gewinnen dadurch nicht nur neurowissenschaftliche Erkenntnisse, sondern lernen auch wieder etwas über Künstliche Intelligenz.

Es gibt bereits viele hervorragende Lehr- und Sachbücher, in denen die verschiedenen Disziplinen jeweils isoliert dargestellt werden. Eine integrierte Darstellung von KI und Hirnforschung existierte bislang aber nicht. Mit dem vorliegenden Buch möchte ich diese Lücke schließen. Anhand spannender und aktueller Forschungsergebnisse werden die grundlegenden Ideen und Konzepte, offene Fragen und zukünftige Entwicklungen im Schnittbereich aus KI und Hirnforschung verständlich dargestellt. Sie werden lernen, wie das menschliche Gehirn aufgebaut ist, auf welchen grundlegenden Mechanismen Wahrnehmen, Denken und Handeln beruhen, wie KI funktioniert und was hinter den spektakulären Leistungen

von AlphaGo, ChatGPT und Co. steckt. Wohlgermerkt geht es mir nicht um eine umfassende Einführung in KI oder Hirnforschung. Sie sollen lediglich mit dem aus meiner Sicht theoretischen Minimum ausgestattet werden, sodass Sie die Herausforderungen, ungelösten Probleme und schließlich die Integration beider Disziplinen verstehen können.

Das Buch ist in vier Teile gegliedert, die zum Teil aufeinander aufbauen, zum Teil aber auch unabhängig voneinander gelesen werden können. Es gibt also verschiedene Möglichkeiten, sich dem Inhalt dieses Buches zu nähern. Am liebsten wäre es mir natürlich, wenn Sie das Buch als Ganzes lesen, am besten zweimal: einmal, um sich einen Überblick zu verschaffen, und ein zweites Mal, um in die Details einzutauchen. Wenn Sie sich einen Überblick über die Funktionsweise des Gehirns verschaffen wollen, dann beginnen sie mit Teil I. Wenn es Ihnen aber eher darum geht, sich einen Überblick über den Stand der Forschung in der Künstlichen Intelligenz zu verschaffen, dann empfehle ich Ihnen, gleich Teil II zu lesen. Die offenen Fragen und Herausforderungen beider Disziplinen werden in Teil III dargestellt. Wenn Sie bereits mit den Grundlagen und offenen Fragen von KI und Hirnforschung vertraut sind und sich vor allem für die Integration beider Forschungsgebiete interessieren, dann lesen Sie Teil IV.

Ich habe versucht, wo immer möglich komplexe Sachverhalte durch anschauliche Abbildungen zu verdeutlichen. Bei der Erstellung der Abbildungen haben mich meine Kinder tatkräftig unterstützt. Englische Zitate wurden, wenn nicht anders vermerkt, von mir selbst übersetzt. Bei der Korrektur von Fehlern und der Verbesserung der Verständlichkeit und Lesbarkeit des Textes haben Kollegen, Freunde und Verwandte sehr geholfen. Ich möchte mich dafür bedanken bei Konstantin Tziridis, Claus Metzner, Holger Schulze, Nathaniel Melling, Tobias Olschewski, Peter Krauß und Katrin Krauß.

Mein besonderer Dank gilt Sarah Koch, Ramkumar Padmanaban und Ken Kissinger vom Springer Verlag, die mich bei der Realisierung dieses Buchprojekts unterstützt haben.

Meine Forschungsarbeiten wurden und werden von der Deutschen Forschungsgemeinschaft unterstützt. Den Verantwortlichen gilt mein Dank. Ohne die inspirierende Arbeitsatmosphäre an der Friedrich-Alexander-Universität Erlangen-Nürnberg und der Uniklinik Erlangen wären viele meiner Ideen und Forschungsprojekte nicht möglich gewesen. Mein besonderer Dank gilt Holger Schulze, Andreas Maier und Thomas Herbst für ihre Unterstützung sowie Claus Metzner und Achim Schilling für die zahllosen inspirierenden Gespräche. Meinem Vater danke ich herzlich für die vielen Diskussionen zu den diversen Themen dieses Buches. Mein größter Dank

**X      Vorwort**

gilt meiner Frau, die in all den Jahren immer alles mitgetragen hat und trägt.  
Was ich ihr verdanke, kann ich nicht in Worte fassen. Ich widme dieses  
Buch meinen Kindern.

Großenseebach  
im April 2023

Patrick Krauss

# Inhaltsverzeichnis

|                                       |  |            |
|---------------------------------------|--|------------|
| <b>1</b>                              | <b>Einführung</b>                          | <b>1</b>   |
| <b>Teil I Hirnforschung</b>           |  |            |
| <b>2</b>                              | <b>Das komplexeste System im Universum</b> | <b>17</b>  |
| <b>3</b>                              | <b>Bausteine des Nervensystems</b>         | <b>23</b>  |
| <b>4</b>                              | <b>Organisation des Nervensystems</b>      | <b>31</b>  |
| <b>5</b>                              | <b>Organisation des Cortex</b>             | <b>45</b>  |
| <b>6</b>                              | <b>Methoden der Hirnforschung</b>          | <b>57</b>  |
| <b>7</b>                              | <b>Gedächtnis</b>                          | <b>63</b>  |
| <b>8</b>                              | <b>Sprache</b>                             | <b>75</b>  |
| <b>9</b>                              | <b>Bewusstsein</b>                         | <b>83</b>  |
| <b>10</b>                             | <b>Freier Wille</b>                        | <b>105</b> |
| <b>Teil II Künstliche Intelligenz</b> |  |            |
| <b>11</b>                             | <b>Was ist Künstliche Intelligenz?</b>     | <b>117</b> |
| <b>12</b>                             | <b>Wie lernt Künstliche Intelligenz?</b>   | <b>125</b> |
| <b>13</b>                             | <b>Spielende Künstliche Intelligenz</b>    | <b>139</b> |
| <b>14</b>                             | <b>Rekurrente neuronale Netze</b>          | <b>145</b> |

**XII Inhaltsverzeichnis**

|                                   |   |            |
|-----------------------------------|---|------------|
| <b>15</b>                         | <b>Kreativität: Generative Künstliche Intelligenz</b> | <b>153</b> |
| <b>16</b>                         | <b>Sprachbegabte KI: ChatGPT und Co.</b>              | <b>163</b> |
| <b>17</b>                         | <b>Woran forschen KI-Entwickler heute?</b>            | <b>177</b> |
| <b>Teil III Herausforderungen</b> |   |            |
| <b>18</b>                         | <b>Herausforderungen der KI</b>                       | <b>187</b> |
| <b>19</b>                         | <b>Herausforderungen der Hirnforschung</b>            | <b>197</b> |
| <b>Teil IV Integration</b>        |   |            |
| <b>20</b>                         | <b>KI als Werkzeug in der Hirnforschung</b>           | <b>209</b> |
| <b>21</b>                         | <b>KI als Modell für das Gehirn</b>                   | <b>217</b> |
| <b>22</b>                         | <b>Mit Hirnforschung die KI besser verstehen</b>      | <b>223</b> |
| <b>23</b>                         | <b>Das Gehirn als Vorlage für KI</b>                  | <b>231</b> |
| <b>24</b>                         | <b>Ausblick</b>                                       | <b>241</b> |
|                                   | <b>Glossar</b>  | <b>253</b> |
|                                   | <b>Stichwortverzeichnis</b>                           | <b>289</b> |



# 1

## Einführung

*Ein Elefant ist wie ein Fächer!  
Der fünfte Blinde*

### ChatGPT besteht den Turing-Test

Im Bereich der Künstlichen Intelligenz (KI) gab es in den letzten etwa 10–15 Jahren eine ganze Reihe von spektakulären Durchbrüchen – von *AlphaGo* über *DALL-E 2* bis *ChatGPT* –, die so bis vor Kurzem noch völlig undenkbar waren.

Das jüngste Ereignis in dieser Reihe ist mit Sicherheit auch das spektakulärste: Bereits jetzt ist klar, dass der 30. November 2022 in die Geschichte eingehen wird. An diesem Tag machte die Firma OpenAI die Künstliche Intelligenz *ChatGPT* für die Öffentlichkeit frei zugänglich. Dieses sogenannte Große Sprachmodell kann in Sekundenschnelle jede beliebige Art von Text generieren, beantwortet Fragen zu jedem beliebigen Thema, gibt Interviews und führt Unterhaltungen, deren Verlauf es sich merkt und somit auch in längeren Konversationen adäquat antwortet. Millionen Menschen konnten sich seither tagtäglich von den erstaunlichen Fähigkeiten dieses Systems selbst überzeugen. Die von *ChatGPT* generierten Antworten und Texte sind dabei von solchen, die von Menschen erzeugt wurden, nicht unterscheidbar. *ChatGPT* besteht damit erstmals in der Geschichte der Künstlichen Intelligenz den Turing-Test, ein Verfahren, das

erdacht wurde, um zu entscheiden, ob eine Maschine über die Fähigkeit zu denken verfügt (Turing, 1950). Ein künstliches System, welches den Turing-Test besteht, galt jahrzehntelang als der Heilige Gral der Forschung auf dem Gebiet der Künstlichen Intelligenz. Auch wenn das Bestehen des Turing-Tests nicht zwingend bedeutet, dass *ChatGPT* tatsächlich denkt, sollten Sie sich den 30. November 2022 dennoch gut merken. Er stellt nicht nur den bisher wohl wichtigsten Meilenstein in der Geschichte der Künstlichen Intelligenz dar, sondern ist in seiner Tragweite sicherlich vergleichbar mit der Erfindung des Webstuhls, der Dampfmaschine, des Automobils, des Telefons, des Internets und des Smartphones, welche sich oft erst im Nachhinein als Game-Changer und entscheidende Wendepunkte in der Entwicklung herausgestellt haben.

## Die nächste Kränkung

Neben den viel diskutierten Konsequenzen, die *ChatGPT* und ähnliche KI-Systeme auf nahezu allen Ebenen unseres gesellschaftlichen Lebens haben werden, fordern die erstaunlichen Leistungen dieser neuen Systeme auch unsere Erklärungen dessen, was grundlegende Konzepte wie Kognition, Intelligenz und Bewusstsein überhaupt bedeuten, stark heraus. Insbesondere ist der Einfluss, den diese neue Art von KI auf unser Verständnis des menschlichen Gehirns haben wird, bereits jetzt immens und in seinen Auswirkungen noch gar nicht völlig absehbar.

Manche sprechen bereits von der nächsten großen Kränkung der Menschheit. Dabei handelt es sich um grundlegende Ereignisse oder Erkenntnisse, die im Laufe der Geschichte das Selbstverständnis des Menschen und sein Verhältnis zur Welt tiefgreifend erschüttert haben.

Die Kopernikanische Kränkung, benannt nach dem Astronomen Nikolaus Kopernikus, bezieht sich auf die Entdeckung, dass die Erde nicht der Mittelpunkt des Universums ist, sondern sich um die Sonne dreht. Diese Erkenntnis im 16. Jahrhundert veränderte das Weltbild grundlegend und führte zu einem Verlust an Selbstbezogenheit und Selbstsicherheit. Mit der Entdeckung von Tausenden von Exoplaneten, also Planeten außerhalb unseres Sonnensystems, in den letzten Jahrzehnten wurde diese Kränkung sogar noch verschärft. Hat dies doch gezeigt, dass Planetensysteme in unserer Galaxie sehr verbreitet sind und dass es möglicherweise sogar viele Planeten gibt, die in der habitablen Zone um ihre Sterne kreisen und somit mögliche Orte für Leben darstellen.

Eine weitere Erkenntnis, die sich auf das Selbstverständnis des Menschen auswirkte, war die Darwin'sche Kränkung. Die Evolutionstheorie von Charles Darwin im 19. Jahrhundert zeigte, dass der Mensch keine von Gott geschaffene Spezies ist, sondern sich wie alle anderen Spezies durch Evolution entwickelt hat. Diese Entdeckung stellte das Selbstverständnis des Menschen als einzigartige, von der übrigen Natur getrennte Spezies infrage.

Eine weitere Kränkung, welche Sigmund Freud wenig bescheiden nach der von ihm entwickelten Theorie die psychoanalytische Kränkung nannte, bezieht sich auf die Entdeckung, dass menschliches Verhalten und Denken nicht immer bewusst und rational gesteuert sind, sondern auch von unbewussten und irrationalen Trieben beeinflusst werden. Diese Erkenntnis erschütterte das Vertrauen des Menschen in seine Fähigkeit zur Selbstkontrolle und Rationalität. Die Libet-Experimente, welche schließlich sogar die Existenz des freien Willens infrage stellen, verschärften die Wucht dieser Kränkung noch weiter.

KI kann als neu hinzugekommene vierte große Kränkung für das menschliche Selbstverständnis angesehen werden. Bislang galt unsere hochentwickelte Sprache als das entscheidende Unterscheidungsmerkmal zwischen Menschen und anderen Spezies. Die Entwicklung großer Sprachmodelle wie *ChatGPT* hat jedoch gezeigt, dass Maschinen prinzipiell dazu in der Lage sind, mit natürlicher Sprache ähnlich wie Menschen umzugehen. Diese Tatsache stellt unser Konzept der Einzigartigkeit und Unvergleichbarkeit als Spezies erneut infrage und zwingt uns, unsere Definition des Menschseins zumindest teilweise zu überdenken.

Diese „KI-Kränkung“ betrifft nicht nur unsere sprachlichen Fähigkeiten, sondern unsere kognitiven Fähigkeiten im Allgemeinen. KI-Systeme sind bereits in der Lage, komplexe Probleme zu lösen, Muster zu erkennen und in bestimmten Bereichen menschenähnliche oder sogar übermenschliche Leistungen zu erbringen (Mnih et al., 2015; Silver et al., 2016, 2017a, 2017b; Schrittwieser et al., 2020; Perolat et al., 2022). Dies zwingt uns zu einer Neuinterpretation der menschlichen Intelligenz und Kreativität, bei der wir uns fragen müssen, welche Rolle der Mensch in einer Welt spielt, in der Maschinen viele unserer bisherigen Aufgaben übernehmen können. Auch zwingt sie uns, über die ethischen, sozialen und philosophischen Fragen nachzudenken, die sich aus der Einführung der KI in unser Leben ergeben. Beispielsweise stellt sich die Frage, wie wir mit der Verantwortung für Entscheidungen umgehen, die von KI-Systemen getroffen werden, und welche Grenzen wir dem Einsatz von KI setzen sollten, um sicherzustellen, dass sie dem Wohl der Menschheit dient (Anderson & Anderson, 2011; Goodall, 2014; Vinuesa et al., 2020).

Kein halbes Jahr nach der Veröffentlichung von *ChatGPT* wurde im März 2023 bereits dessen Nachfolger *GPT-4* veröffentlicht, welcher die Leistungsfähigkeit seines Vorgängers noch einmal deutlich übertrifft. Dies veranlasste einige der einflussreichsten Vordenker auf diesem Gebiet sogar dazu, in einem viel beachteten offenen Brief<sup>1</sup> eine vorübergehende Pause in der weiteren Entwicklung von KI-Systemen, welche noch leistungsfähiger als *GPT-4* sind, zu fordern, um einem möglicherweise drohenden Kontrollverlust vorzubeugen.

## Künstliche Intelligenz und Hirnforschung

Die erstaunlichen Leistungen von *ChatGPT* und *GPT-4* haben auch direkte Auswirkungen auf unser Verständnis des menschlichen Gehirns und seiner Funktionsweise. Sie fordern daher die Hirnforschung nicht nur heraus, sondern haben sogar das Potenzial, sie zu revolutionieren. In der Tat waren KI und Hirnforschung in ihrer Geschichte schon immer eng miteinander verflochten. Die sogenannte kognitive Revolution Mitte des letzten Jahrhunderts kann auch als Geburtsstunde der Forschung auf dem Gebiet der KI angesehen werden, wo sie sich als integraler Bestandteil der neu entstandenen Forschungsagenda der Kognitionswissenschaften als eigenständige Disziplin entwickelte. Tatsächlich ging es in der KI-Forschung nie nur darum, Systeme zu entwickeln, die uns lästige Arbeit abnehmen. Von Anfang an ging es auch darum, Theorien über natürliche Intelligenz zu entwickeln und zu testen. Wie wir sehen werden, konnten gerade in jüngster Zeit einige erstaunliche Parallelen zwischen KI-Systemen und Gehirnen aufgedeckt werden. KI spielt daher in der Hirnforschung eine immer größere Rolle, und zwar nicht nur als reines Werkzeug zur Analyse von Daten, sondern insbesondere auch als Modell für die Funktion des Gehirns.

Umgekehrt haben auch die Neurowissenschaften in der Geschichte der Künstlichen Intelligenz eine Schlüsselrolle gespielt und die Entwicklung neuer KI-Methoden immer wieder inspiriert. Die Übertragung von Design- und Verarbeitungsprinzipien aus der Biologie auf die Informatik hat das Potential, neue Lösungen für aktuelle Herausforderungen im Bereich der KI bereitzustellen. Auch dabei spielt die Hirnforschung nicht nur die Rolle, mit dem Gehirn ein Vorbild für neue KI-Systeme zur Verfügung zu stellen. Vielmehr wurde in den Neurowissenschaften eine Vielzahl von Methoden

---

<sup>1</sup> <https://futureoflife.org/open-letter/pause-giant-ai-experiments/>

zur Entschlüsselung der Repräsentations- und Rechenprinzipien natürlicher Intelligenz entwickelt, die jetzt wiederum als Werkzeug zum Verständnis Künstlicher Intelligenz eingesetzt und damit zur Lösung des sogenannten Black-Box-Problems beitragen können. Ein Unterfangen, welches gelegentlich als Neurowissenschaft 2.0 bezeichnet wird. Es zeichnet sich ab, dass beide Disziplinen in der Zukunft immer mehr miteinander verschmelzen werden (Marblestone et al., 2016; Kriegeskorte & Douglas, 2018; Rahwan et al., 2019; Zador et al., 2023).

## Zu blind, um den Elefanten zu sehen

Die Erkenntnis, dass verschiedene Disziplinen zusammenarbeiten müssen, um etwas derart Komplexes wie Kognition auf menschlichem Niveau zu verstehen, ist natürlich nicht neu und wird in der bekannten Metapher von den sechs Blinden und dem Elefanten anschaulich illustriert (Friedenberg et al., 2021):

Es waren einmal sechs blinde Wissenschaftlerinnen und Wissenschaftler, die noch nie einen Elefanten gesehen hatten und erforschen wollten, was ein Elefant ist und wie er aussieht. Jeder untersuchte einen anderen Körperteil und kam entsprechend zu einer anderen Schlussfolgerung.

Die erste Blinde näherte sich dem Elefanten und berührte seine Seite. „Ah, ein Elefant ist wie eine Wand“, sagte sie.

Die zweite Blinde berührte den Stoßzahn des Elefanten und rief: „Nein, ein Elefant ist wie ein Speer!“

Die dritte Blinde berührte den Rüssel des Elefanten und sagte: „Ihr irrt euch beide! Ein Elefant ist wie eine Schlange!“

Der vierte Blinde berührte ein Bein des Elefanten und sagte: „Ihr irrt euch alle. Ein Elefant ist wie ein Baumstamm.“

Der fünfte Blinde berührte das Ohr des Elefanten und sagte: „Keiner von euch weiß, wovon ihr redet. Ein Elefant ist wie ein Fächer.“

Schließlich näherte sich der sechste Blinde dem Elefanten und berührte seinen Schwanz: „Ihr irrt euch alle“, sagte er. „Ein Elefant ist wie ein Seil.“

Hätten die sechs Wissenschaftlerinnen und Wissenschaftler ihre Erkenntnisse kombiniert, wären sie der wahren Natur des Elefanten viel näher gekommen. In dieser Geschichte steht der Elefant für den menschlichen Geist, und die sechs Blinden stehen für die verschiedenen wissenschaftlichen Disziplinen, die versuchen, seine Funktionsweise jeweils aus verschiedenen Perspektiven zu ergründen (Abb. 1.1). Die Pointe der



**Abb. 1.1 Die Blinden und der Elefant.** Jeder untersucht einen anderen Körperteil und kommt entsprechend zu einer anderen Schlussfolgerung. Der Elefant steht für Geist und Gehirn, und die sechs Blinden stehen für verschiedene Wissenschaften. Die Sichtweise jeder einzelnen Disziplin ist wertvoll, ein umfassendes Verständnis kann jedoch nur durch die Zusammenarbeit und den interdisziplinären Austausch erreicht werden

Geschichte ist, dass die Sichtweise jedes Einzelnen zwar wertvoll ist, dass aber ein umfassendes Verständnis von Kognition nur erreicht werden kann, wenn die unterschiedlichen Wissenschaften zusammenarbeiten und sich austauschen.

Dies ist der Gründungsgedanke der Kognitionswissenschaften, die in den 1950er-Jahren als intellektuelle Bewegung begannen, welche als kognitive Revolution bezeichnet wurde (Sperry, 1993; Miller, 2003). In dieser Zeit kam es zu großen Veränderungen in der Arbeitsweise von Psychologen und Linguisten und zur Entstehung neuer Disziplinen wie Informatik und Neurowissenschaften. Die kognitive Revolution wurde durch eine Reihe von Faktoren vorangetrieben, darunter die rasche Entwicklung von Personal Computern und neuen bildgebenden Verfahren für die Hirnforschung. Diese technologischen Fortschritte ermöglichten es den Forschern, besser zu verstehen, wie das Gehirn funktioniert und wie Informationen verarbeitet, gespeichert und abgerufen werden. Als Folge dieser Entwicklungen ent-

stand in den 1960er-Jahren ein interdisziplinäres Gebiet, das Forscher aus den unterschiedlichsten Disziplinen zusammenführte. Dieses Gebiet trug verschiedene Namen, darunter Psychologie der Informationsverarbeitung, Kognitionsforschung und eben auch Kognitionswissenschaft.

Die kognitive Revolution markierte einen wichtigen Wendepunkt in der Geschichte der Psychologie und verwandter Disziplinen. Sie hat die Art und Weise, wie Forscher Fragen der menschlichen Kognition und des menschlichen Verhaltens angehen, grundlegend verändert und den Weg für zahlreiche Durchbrüche in Bereichen wie der Künstlichen Intelligenz, der Kognitiven Psychologie und den Neurowissenschaften geebnet.

Heute versteht man unter Kognitionswissenschaft ein interdisziplinäres wissenschaftliches Unterfangen zur Erforschung der unterschiedlichen Aspekte von Kognition. Dazu gehören Sprache, Wahrnehmung, Gedächtnis, Aufmerksamkeit, logisches Denken, Intelligenz, Verhalten und Emotionen. Hierbei konzentriert man sich vor allem auf die Art und Weise, wie natürliche oder künstliche Systeme Informationen repräsentieren, verarbeiten und umwandeln (Bermúdez, 2014; Friedenberget al., 2021).

Die Schlüsselfragen sind: Wie funktioniert der menschliche Geist? Wie funktioniert Kognition? Wie ist Kognition im Gehirn implementiert? Und wie kann Kognition in Maschinen umgesetzt werden?

Damit widmen sich die Kognitionswissenschaften einigen der schwierigsten wissenschaftlichen Probleme überhaupt, da das Gehirn unglaublich schwer zu beobachten, zu messen und zu manipulieren ist. Viele Wissenschaftler halten das Gehirn sogar für das komplexeste System im bekannten Universum.

Zu den beteiligten Disziplinen der Kognitionswissenschaften gehören heute Linguistik, Psychologie, Philosophie, Informatik, Künstliche Intelligenz, Neurowissenschaft, Biologie, Anthropologie und Physik (Bermúdez, 2014). Zwischenzeitlich waren die Kognitionswissenschaften etwas aus der Mode gekommen, insbesondere die Idee der integrativen Zusammenarbeit der unterschiedlichen Disziplinen geriet teilweise in Vergessenheit. Speziell KI und Neurowissenschaft entwickelten sich eigenständig weiter und somit auch voneinander weg. Erfreulicherweise erlebt die Idee, dass KI und Hirnforschung komplementär zueinander sind und viel von der jeweils anderen Disziplin profitieren können, derzeit eine regelrechte Renaissance, wobei der Terminus „Kognitionswissenschaft“ anscheinend heute in manchen Communities entweder anders interpretiert wird oder als zu unmodern gilt, weshalb stattdessen Begriffe wie *Cognitive Computational Neuroscience* (Kriegeskorte & Douglas, 2018) oder *NeuroAI* (Zador et al., 2023) vorgeschlagen wurden.

Das Erbe der kognitiven Revolution zeigt sich in den vielen innovativen und interdisziplinären Ansätzen, die unser Verständnis des menschlichen Geistes und seiner Funktionsweise weiterhin prägen. Ob mithilfe modernster bildgebender Verfahren des Gehirns, ausgefeilter Computermodelle oder neuer theoretischer Rahmenkonzepte – die Forscherinnen und Forscher verschieben immer wieder die Grenzen dessen, was wir über das menschliche Gehirn und seine komplexen Prozesse wissen.

## Gehirn-Computer-Analogie

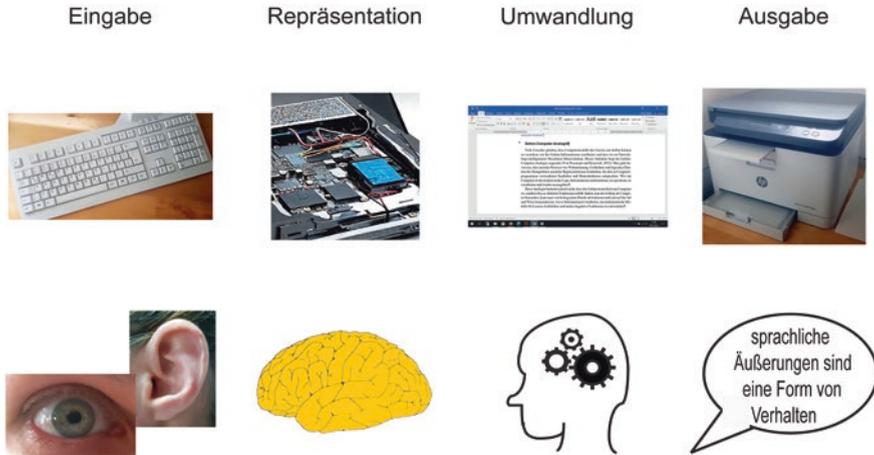
Viele Forscher glauben, dass Computermodelle des Geistes uns helfen können zu verstehen, wie das Gehirn Informationen verarbeitet, und dass sie zur Entwicklung intelligenterer Maschinen führen können. Dieser Annahme liegt die Gehirn-Computer-Analogie zugrunde (Von Neumann & Kurzweil, 2012). Man geht davon aus, dass mentale Prozesse wie Wahrnehmung, Gedächtnis und logisches Denken die Manipulation mentaler Repräsentationen beinhalten, die den in Computerprogrammen verwendeten Symbolen und Datenstrukturen entsprechen (Abb. 1.2). Wie ein Computer ist das Gehirn in der Lage, Informationen aufzunehmen, zu speichern, zu verarbeiten und wieder auszugeben.<sup>2</sup>

Diese Analogie bedeutet jedoch nicht, dass das Gehirn tatsächlich ein Computer ist, sondern dass es ähnliche Funktionen erfüllt. Indem man das Gehirn als Computer betrachtet, kann man von biologischen Details abstrahieren und sich auf die Art und Weise konzentrieren, wie es Informationen verarbeitet, um mathematische Modelle für Lernen, Gedächtnis und andere kognitive Funktionen zu entwickeln.

Die Gehirn-Computer-Analogie stützt sich auf zwei zentrale Annahmen, welche den Kognitionswissenschaften zugrunde liegen. Diese sind Computationalismus und Funktionalismus.

---

<sup>2</sup>Ein fundamentaler Unterschied ist, dass ein Computer Informationen mit anderen Bauteilen verarbeitet als denen, mit denen er die Informationen speichert. Im Gehirn machen beides die – mitunter selben – Neurone.



**Abb. 1.2 Gehirn-Computer-Analogie.** Informationsverarbeitung umfasst die Eingabe, Repräsentation, Umwandlung und Ausgabe von Informationen. Beim Computer kann die Eingabe beispielsweise von der Tastatur kommen, beim biologischen Organismus von den Sinnesorganen. Anschließend muss diese Eingabe repräsentiert werden: durch Speicherung auf einer Festplatte oder im RAM des Computers, oder im Gehirn als momentane neuronale Aktivität im Kurzzeitgedächtnis oder im Langzeitgedächtnis in der Verschaltung der Neuronen. Danach findet eine Umwandlung oder Verarbeitung statt, d. h., mentale Prozesse oder Algorithmen müssen auf die gespeicherte Information einwirken und sie verändern, um neue Information zu erzeugen. Beim Computer könnte das z. B. eine Textverarbeitung sein, beim Menschen z. B. logisches Schlussfolgern. Schließlich wird das Ergebnis der Informationsverarbeitung ausgegeben. Die Ausgabe kann beim Computer z. B. über einen Drucker erfolgen. Bei Lebewesen entspricht der Output dem beobachtbaren Verhalten oder auch, als Spezialfall von Verhalten, beim Menschen sprachlichen Äußerungen

## Computationalismus

Im Computationalismus geht man davon aus, dass Kognition gleichbedeutend mit Informationsverarbeitung ist, d. h., dass mentale Prozesse als Berechnungen verstanden werden können und dass das Gehirn im Wesentlichen ein informationsverarbeitendes System ist (Dietrich, 1994; Shapiro, 1995; Piccinini, 2004, 2009). Wie jedes derartige System muss demnach auch das Gehirn Informationen repräsentieren und diese repräsentierten Informationen anschließend transformieren, d. h., es muss mentale Repräsentationen von Informationen geben und es muss mentale Prozesse geben, die auf diese Repräsentationen einwirken und sie manipulieren können. Der Computationalismus hat die Art und Weise, wie Kognitions-

wissenschaftler und Forscher im Bereich der Künstlichen Intelligenz über Intelligenz und Kognition denken, stark beeinflusst.

Es gibt jedoch auch Kritik an dieser Sichtweise, was zahlreiche bis heute andauernde Debatten in der Philosophie und den Kognitionswissenschaften belegen. Einige Kritiker argumentieren beispielsweise, dass das Computermodell des Geistes zu einfach ist und die Komplexität und den Reichtum der menschlichen Kognition nicht vollständig erfassen kann. Andere argumentieren, es sei unklar, ob mentale Prozesse wirklich als Berechnungen verstanden werden können oder ob sie sich nicht doch grundlegend von der Art und Weise unterscheiden, wie Prozesse in Computern ablaufen.

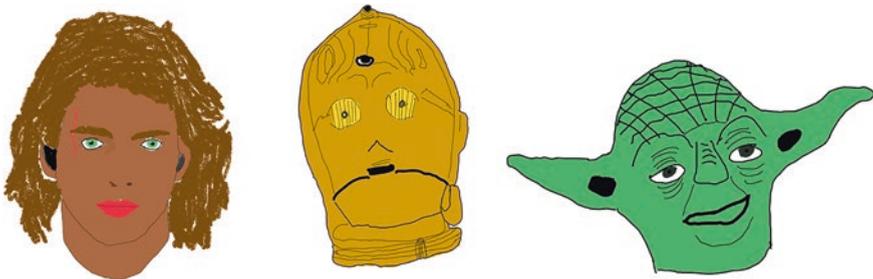
## Funktionalismus

Ist Kognition nur in einem (menschlichen) Gehirn möglich? Der Funktionalismus beantwortet diese Frage ganz klar mit Nein. Demnach werden mentale Zustände und Prozesse ausschließlich durch ihre Funktionen oder ihre Beziehung zum Verhalten und nicht durch ihre physikalischen oder biochemischen Eigenschaften definiert (Shoemaker, 1981; Jackson & Pettit, 1988; Piccinini, 2004). Was bedeutet das konkret?

Stellen Sie sich jetzt bitte vor Ihrem geistigen Auge ein Auto vor. Und nun erinnern Sie sich an die letzte Situation, in der Sie Schokolade gegessen haben, und stellen Sie sich den Geschmack so genau wie möglich vor. Ist es Ihnen gelungen? Ich nehme an, das ist es. Während ich diese Zeilen schreibe, habe ich mir dieselben beiden mentalen Zustände „*Sehen eines Autos*“ und „*Schmecken von Schokolade*“ vergegenwärtigt. Offensichtlich kann jeder von uns die entsprechenden mentalen Repräsentationen in seinem Gehirn aktivieren und dies, obwohl Sie, ich und jeder andere Leser dieser Zeilen ein völlig anderes Gehirn hat. Alle menschlichen Gehirne sind sich natürlich in ihrem grundlegenden Aufbau ähnlich. Aber sie sind sicherlich nicht bis ins kleinste Detail identisch und schon gar nicht in der exakten Verschaltung der Neuronen, alleine schon deshalb, weil jeder Mensch völlig andere, individuelle Erfahrungen in seinem Leben gemacht hat, welche sich auf das Verschaltungsmuster des Gehirns auswirken. In der Terminologie der Informatik würde man sagen, jeder Mensch hat eine andere, individuelle Hardware. Dennoch können wir alle uns den gleichen mentalen Zustand vergegenwärtigen.

Während im vorherigen Beispiel die Systeme trotzdem irgendwie sehr ähnlich waren – es handelte sich immer um menschliche Gehirne –, mag das folgende Beispiel illustrieren, wie stark sich die verschiedenen physikalischen Implementationen desselben Algorithmus voneinander unterscheiden können. Betrachten wir die Addition zweier Zahlen. Die Repräsentation dieser Zahlen sowie der dazugehörige Prozess oder Algorithmus, um sie zu addieren, können in Ihrem Gehirn implementiert sein, wenn sie „im Kopf rechnen“, oder z. B. auch in einem Laptop mit Tabellenkalkulationsprogramm, einem Rechenschieber, einem Taschenrechner oder einer Taschenrechner-App auf Ihrem Smartphone. Jedes Mal werden dieselben Zahlen repräsentiert und addiert, wobei die informationsverarbeitenden Systeme völlig verschieden sind. Das ist das *Konzept der multiplen Realisierbarkeit*.

Demnach kann derselbe mentale Zustand oder Prozess prinzipiell durch völlig verschiedene natürliche oder künstliche Systeme realisiert werden. Vereinfacht ausgedrückt bedeutet das, dass Kognition und vermutlich auch Bewusstsein prinzipiell in jedem physikalischen System implementiert sein können, welches in der Lage ist, die erforderlichen Berechnungen zu unterstützen. Wenn also bereits viele verschiedene menschliche Gehirne dazu in der Lage sind, warum sollte diese Fähigkeit auf den Menschen oder auf biologische Systeme beschränkt sein? Aus Sicht des Funktionalismus ist es daher durchaus möglich, dass die Fähigkeit zu menschenähnlicher Kognition auch in entsprechend hoch entwickelten Maschinen oder außerirdischen Gehirnen implementiert sein kann (Abb. 1.3).



**Abb. 1.3 Funktionalismus.** Kognition auf menschlichem Niveau ist nicht auf ein menschliches Gehirn beschränkt, sondern könnte prinzipiell auch in jedem anderen System implementiert sein, welches die erforderlichen Berechnungen unterstützt, beispielsweise entsprechend hochentwickelte Roboter oder Außerirdische

## Fazit

In den letzten Jahren haben spektakuläre Fortschritte in der Künstlichen Intelligenz unser Verständnis von Kognition, Intelligenz und Bewusstsein auf den Kopf gestellt und werden tiefgreifende Auswirkungen auf die Gesellschaft und unser Verständnis des menschlichen Gehirns haben. Die Kognitionswissenschaften sind der Schlüssel zu einem tieferen Verständnis von Gehirn und Geist, und Computermodelle des Geistes können uns helfen zu verstehen, wie das Gehirn Informationen verarbeitet, und zur Entwicklung intelligenterer Maschinen beitragen. Diese Modelle basieren auf den zentralen Annahmen des Computationalismus und des Funktionalismus, die die Äquivalenz von Kognition und Informationsverarbeitung sowie die Unabhängigkeit kognitiver Prozesse von ihrer physischen Implementierung betonen.

Die Fortschritte in der Künstlichen Intelligenz haben ebenfalls dazu geführt, dass die Bereiche der Neurowissenschaften und der Informatik immer enger zusammenwachsen. Die Übertragung von Konstruktions- und Verarbeitungsprinzipien aus der Biologie auf die Informatik verspricht neue Lösungen für aktuelle Herausforderungen der Künstlichen Intelligenz. Umgekehrt wird die enge Zusammenarbeit dieser Disziplinen in Zukunft immer wichtiger werden, um komplexe Systeme wie das menschliche Gehirn zu verstehen.

Die jüngsten Fortschritte in der Künstlichen Intelligenz und ihre Anwendungen haben in den Neurowissenschaften die Tür zu neuen Erkenntnissen und Technologien weit über das bisher Mögliche hinaus geöffnet. Wir stehen erst am Anfang einer neuen Ära der Forschung und Innovation, und es bleibt abzuwarten, welche faszinierenden Entdeckungen und Entwicklungen uns in Zukunft erwarten.

## Literatur

- Anderson, M., & Anderson, S. L. (Hrsg.). (2011). *Machine ethics*. Cambridge University Press.
- Bermúdez, J. L. (2014). *Cognitive science: An introduction to the science of the mind*. Cambridge University Press.
- Dietrich, E. (1994). *Computationalism*. In *thinking computers and virtual persons* (S. 109–136). Academic Press.
- Friedenberg, J., Silverman, G., & Spivey, M. J. (2021). *Cognitive science: An introduction to the study of mind*. Sage Publications.

- Goodall, N. J. (2014). Machine Ethics and Automated Vehicles. In: Meyer, G., Beiker, S. (eds) *Road Vehicle Automation. Lecture Notes in Mobility*. Springer, Cham. [https://doi.org/10.1007/978-3-319-05990-7\\_9](https://doi.org/10.1007/978-3-319-05990-7_9)
- Jackson, F., & Pettit, P. (1988). Functionalism and broad content. *Mind*, 97(387), 381–400.
- Kriegeskorte, N., & Douglas, P. K. (2018). Cognitive computational neuroscience. *Nature neuroscience*, 21(9), 1148–1160.
- Marblestone, A. H., Wayne, G., & Kording, K. P. (2016). Toward an integration of deep learning and neuroscience. *Frontiers in computational neuroscience*, 10, 94.
- Miller, G. A. (2003). The cognitive revolution: A historical perspective. *Trends in cognitive sciences*, 7(3), 141–144.
- Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., ... & Hassabis, D. (2015). Human-level control through deep reinforcement learning. *Nature*, 518(7540), 529–533.
- Perolat, J., De Vylder, B., Hennes, D., Tarassov, E., Strub, F., de Boer, V., ... & Tuyls, K. (2022). Mastering the game of Stratego with model-free multiagent reinforcement learning. *Science*, 378(6623), 990–996.
- Piccinini, G. (2004). Functionalism, computationalism, and mental contents. *Canadian Journal of Philosophy*, 34(3), 375–410.
- Piccinini, G. (2009). Computationalism in the philosophy of mind. *Philosophy Compass*, 4(3), 515–532.
- Rahwan, I., Cebrian, M., Obradovich, N., et al. (2019). Machine behaviour. *Nature*, 568, 477–486.
- Schrittwieser, J., Antonoglou, I., Hubert, T., Simonyan, K., Sifre, L., Schmitt, S., ... & Silver, D. (2020). Mastering Atari, Go, Chess and Shogi by planning with a learned model. *Nature*, 588(7839), 604–609.
- Shapiro, S. C. (1995). Computationalism. *Minds and Machines*, 5, 517–524.
- Shoemaker, S. (1981). Some varieties of functionalism. *Philosophical topics*, 12(1), 93–119.
- Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., Van Den Driessche, G., ... & Hassabis, D. (2016). Mastering the game of Go with deep neural networks and tree search. *Nature*, 529(7587), 484–489.
- Silver, D., Schrittwieser, J., Simonyan, K., Antonoglou, I., Huang, A., Guez, A., ... & Hassabis, D. (2017a). Mastering the game of Go without human knowledge. *Nature*, 550(7676), 354–359.
- Silver, D., Hubert, T., Schrittwieser, J., Antonoglou, I., Lai, M., Guez, A., ... & Hassabis, D. (2017b). *Mastering Chess and Shogi by self-play with a general reinforcement learning algorithm*. arXiv preprint [arXiv:1712.01815](https://arxiv.org/abs/1712.01815).
- Sperry, R. W. (1993). The impact and promise of the cognitive revolution. *American Psychologist*, 48(8), 878.
- Turing, A. M. (1950). Computing machinery and intelligence. *Mind*, 59(236), 433–460.

- Vinuesa, R., Azizpour, H., Leite, I., Balaam, M., Dignum, V., Domisch, S., ... & Fuso Nerini, F. (2020). The role of artificial intelligence in achieving the sustainable development goals. *Nature Communications*, *11*(1), 233.
- Von Neumann, J., & Kurzweil, R. (2012). *The computer and the brain*. Yale University Press.
- Zador, A., Escola, S., Richards, B., Ölveczky, B., Bengio, Y., Boahen, K., ... & Tsao, D. (2023). Catalyzing next-generation artificial intelligence through NeuroAI. *Nature Communications*, *14*(1), 1597.

# Teil I

# Hirnforschung

Im ersten Teil des Buches soll es darum gehen, Sie mit den wichtigsten Aspekten zu Aufbau und Funktion des Gehirns vertraut zu machen. Hierbei wird bewusst auf eine eingehende und systematische Beschreibung vieler molekularbiologischer, physiologischer und anatomischer Details verzichtet. Auch erhebt die Darstellung keinerlei Anspruch auf Vollständigkeit. Interessierte Leserinnen und Leser mögen ihr Wissen mit einem der zahlreich verfügbaren sehr guten Lehrbücher zu Psychologie und Neurowissenschaft vertiefen. Vielmehr sollen diese ersten Kapitel dazu dienen, die aus der Sicht des Autors notwendigen Grundlagen zu vermitteln, anhand derer wir in späteren Kapiteln die zahlreichen Querverbindungen zur Künstlichen Intelligenz aufzeigen wollen.



# 2

## Das komplexeste System im Universum

*Es gibt immer einen noch größeren Fisch.*

*Qui-Gon Jinn*

### Das Gehirn in Zahlen

Das menschliche Gehirn besteht aus ca. 86 Mrd. Nervenzellen, den sogenannten Neuronen (Herculano-Houzel, 2009). Das sind die fundamentalen Verarbeitungseinheiten, die für die Aufnahme, Verarbeitung und Weiterleitung von Informationen im gesamten Körper verantwortlich sind. Die Neuronen sind über sogenannte Synapsen verbunden und bilden ein gigantisches neuronales Netzwerk. Im Mittel empfängt jedes Neuron von etwa 10.000 anderen Neuronen seinen Input und sendet, ebenfalls im Mittel, seinen Output an etwa 10.000 nachfolgende Neurone (Kandel et al., 2000; Herculano-Houzel, 2009). Die tatsächliche Anzahl von Verbindungen pro Neuron kann hierbei erheblich, und zwar über mehrere Größenordnungen, variieren, weshalb man auch von einer breiten Verteilung der Verbindungen pro Neuron spricht. Manche Neurone, etwa im Rückenmark, sind nur mit einem einzigen anderen Neuron verbunden, während andere, beispielsweise im Kleinhirn, mit bis zu einer Million anderen Neurone verbunden sein können.

Aufgrund der Gesamtzahl der Neurone und der mittleren Anzahl von Verbindungen pro Neuron lässt sich die Gesamtzahl der synaptischen Verbindungen im Gehirn auf grob eine Billion abschätzen. Das ist eine Zahl

mit 15 Nullen und kann auch geschrieben werden als  $10^{15}$ . In den vergangenen Jahren haben wir uns im Kontext von Politik und Wirtschaft inzwischen einigermaßen an Beträge jenseits der Tausend Milliarden, also im Billionenbereich gewöhnt. Die ungefähre Zahl der Synapsen im Gehirn ist noch einmal um den Faktor tausend größer!

Das klingt erstmal nach sehr viel, aber überlegen wir, wie viele Synapsen maximal im Gehirn theoretisch möglich wären. Jedes der ca.  $10^{11}$  Neurone könnte im Prinzip mit jedem anderen in Verbindung stehen, wobei die Information zwischen zwei beliebigen Neuronen grundsätzlich auch in zwei Richtungen laufen kann: entweder von Neuron A zu Neuron B oder umgekehrt. Zusätzlich kann jedes Neuron tatsächlich auch mit sich selbst verbunden sein. Diese spezielle Art von Verbindungen nennt man Autapsen. Rein kombinatorisch ergibt sich somit  $10^{11}$  mal  $10^{11}$ , also  $10^{22}$  als mögliche Anzahl an Synapsen. Ein Vergleich mit der Anzahl real existierender Synapsen ergibt, dass nur etwa eine von 10 Mio. theoretisch möglichen Verbindungen tatsächlich realisiert ist. Das Netzwerk, das die Neurone im Gehirn bilden, ist also alles andere als dicht (*dense*), sondern im Gegenteil extrem dünn (*sparse*) (Hagmann, 2008).

## Wie viele verschiedene Gehirne kann es geben?

In Anlehnung an das Genom, welches die Gesamtheit aller Gene eines Organismus bezeichnet, ist das Konnektom die Gesamtheit aller Verbindungen im Nervensystems eines Lebewesens (Sporns et al., 2005). Um die Frage, wie viele verschiedene Gehirne es geben kann, zu beantworten, muss man abschätzen, wie viele verschiedene Konnektome rein kombinatorisch möglich sind. An dieser Stelle sei einschränkend angemerkt, dass nicht jedes theoretisch mögliche Konnektom auch ein funktionierendes lebensfähiges Nervensystem ergeben muss. Es stellt sich heraus, dass es ziemlich kompliziert ist, die exakte Anzahl zu berechnen, weshalb wir uns hier mit einer Abschätzung für die untere Grenze der tatsächlichen Anzahl auf Basis einiger Vereinfachungen begnügen wollen. Nehmen wir vereinfachend an, dass jede der  $10^{22}$  theoretisch möglichen Verbindungen entweder vorhanden oder nicht vorhanden sein kann. Wir gehen also von binären Verbindungen aus, wobei Eins einer vorhandenen und Null einer nicht vorhandenen Verbindung entspricht. Wie wir später noch sehen werden, ist die Realität noch deutlich komplizierter. Aber selbst unter dieser starken Vereinfachung ergibt sich bereits eine absurd hohe Anzahl von  $(2^{10})^{22}$  (sprich „zwei hoch zehn hoch 22“). Dies entspricht einer Zahl mit

einer Trilliarde Nullen. Natürlich führt nicht jedes dieser Konnektome zu einem leistungsfähigen und lebensfähigen Nervensystem, sodass die realisierbare Anzahl nur einer winzigen Untermenge aller theoretisch möglichen Konnektome entsprechen dürfte. Andererseits sind die Synapsengewichte aber keine Binärzahlen, sondern können jeden beliebigen kontinuierlichen Wert annehmen, was die Anzahl der Möglichkeiten wiederum noch einmal deutlich erhöht.

## Wie viele verschiedene Geisteszustände sind möglich?

Überlegen wir uns nun, wie viele verschiedene Gehirn- oder Geisteszustände es geben kann. Warum ist das wichtig, fragen Sie sich? Nun, eine der zentralen Annahmen der modernen Kognitions- und Neurowissenschaften ist, dass jedem mentalen Zustand (Geisteszustand) ein neuronaler Zustand (Gehirnzustand) oder eine ganze Abfolge von Gehirnzuständen zugrunde liegt. Oder anders ausgedrückt: Alles, was wir denken, fühlen und erleben können, hat ein bestimmtes neuronales Korrelat, also eine Aktivierung oder auch eine zeitliche Sequenz von Aktivierungen von Neuronen in unserem Gehirn. Man spricht in diesem Zusammenhang auch von raumzeitlichen Aktivierungsmustern.

Auch bei dieser Abschätzung gehen wir wieder von Vereinfachung aus. Zu jedem gegebenen Zeitpunkt kann ein bestimmtes Neuron entweder aktiv sein, also ein Aktionspotential aussenden, was einer Eins entspricht, oder es tut dies nicht und ist somit inaktiv, was einer Null entspricht. Auch hier landen wir also wieder beim Binärsystem. Als weitere Vereinfachung unterteilen wir den kontinuierlichen Zeitstrom in kleinste sinnvolle Einheiten. Die typische Dauer eines Aktionspotentials beträgt eine Millisekunde, also eine tausendstel Sekunde. Dies ist die charakteristische Zeitskala des Gehirns. In jeder Millisekunde kann demnach jedes der  $10^{11}$  Neurone im Gehirn entweder aktiv oder inaktiv sein. Dies ergibt rein kombinatorisch  $(2^{10})^{11}$  verschiedene Aktivierungsmuster pro Millisekunde. Nun haben wir natürlich nicht jede Millisekunde oder tausendmal pro Sekunde einen neuen Gedanken, Wahrnehmungs- oder Gefühlseindruck. Man geht davon aus, dass das, was wir als Jetzt, die Gegenwart oder den Moment erleben, eine Dauer von ca. drei Sekunden hat (Pöppel, 1997), was 3000 Millisekunden entspricht. Jeder mentale Zustand entspräche demnach innerhalb unserer Vereinfachungen einer Sequenz von 3000 verschiedenen Aktivierungsmustern.