

2 Neuronenmodelle. Die Subsumierung neuronaler Aktivität unter die Logik Neuronaler Netze

Anfang des 20. Jahrhunderts, etwa 150 Jahre nach Leibniz' ersten Versuchen einer universalen Sprache und 50 Jahre nach Fouriers Entdeckung, feste Körper durch ihre physikalischen Eigenschaften beschreiben zu können, entwickelten sich die zunächst nur addierenden und subtrahierenden, später auch multiplizierenden und dividierenden Rechenmaschinen durch die Lösung eines grundlegenden Problems weiter: Das Entscheidungsproblem wurde von Alan Turing so gelöst, dass es in Maschinen implementiert werden konnte. Ab den 1940er- und 1950er-Jahren verbinden sich dann verschiedene Konzepte ungleicher Disziplinen, bringen die Informatik und ein neues Neuronenmodell hervor.

2.1 Die Lösung des Entscheidungsproblems – Turing

Zunächst galt es, das Entscheidungsproblem, das bereits Mitte des 19. Jahrhunderts als mathematisches ›Problem‹ konstatiert wurde, zu lösen. Das Entscheidungsproblem stellte sich die Frage, ob es einen Algorithmus gibt, mit dem alle Entscheidungsfragen so ausgedrückt werden können, dass man sie einfach beantworten kann. Das Entscheidungsproblem wurde von dem Mathematiker und Informatiker Alan Turing (1912–1954) durch den Bau einer Maschine und durch einen Entscheidungsalgorithmus in eine wegweisende Richtung gebracht.

Das Entscheidungsproblem bezieht sich prinzipiell auf die Schwierigkeit, Aussagen so zu definieren, dass sie maschinell berechnet werden können. Es musste ein allgemeines Verfahren festgelegt werden, um zu entscheiden, ob ein theoretischer Satz als korrekt identifiziert werden kann oder nicht. Turing hat hierfür die Frage, was einen Algorithmus ausmacht, präzisiert und ermöglichte dadurch die sequentielle Berechnung einer Kette von Algorithmen ohne menschliches Eingreifen durch den Bau einer Maschine, die nach Beantwortung eines Algorithmus weitere nachfolgende Optionen vorschlug. Turing war sich bewusst, dass das menschliche Entscheidungsproblem nicht generell gelöst werden konnte, und konzentrierte sich auf einen Vorschlag, wie Entscheidungen formalisiert werden müssen, um sie mithilfe von Algorithmen zu verarbeiten.

Seine Lösung basierte auf Überlegungen der formalen Mathematik, die zwei Haupttypen von Algorithmen definiert: »Fragen, die eine ›Ja/Nein-«

Antwort erfordern (Entscheidungsalgorithmen), und ›Was‹-Fragen (Berechnungsalgorithmen).« (Heintz 1993, 74) Die Implementierung formaler Logik in die Verrechnungsschritte der Maschine ermöglichte es Turing, den Entscheidungsprozess auf »Ja/Nein«-Fragen zu beschränken und damit den Radius programmierbarer und damit auch beantwortbarer Entscheidungsfragen zu reduzieren. Jede Antwort auf eine »Ja/Nein«-Frage führte zu einer neuen »Ja/Nein«-Frage, hier wurden menschliche Handlungsweisen als Vorlagen für berechnete Aktionen herangezogen. Dazu reduzierte Turing mechanische Vorgehensweisen auf einfache Schritte, um ihre Komplexität zu minimieren und in einer schematischen Reihenfolge zu speichern. Die Frage, was mit einer Rechenmaschine beantwortet werden kann und auf welche Fragen Maschinen Lösungen bereithalten können, wurde von Turing damals ziemlich klar beantwortet. Er war davon überzeugt, dass Maschinen quasi unendlich lange Ketten von Algorithmen berechnen können, die Fragen aber in einer klar binären Ja/Nein-, Like/Dislike- oder On/Off-Logik verbleiben müssen. Turings für Maschinen begrenzte und zweckgebundene Lösung des Entscheidungsproblems hat durch die digitale Verbreitung computergesteuerter Prozesse zu binären Denkweisen und einer Mathematisierung der Wahrnehmung in Erkenntnisprozessen geführt. Seine Lösung beinhaltet zugleich die Vorstellung, dass komplexe Prozesse wie das menschliche Denken selbst einer algorithmischen Prozessualität folgen.

Turing war sich darüber bewusst, dass nicht alle Entscheidungen beziehungsweise Fragen so formuliert werden können, dass sie rechnerisch zu vermitteln sind. Gleichzeitig stand für ihn fest, dass alle Fragen, die mit ja/nein beantwortet werden können, in einen mathematischen Algorithmus gebracht und in Reihe geschaltet werden können, sodass man einen Entscheidungsprozess simulieren kann. Diese auf zwei mögliche Antworten gebrachte Vereinfachung menschlicher Entscheidungsprozesse wurde kurz darauf in kybernetischen und konnektionistischen Theorien der Kognition weiterverwendet; schon bald ergänzt um die Konzepte der Wahrscheinlichkeitsrechnung, also durch die Einberechnung mathematischer Zufälle.

2.2 Elektrotechnisches Neuronenmodell – McCulloch und Pitts

Die Funktionsweise des Gehirns mithilfe Neuronaler Netzwerke organisch zu erklären, fand zunächst, wie oben aufgezeigt, in der Physiologie und der Anatomie, namentlich bei Exner und Cajal, statt. Nachdem diese Entwürfe zunächst überwiegend ohne weitere wissenschaftliche Beachtung blieben, gab

es 1943 den ersten Vorstoß, neuronale Geflechte als elektrotechnische Reihenschaltungen zu modellieren, was, aus heutiger Sicht, die Geburt künstlicher Neuronaler Netze darstellt. Initiatoren dieses neuen Neuronenmodells, das Neuronen als Schaltelement begreift, sind der Neurophysiologe Warren McCulloch (1898–1969) und der Mathematiker, später Neuroinformatiker, Walter Pitts (1923–1969), als Schüler Carnaps spezialisiert auf die Mathematische Logik. McCulloch und Pitts entspinnen in ihrem 1943 veröffentlichten Text *A logical calculus of ideas immanent in nervous activity* ein an die Turing-Maschine angelehntes Modell, das Aussagenlogik und mathematische Formalisierung zum Verständnis neuronaler und mentaler Aktivität heranzieht. In ihrem Text verbinden die beiden biochemische und elektrophysikalische Überlegungen und stellen das ›Netz‹ als eine Möglichkeit vor, in sich abgeschlossene synaptische Elemente hintereinander zu schalten, um ihre Aktivität auf diese Weise berechnen zu können.

In ihrem Artikel liefern McCulloch und Pitts ein Modell, mit dem synaptische Feuerungsraten, zusammengefasst als charakterisierbare Knotenpunkte, in eine Reihenschaltung übertragen werden können. Eine dieser charakteristischen Eigenschaften des Neurons ist das Alles-oder-nichts-Prinzip, das in diesem ersten Neuronenmodell die wesentlichste Vorannahme darstellt. Das Prinzip beschreibt das Verhalten einer einzelnen Nervenfaser, die bei Stimulation entweder eine maximale Antwort mit der gleichen Schwingungsweite produziert oder gar keine. Auch wenn die Intensität des Reizes erhöht wird, bleibt der Ausschlag, so heißt die synaptische Reaktion, gleich. Der sich ausbreitende Impuls in einer einzelnen Faser kann nicht durch Abstufung der Intensität oder durch die Dauer des Reizes abgeschwächt werden. Aufgrund des Alles-oder-nichts-Charakters der Nervenaktivität können, so McCulloch und Pitts, neuronale Ereignisse und die Beziehungen zwischen ihnen mithilfe der Aussagenlogik berechnet werden (1943, 99). Mit der Alles-oder-nichts-Charakterisierung konnte neuronale Aktivität als klar abgegrenztes Ereignis in dem jeweiligen Knotenpunkt schematisiert werden und in eine regelkonforme, hintereinander geschaltete Rechenfolge gleichmäßiger Logik gesetzt werden. Allein mit der Behauptung, dass sich Netze auf diese festgelegte Art und Weise verhalten, wurde die Annahme auf das Verhalten aller Netze übertragen:

It is found that the behavior of every net can be described in these terms, with the addition of more complicated logical means for nets containing circles; and that for any logical expression satisfying certain conditions, one

can find a net behaving in the fashion it describes. It is shown that many particular choices among possible neurophysiological assumptions are equivalent, in the sense that for every net behaving under one assumption, there exists another net which behaves under the other and gives the same results, although perhaps not in the same time. (Ebd.)

McCulloch und Pitts gehen aber noch einen Schritt weiter und stellen mit ihrem ersten Theorem eine weitreichende Hypothese Neuronaler Netzwerke auf: »Every net of order 0 can be solved in terms of temporal propositional expressions.« (Ebd., 103) Sie verallgemeinern diese Behauptung in ihrem zweiten Theorem: »Every TPE [temporal propositional expression] is realizable by a net of order zero.« (Ebd., 104) TPE verweist auf Theorem 1 und aus der behaupteten Formel »Jedes Netz mit der Ordnung 0 kann durch die Aussagenlogik gelöst werden« wird wiederum die Formel abgeleitet, dass, wenn Theorem 1 wahr ist, die Aussagenlogik durch ein Netz der Ordnung 0 realisiert werden kann. Das Alles-oder-nichts-Prinzip lässt, wie die Aussagenlogik, nur zwei Antworten (wahr oder falsch) zu und verweist damit auf die Arbeiten von Alan Turing, der mit der Implementierung eines Entscheidungsalgorithmus in eine Maschine auch für die weitere Theoretisierung logischer Operationen des Geistes wegweisend war. Das Entscheidungsproblem bezieht sich auf die Schwierigkeit, Entscheidungen so zu identifizieren und klar zu definieren, dass sie maschinell berechnet werden können. Alles oder nichts, ja oder nein, like or dislike, 1 oder 0. Mit diesen Annahmen ließ sich neuronale Aktivität so formalisieren, dass sie in Analogie zu den logischen Operationen einer Rechenmaschine erfolgt.

Durch die klare Definition der immanenten Eigenschaften der hintereinander geschalteten Einheiten im Nervennetz ergibt sich ein notwendiger Zusammenhang, aus dem sich durch die Beschreibung eines beliebigen Zustands der nachfolgende Zustand berechnen lässt. Allerdings kann dieser Prozess nur in eine Richtung vorgenommen werden, da die Einbeziehung disjunktiver Beziehungen eine vollständige Bestimmung des vorherigen Zustands verhindert. »To compute« beschreibt hier nicht den Rekurs auf den Computer, denn den gab es zu der Zeit so noch nicht, sondern auf das elektrotechnische Rechnen, das sich das Netz als Schaltkreis imaginiert, bestehend aus singulären, abgegrenzten Knotenpunkten, die als Events berechnet werden können und die am Ende addiert werden. Neben dem Alles-oder-nichts-Prinzip entscheidet ein weiteres Merkmal über die im jeweiligen Knotenpunkt angestellte Berechnung. Neben der Aktivierung von Feuerungsraten

und deren Weitergabe kann die Synapse auch die Weitergabe von Aktivität hemmen. Ausgehend von der Funktionsweise des einzelnen Neurons, wurde das Denken als ein System von Befehl und Kontrolle modelliert. Der Text greift hierfür auf einen epistemologischen Umkehrschluss zurück, der der Theorie einen immensen Einfluss verschaffte und gänzlich neue Disziplinen wie den Konnektionismus und die Kognitionswissenschaften erst ermöglichte. Er hatte auch großen Einfluss auf die Entstehung der Informatik und die Erforschung künstlicher Intelligenz. In der Geschichte der Übersetzung physikalischer und biochemischer Prozesse mittels mathematisch-mechanistischer Formalisierungen wurde durch McCullochs und Pitts' Artikel ein weiterer Meilenstein erreicht, der wies aber gleichzeitig an einigen Punkten über die reine Anwendung Mathematischer Logik hinaus.

Die von McCulloch und Pitts vorgenommene Beschreibung des Neurons und ihre vorgeschlagenen Berechnungsweisen bestachen nicht durch ihre Komplexität, sondern waren bereits zu dem Zeitpunkt, an dem sie angestellt wurden, stark vereinfacht. Sie fielen hinter die bekannten Eigenschaften von Neuronen und Neuronalen Netzen, wie sie von Neurophysiologen schon vor den 1940er-Jahren beschrieben worden waren, zurück. Sie waren nur sehr vordergründig »neural« in dem Sinne, dass die An- und Auswerte ihrer Einheiten durch den Alles-oder-nichts-Charakter neuronaler Aktivität inspiriert waren. Heute sind die angenommenen neuronalen und synaptischen Eigenschaften längst überholt. Das All-or-none-Prinzip wurde als viel zu ungenau verworfen, auch konnten neuere Messmöglichkeiten zeigen, dass manche Hemmungen weniger als eine Millisekunde benötigen. Generell lässt sich aber sagen, dass diese Ungenauigkeiten der Idee einer Mechanik des Geistes keinerlei Abbruch getan haben. In der Neurophysiologie hatte das McCulloch-Pitts-Modell zunächst ohnehin kaum bis gar keinen Eindruck hinterlassen. Es waren eben jene Verallgemeinerungen und Ungenauigkeiten, die es der empirischen Neurophysiologie verunmöglichte, mit dem Modell zu arbeiten, die aber dem Grundgedanken einer auf Aussagenlogik regelhaft hintereinander angeordneten Reihenschaltung zu einem Siegeszug sondergleichen verhalfen.

Die Theorie von McCulloch und Pitts fand sein Publikum bei Mathematikern wie Norbert Wiener und John von Neumann, die sich für Erkenntnistheorie und Logik interessierten, aber eher in Mathematik oder Ingenieurwesen als in Neurophysiologie ausgebildet waren. Zum einen gefiel ihnen die Behauptung, dass der Geist auf das Gehirn reduziert werden könne; einige ihrer intellektuellen Erben sehen noch heute die Lösung des Geist-Körper-

Problems als den großen Beitrag von McCulloch und Pitts an. Zum anderen gefiel ihnen der operationale Geist der Theorie, nach dem das Design des mathematischen Netzes als Ausgangspunkt für alles gesehen wurde, und zwar zur Durchführung von Inferenzen und als Beschreibung mentaler Phänomene. Aus dem Versuch, mit mathematischen Regeln die Verarbeitungsprozesse des Gehirns zu erklären, wurde im Umkehrschluss behauptet, dass das Gehirn Prozesse rechnend verarbeitet. Der Kreis des Arguments schließt sich hier: Viele der sehr vereinfachenden Annahmen über Neuronen und neuronale Netzwerke wurden vorgenommen, um das Geist-Gehirn-Problem zu lösen, indem sie Mathematische Logik und Turingmaschinen als Beschreibungen des Nervensystems verwendeten. Aber mit dem Erfolg dieser Theorie wurde sie als Beweis dafür herangezogen, dass das Gehirn tatsächlich auf Rechenprozessen beruht. Die mathematischen Werkzeuge wurden seither ausgefeilter, modifiziert und differenziert, aber der Grundgedanke bleibt: Der Geist sollte als rechnender Mechanismus erklärt werden, jeder mentale Prozess wurde zu einem Rechenprozess und jedes Verhalten zur Ausgabe einer Berechnung.

Dass dieser »computationale Ansatz der Erklärung des menschlichen Geistes [...] nicht auf der Methode strenger, in sich geschlossener logischer Deduktion [gründet], sondern sich in erster Linie auf Metaphern und Analogien« (Greif 2014, 29) stützt und keinerlei überprüfbare Vorhersagen oder Erklärungen für beobachtbare neuronale Aktivität bereithielt, tat seinem Erfolg keinen Abbruch. Die Ansichten von McCulloch und Pitts, dass neuronale Netze Berechnungen durchführen (im Sinne der Berechenbarkeitstheorie) und dass neuronale Berechnungen mentale Phänomene erklären, setzten sich durch und wurden zur Mainstreamtheorie über Gehirn und Geist. Es ist also nicht mehr nur so, dass das formal-logische Neuronenmodell Modelle der Kognition inspirierte, sondern in einer kontinuierlichen Dialektik zwischen dem Nervensystem und der Verhaltensmaschine wurden der Geist und der Computer zu Modellen füreinander.

2.3 Mechanisierung des Geistes – Kybernetik – Norbert Wiener und Marvin Minsky

McCullochs und Pitts' wissenschaftlicher Werdegang war eng mit der in den 1940er- und 1950er-Jahren sich etablierenden Kybernetik verknüpft. Mit einem Namen ist die Kybernetik besonders verbunden: mit ihrem Namensgeber Norbert Wiener (1894–1964) und dessen 1948 erstmals erschienenen Werk

Kybernetik. Regelung und Nachrichtenübertragung im Lebewesen und in der Maschine. Die Kybernetik ist aus der Nachrichtentechnik und den Errungenschaften, die sie zum Großteil während des Zweiten Weltkriegs auf dem Gebiet der Kryptologie generieren konnte, hervorgegangen. Sie ging aber schnell weit über die Nachrichtentechnik und die Frage der Nachrichtenübermittlung hinaus und wurde zu einer Informationstheorie, die sich mit der Informationsbe- und -verarbeitung beschäftigte. Informationsübertragung in Organismen, so die Annahme, liegt nie ohne Informationsverarbeitung im Empfänger vor und die Übertragung erfolgt nur in den seltensten Fällen in nur eine Richtung. Das heißt, es brauchte für die Kommunikation von Information mehr als nur einen Sender und einen Empfänger, zumindest einen bidirektionalen Nachrichtenkanal, sodass ein Sender immer auch ein Empfänger war und umgekehrt. Alle Naturvorgänge und alle Organismen wurden von nun an von ihrem Aspekt her betrachtet, Information zu sein. Die Kybernetik ist die Wissenschaft der Steuerung und Regelung von Maschinen und ihren Prozessen. Gleichzeitig begründet sie die vormals bereits angelegte, aber nun theoretisch plausibilisierte Übertragung maschineller Prozesse als Erklärung auch von lebenden Organismen und sozialen Strukturen wie Organisationsweisen und geht hier von einer engen Verzahnung von Mensch und Maschine aus. Aus heutiger Sicht stellt die Kybernetik den theoretischen Vorläufer für die Durchsetzung der Mathematischen Logik in Erkenntnisprozessen dar und bereitete damit den Boden für den enormen Erfolg von vermeintlich intelligenten Rechenmaschinen.

Norbert Wiener selbst war sich der erkenntnistheoretischen Fragen, die mit seiner neu formierten Wissenschaft verbunden wurden, sehr wohl bewusst. Das Umdeuten des Begriffs der Information als inhaltsleere Sequenz, rein auf die Frage ausgerichtet, ob etwas ankommt, bedeutete, dass die vom Informationsbegriff abhängige Vorstellung von Kommunikation ebenfalls einer Neudefinition unterlag. Künstliche Intelligenz führte uns demnach nicht in ein Informationszeitalter, sondern in ein Kommunikationszeitalter, in dem der Inhalt der Kommunikation auf stochastische, mathematische Prozesse ausgerichtet wurde und nicht mehr dem Austausch im weitesten Sinne dient. Die Kybernetik führte mathematische, regelorientierte und biologische sowie geisteswissenschaftliche Theorien zusammen und überführte ein neues Wertesystem wie die künstliche Intelligenz und die instrumentelle, mathematische Vernunft aus dem Industriezeitalter ins Informationszeitalter, genauer bestimmt als Computerzeitalter. Orit Halpern arbeitet in ihrem Buch *Beautiful Data* die Auswirkungen der Kybernetik und der Kommunikations-

wissenschaften nach dem Zweiten Weltkrieg für die Sozial- und Geisteswissenschaften heraus und zeichnet hierfür den radikalen Wandel im Verständnis, der Aufzeichnung und Verwendung von Daten nach, die neue Formen der Beobachtung und Überwachung, der Rationalität und der Ökonomie hervorgebracht haben und sich im Anschluss an Halpern als »Kommunikative Objektivität« (Halpern 2014, 1) fassen lassen. Diese historischen Veränderungen, wie wir Wahrnehmung steuern und schulen und wie wir Vernunft und Intelligenz definieren, zeigen sich, so Halpern, auch in einer Veränderung technisch induzierter Regierungs- und Steuerungsweisen von Individuen.

Die Kybernetik wartete mit neuen und bis heute wirkmächtigen Begrifflichkeiten auf, wie Black Box, Information, Prozess, Feedback, Kausalität, Antwortzeit und Zeitlichkeit. Nicht mehr die Essenz, im Falle der Neuronenmodelle die des Gehirnorgans, begründet ontologische Überlegungen, sondern der Prozess und die Prozessverarbeitung werden essenziell.

In this new [cybernetic] imaginary, response time was the critical feature driving the system, and this produced a new understanding of teleology that was not about progress, linearity, or conscious humanist effort, but rather a new mode of technical thought. [...] Cyberneticians concerned themselves with process, not essence. (Ebd., 44)

Im Kontrast zu ihren weitreichenden Resultaten gab es die Kybernetik als wissenschaftliche Disziplin nicht wirklich lang – grob von den 1940ern bis in die 1960er-Jahre – aber die aus ihr hervorgegangenen Anstöße finden sich in anderen Bereichen wieder, wie etwa der Informatik, heute der Robotik und der künstlichen Intelligenzforschung.

Die Kybernetik Norbert Wieners baut auf den Arbeiten des Mathematikers und Elektrotechnikers Claude Shannon (1916–2001) und des Mathematikers und Logikers Alan Turing (1912–1954) auf. Shannon begründete eine mathematische Kommunikationstheorie und beschreibt in *A Mathematical Theory of Communication* eine Anwendung für die klassische boolsche Algebra auf die Untersuchung der Aussagenlogik mithilfe von elektrotechnischen Schaltsystemen. Turing wiederum ist »vielleicht der erste, der die logischen Möglichkeiten der Maschine als intellektuelles Experiment untersucht hat [...]« (ebd., 41). Mit Claude Shannons mathematischer Informationstheorie und Norbert Wieners Postulat »Information is information, not matter or energy« war 1948 eine dritte physikalische Untersuchungskategorie neben Energie und Materie in die Physik getreten, und Information wurde mathematisch berechenbar. Der Erlanger Physiologe und Kybernetiker Wolf-Dieter Keidel definierte In-

formation nach Shannon als »ein quantitatives Maß von Unwahrscheinlichkeit einer gewissen Ereignisfolge« (Keidel, zit. n. Aumann 2009, 34). Information wurde quantifizierbar und »besaß eine eigene Maßeinheit, so dass Informationsübermittlung und Codierung naturwissenschaftlich-technisch beschreibbar waren. Information war nicht mehr allein Gegenstand der Geisteswelt und der Technik, sondern auch eine naturwissenschaftliche Größe geworden.« (Ebd., 35) Im nächsten Schritt wurde Information fragmentiert und in zwei Ebenen übersetzt. Unterschieden wird zwischen »dem Signal als materiellem Träger der Nachricht und dem Zeichen als Träger der Bedeutung, dem Sender und dem Empfänger, zwischen denen Information fließt, sowie deren gemeinsamem Zeichenvorrat. Diese Elemente bestimmten wiederum die Informationskapazität des Systems, die berechenbar und damit für die Technik manipulierbar und optimierbar seien.« (Ebd.)

Wiener beschreibt zunächst die Probleme der Regeltechnik und der Nachrichtentechnik als untrennbar und hebt vorvor,

daß sie sich nicht auf die Elektrotechnik konzentrierten, sondern auf den fundamentaleren Begriff der Nachricht, ob diese nun durch elektrische, mechanische oder nervliche Mittel übertragen wird. Die Nachricht ist eine zeitlich diskret oder stetig verteilte Folge meßbarer Ereignisse – genau das, was von den Statistikern ein Zufallsprozeß genannt wird. Die Vorhersage der Zukunft einer Nachricht geschieht durch irgendeine Operation auf ihre Vergangenheit, gleichgültig, ob dieser Operator durch ein mathematisches Rechenschema oder durch einen mechanischen oder elektrischen Apparat verwirklicht wird. (1992, 35)

Der empirisch-analytische Auftrag änderte sich durch die Kybernetik. Wiener spricht es bereits an: Die Qualität eines Modells beruht sich nicht länger auf die möglichst adäquate Beschreibung der Übertragung beziehungsweise des Prozesses, sondern auf die möglichst genaue Bestimmung eines zukünftigen Verhaltens. Schafft man es aufgrund der Berechnungen eines Modells, möglichst genaue Vorhersagen eintretender Ereignisse zu treffen, so die Idee, ist die Qualität der geschalteten Entscheidungsmodalitäten bewiesen. Wiener erkannte, dass die »optimale Vorhersage nur durch eine Verarbeitung der Statistik der Zufallsprozesse, die vorhergesagt werden sollen, zu erhalten war [...]« (ebd., 36). Dies einmal erfasst, war es nicht schwer, die ursprüngliche Schwierigkeit in der Theorie zu etwas zu machen, was tatsächlich ein wirk-sames Werkzeug zur Lösung des Problems der Prognose war. Das Problem der optimalen Vorhersage wurde in die Anordnung eines ganz bestimmten

Operators überführt, mit deren Hilfe die minimalsten Werte einer Funktion errechnet werden können. Dieser Operator ist die Variationsrechnung, die sich bereits im 18. Jahrhundert aus der Analyse mechanischer Systeme entwickelt und von Leonard Euler (1707–1783), basierend auf einem Rechenverfahren, das Jakob Bernoulli (1655–1705) zur Lösung eines Problems vorschlug, als Variationsrechnung bezeichnet wurde. Bernoulli hatte mit der Variationsrechnung den kürzesten Weg einer Kurve zwischen den Punkten A und B berechnet. Das Rechenverfahren der Variationsrechnung nähert sich durch eine unendlich dimensionale Variation (daher der Name Variationsrechnung) und Optimierung von »Argumenten«, bis das gesuchte Integral das minimalste ist.

Nachdem die Nachrichtentechnik zu einer statistischen Wissenschaft umfunktioniert wurde, brauchte es auch eine »statistische Theorie des Informationsgehalts« (ebd., 37). Die statistische Theorie des Informationsgehalts sieht vor, dass die »Einheit der Information diejenige ist, die bei einer Entscheidung zwischen zwei gleichwahrscheinlichen, einfachen Alternativen übertragen wird« (ebd., 38). Den Ablauf einer idealen Rechenmaschine beschreibt der Kybernetiker als abgeschlossenes System, in dem alle Informationen am Anfang eingegeben sein müssen und das »bis ganz zum Schluss von menschlicher Einwirkung so frei wie möglich sein« (ebd., 174) sollte. Nicht nur alle numerischen Daten müssen am Anfang eingegeben sein, sondern auch alle ihre »Verknüpfungsregeln [...] in der Form von Anweisungen, die jede Situation einschließen, die sich im Laufe der Rechnung ergibt« (ebd.). Um dies zu programmieren, wird auf die Aussagenlogik der booleschen Algebra zurückgegriffen:

Während es viele Algorithmen gibt, die dazu benutzt werden können, Möglichkeiten zu kombinieren, ist ihre einfachste als die logische Algebra par excellence oder als Boolesche Algebra bekannt. Dieser Algorithmus ist wie die binäre Arithmetik auf der Dichotomie zwischen Ja und Nein begründet, auf der Wahl innerhalb oder außerhalb einer Klasse zu liegen. (Ebd.)

Das heißt, jede Möglichkeit, die in die Rechenmaschine eingespeist wird, muss mit 0 oder 1 beziehungsweise mit ja oder nein beantwortbar sein. Daraus ergibt sich die Struktur einer Reihenschaltung der Maschine, »jene einer Reihe von Relais, die in der Lage sind, zwei Stellungen, z.B. ›Ein‹ und ›Aus‹ anzunehmen, wobei in jedem Zustand der Rechnung jedes Relais eine Stellung einnimmt, die von den Stellungen eines oder aller Relais der Reihe in einem früheren Zustand der Operation bestimmt wird« (ebd., 175). Die Quintessenz aus diesen Überlegungen bedeutet laut Wiener, dass »[d]ie Übertragung von

Information [...] nur als eine Übertragung von Alternativen möglich« (ebd., 37) ist.

Um eine neue Theorie solchen Ausmaßes aufzustellen, rekurriert die Kybernetik auf Leibniz und dessen Suche nach einer universalen symbolischen Logik, die sich gleichzeitig in den Abläufen geistiger Prozesse wiederfinden lässt. Wiener schlussfolgert: »Es ist deshalb nicht im mindesten überraschend, daß der gleiche intellektuelle Impuls, der zur Entwicklung der Mathematischen Logik geführt hat, gleichzeitig zur idealen oder tatsächlichen Mechanisierung der Prozesse des Denkens geführt hat.« (Ebd.)

Weitere Aspekte aus der Mathematischen Logik, derer sich die Kybernetik bedient, sind die boolesche Algebraisierung der Logik und die Trennung von Semiotik und Semantik in ihrer formal-symbolisch angelegten Sprache. Gleichzeitig schwingt Exners Gedanke mit, dass die Verschaltungsarchitektur der Synapsen und Neuronen über ihrer Funktionsweise stehen. Die Suche nach dem richtigen Aufbau informationsverarbeitender Schaltkreise ist in der Kybernetik zentral, hängt hier doch die Informationskapazität eines Systems »ausschließlich von der geschickten Verteilung der Informationseinheiten« (Meyer-Eppler 1959, 344) ab. Gleichzeitig werden alle biologischen Prozesse und Bedingungen formalisiert, sodass sie von nun an in ihrer Prozesshaftigkeit außen vorgelassen werden konnten und nur mehr die Angabe ihres numerischen Werts von Bedeutung ist – etwa die Neuronentätigkeit, die sich mit einer 0 oder 1 angeben lässt, später in den Computational Neurosciences sind es detailliertere Angaben wie die Membrandurchlässigkeit, die Ionenverteilung etc. Diese Möglichkeit, biologische Prozesse in Zahlen anzugeben, führt zu einem Übertragungswettlauf in Rechenmaschinen, die in ihrer Gleichsetzung als Neuronale Netzwerke schon bald analog zu den morphologischen Vorgängen im Gehirn behandelt werden. Wiener schreibt hierzu, rekurrierend auf die Arbeiten von McCulloch und Pitts:

Von dieser Zeit an wurde es uns klar, daß die ultraschnelle Rechenmaschine, so wie sie abhängig war von aufeinanderfolgenden Schaltern, beinahe ein ideales Modell der sich aus dem Nervensystem ergebenden Probleme darstellen mußte. Der Alles-oder-nichts-Charakter der Neuronenentladung ist völlig analog zur Auswahl einer binären Ziffer; und schon mehr als einer von uns hatte das binäre Zahlensystem als beste Basis des Rechnens in der Maschine erkannt. Die Synapse ist nichts als ein Mechanismus, der bestimmt, ob eine gewisse Kombination von Ausgängen von anderen Elementen ein ausreichender Anreiz für das Entladen des nächsten Elementes

ist oder nicht und muß ein genaues Analogon in der Rechenmaschine haben. (Wiener 1992, 42)

Die weiter oben bereits angesprochene Alternative, die zur Übertragung von Information und Informationsgehalt benötigt wird, liegt im Finden universeller Symbole, also einer ganzen Sprache, in der sich auch der mathematische Beweis ausdrückt, »einer, der in einer endlichen Anzahl von Symbolen geschrieben werden kann« (ebd.). Das mathematische Kalkül wird in Wieners Überlegungen über maschinelle und organische Prozesse zum Beweis. Das ist deswegen interessant, da die Mathematik als deduktive Methode bekannt ist, hier aber durch das Hinzuziehen von Statistik und der Wahrscheinlichkeitsrechnung zu einer induktiven Methodik wird. Ihren Vorteil, objektive Aussagen zu treffen, schöpft die Mathematik aus der klar festgelegten Abfolge spezifischer Schritte, des mathematischen Beweises. Diese ist nun geprägt durch Hilberts eingeführte Selbstbezüglichkeit von Axiomen, deren Beurteilung als wahr allein durch ihre Charakteristik herröhrt, dass die Axiome sich nicht selbst widersprechen, und die ich ausführlich in Kapitel 1 beschrieben habe. Der mathematische Beweis aber folgt nun ähnlichen Gesetzen: Seine Validität erhält ein mathematisches Modell dadurch, dass es eine möglichst exakte Vorhersage trifft. Worüber Aussagen getroffen werden, zu welchem Zweck und ob es sich hierbei um ›wahre‹ Aussagen handelt oder einfach nur um möglicherweise unter bestimmten Bedingungen wahrscheinlichste Ereignisse, die eintreten, darüber kann keine Aussage mehr getroffen werden.

In seinem Buch *An Introduction to Probability and Inductive Logic* warnt der Mathematiker und Wissenschaftshistoriker Ian Hacking davor, nicht zu vergessen, dass die Begriffe ›wahr‹ und ›valide‹ nicht bedeutungsgleich sind: »Be careful about true and valid. In logic propositions are true or false. Arguments are valid or invalid. Valid arguments are risk-free arguments.« (Hacking 2001, 4f.) Und er ergänzt: »Inductive logic is about risky arguments. It analyses inductive arguments using probability. There are other kinds of risky arguments. There is inference to the best explanation, and there are arguments that are based on testimony.« (Ebd., 11) Diese wichtige Unterscheidung, die durch die Implementierung statistischer und stochastischer Paradigmen in die mathematischen Wirklichkeitsräume durcheinandergewirbelt wurde, wird uns im Weiteren noch beschäftigen.

Ich durchbreche an dieser Stelle die Chronologie und springe zeitlich zu einer Theorie, die erst 1986 veröffentlicht wurde, die aber direkt an McCulloch und Pitts und die Kybernetik anknüpft und sich ebenfalls der harten künst-

lichen Intelligenz zurechnen lässt. Die Theorie Marvin Minskys, auf die ich nun eingehen werde, beschreibt den Geist und das Gehirn als das Produkt einer Gesellschaft von konkurrierenden Agenten.

Die Netzmetapher war mit McCullochs und Pitts' Artikel ins Leben getreten, weiterentwickelt und radikal zugespitzt findet sie sich bei Marvin Minsky (1927–2016) wieder. Minsky fasst in *Society of Mind* (1986) sein Verständnis vom Gehirn zusammen, indem er einerseits die langjährigen Debatten über den Zusammenhang von Leib und Seele auf die praktische Formel bringt, dass der Geist ganz einfach das ist, was das Gehirn tut. Andererseits übersetzt er den Geist allein mit ›Intelligenz‹, sodass das Gehirn bei Minsky sich als verwobenes Netz von unintelligenten Agenten, heißt simpler Prozesse, beschreiben lässt, das allein durch die Zusammenarbeit und die Art ihrer Vernetzung Intelligenz hervorbringt und das, was den menschlichen Geist ausmacht, entwickelt. *Society of Mind* wird heute eher als erkenntnistheoretischer Beitrag zur Frage des Verhältnisses von Gehirn und Geist bewertet und nicht als Grundlagentext der Informationstechnik. Das sollte aber nicht darüber hinwegtäuschen, dass Minskys Erfolg und viele seiner Anwendungen vor allem in den Computer Sciences Anwendung fanden. Bereits in den 1950er-Jahren entwickelte Minsky gemeinsam mit dem Mathematiker Claude Shannon und anderen² den Begriff der künstlichen Intelligenz. Minsky begründete bereits vor der Veröffentlichung seines Buches das Labor für Künstliche Intelligenz am Massachusetts Institute of Technology (*AI Lab*) und wird als Verfechter einer starken KI gewürdigt, einer Stoßrichtung, die wichtig für die Kognitionswissenschaft und den Konnektionismus war. Starke und schwache KI bezeichnet unterschiedliche Einschätzungen über die Tragweite künstlicher Intelligenzen. Schwache KI gilt als begrenzt in ihren Möglichkeiten, hier wird der Endpunkt künstlicher Intelligenz im Bau einer Maschine gesehen, die in der Lage ist, menschliche Fähigkeiten zu entwickeln und den Menschen bei einer bestimmten Aufgabe, die Intelligenz erfordert, erreicht oder sogar übertrifft. Vertreter der Idee einer starken KI gehen deutlich über diese Einschätzung hinaus, sie sind überzeugt »that thinking machines will eventually be minds, or at least that they will have all the hallmarks of minds, such as consciousness« (Walsh 2018, 126).

In seinem Buch *Society of Mind* offeriert Marvin Minsky ein in seiner Einfachheit bestechendes Konzept, das sich durch die Fragmentierung neurona-

² John McCarthy oder Nathaniel Rochester; als Ursprung des Wortes wird meist die Dartmouth Conference 1956 genannt.

ler Abläufe in kleine, abgeschlossene Einheiten auszeichnet und das erst in der Art seiner Zusammensetzung seine Form und Funktionsfähigkeit erhält. Dabei geht er nach dem Baukastenprinzip vor, dessen Bausteine als unintelligente Miniaturprozesse zu verstehen sind. Die Blocks (Minsky 1986, 21), aus denen jedes Gehirn zusammengesetzt ist, nennt er Agenten: »Each mental agent by itself can only do some simple thing that needs no mind or thought at all. Yet when we join these agents in societies – in certain very special ways – this leads to true intelligence.« (Ebd., 17) Diese Fragmentierung in kleinste Bestandteile und Systeme als Ausgangspunkt für seine Überlegungen in *Society of Mind* wird von ihm selbst in die neuesten Entdeckungen seiner Zeit eingereiht. In dieser Zeit werden in der Physik, der Biologie und der Psychologie, aufbauend auf neuen Erkenntnissen aus den empirischen Vermessungsdaten, wichtige neue Theorien über die Ursprünge des Lebens aufgestellt. Gleichzeitig werden diese neuen empirischen Vermessungsdaten in die als mechanisch imaginierten, biologischen, physiologischen und mentalen Modelle eingewoben. Diese Verknüpfung bringt die erste Welle artifizeller Neuronaler Netze und künstliche Intelligenzforschung hervor,

als Warren McCulloch und Walter Pitts zu zeigen begannen, wie Maschinen dazu gebracht werden könnten, zu sehen, zu denken und sich zu erinnern. Die Forschung in der modernen Wissenschaft der Künstlichen Intelligenz begann erst in den 1950er-Jahren, angeregt durch die Erfindung moderner Computer. Dies inspirierte eine Flut von neuen Ideen, wie Maschinen das tun könnten, was zuvor nur der Verstand getan hatte. (Ebd., 19)

Mit diesem Konzept prozessgesteuerter Agenten als Grundausstattung des Gehirns wird der Denkapparat bei Minsky zu einer Fleischmaschine (*meat machine*), mit der alle bisher unbeantwortbaren Fragen rund um das menschliche Gehirn und Bewusstsein nicht nur neu gestellt, sondern schlüssig beantwortet werden können. Für Minsky lässt sich durch die von ihm vorgenommene Klärung der Funktionsweise der Agenten an weiterführende soziale wie philosophische Fragen über das Verhalten des Menschen anknüpfen. Mit der allgemeinen Beschreibung der Beschaffenheit dieser Agenten klären sich, fast automatisch, Fragen der Beziehung von Gehirn und Geist (*Embodiment: What are they made of?*), der Kommunikationsweise der Neuronen (*Interaction: How do they communicate?*), des Verhältnisses von Biologie und Evolution (*Origins: Where do the first agents come from?*), Fragen von Intelligenz (*Heredity: Are we all born with the same agents?*), des Lernens (*Learning: How do we make new agents and change old ones?*), Bedeutungszuschreibung und Sinnstiftung (*Meaning: How*

could they understand anything?); weitere sind: sinnliche Empfänglichkeit (Sensitivity: How could they have feelings and emotions?), Bewusstsein (Awareness: How could they be conscious or self-aware?) (vgl. Minsky 1988, 18). Damit sind nur einige der von Minsky aufgeworfenen vorgestellt, die in seinem Buch allerdings nicht näher inhaltlich beantwortet werden. Ihre Beantwortung ergibt sich allein durch das in Aussicht gestellte Modell, bestehend aus dem kleinsten gemeinsamen Nenner – den unintelligenten Agenten –, die erst in ihrem (mathematisch formalisierten) Zusammenwirken Intelligenz ausbilden: »We want to explain intelligence as a combination of simpler things« (ebd., 23), hält Minsky fest, und: »Once we see the mind as a society of agents, each answer will illuminate the rest.« (Ebd., 18) Dieser erkenntnistheoretische Kniff, Prozesse nicht in ihrer Komplexität zu beschreiben, sondern ein Modell zu entwickeln, dessen Funktionsweise von dem Vorhandensein und der Interaktion kleinstmöglicher Agenten ausgeht, deren Beschaffenheit wiederum nicht mehr genauer beschrieben werden muss, erinnert sehr an die Monadentheorie Leibniz', auf die ich bereits in Kapitel 1.1 näher eingegangen bin.

2.4 Hebb und die Rückkopplung

Nach diesem kurzen historischen Vorgriff auf Minsky, der in der Tradition der Kybernetik steht und zu den Befürwortern einer starken KI gezählt werden kann, gehe ich kurz zurück in das Jahr 1949, um eine weitere Theorie zu beleuchten, die eine zentrale Rolle für die Durchsetzung des Konzepts Neuronaler Netzwerke gespielt hat.

Wenige Jahre, nachdem McCulloch und Pitts in ihrem Artikel erstmals ihre Vorstellung von künstlichen Neuronalen Netzwerken als bestmögliche Art, die Kommunikation zwischen Nervenzellen zu beschreiben, darlegten, knüpfte der Kognitionspsychologe Donald Hebb (1904–1985) an das Konzept an und erweiterte dieses maßgeblich um Forschungsergebnisse aus der Psychologie und der Neuropathologie. Im selbstaufgerufenen Kampf gegen die Theorien behavioristischer Zeitgenossen, die sich weigerten, Verhalten mit gedanklichen Assoziationen oder Nervenzellaktivität zu erklären, vertrat Hebb als Schüler von Karl Spencer Lashley (1890–1958) die These, dass menschliches Verhalten nur erklärbar würde, wenn man die Funktionsweise des Gehirns verstünde. In seiner 1949 erschienenen Arbeit *The Organization of Behavior* beschreibt Hebb das Gehirn als Assoziationsprogramm, in dem der synaptische Kontaktbereich den funktionalen Grundbaustein stelle und dass

»lokale Veränderungen der Erregungskopplung von Zellen, die gleichzeitig erregt werden, die Einheit bilden« (Breidbach 1997, 348).

Hebb beschreibt die synaptische Kommunikation so:

Let us assume [...] that the persistence or repetition of a reverberatory activity (or trace) tends to induce lasting cellular changes that add to its stability. The assumption can be precisely stated as follows: When an axon of cell A is near enough to excite a cell B and repeatedly or persistently takes part in firing it, some growth process or metabolic changes take place in one or both cells such that A's efficiency, as one of the cells firing B, is increased. (1949, 62)

Das hebb'sche Modell neurophysiologischer Bedingungen verlangt die räumliche Nähe der durch synaptische Feuerungsraten miteinander in Kontakt tretenden Neuronen und beschreibt, dass je häufiger ein Neuron A gleichzeitig mit Neuron B aktiv ist, umso bevorzugter werden die beiden Neuronen aufeinander reagieren (»what fires together, wires together«). Für seine Theorie geht Hebb von der Verarbeitung sensorischer Reize aus:

Die ersten Schritte in der Aufnahme und Verarbeitung sensorischer Reize, das »Anspringen« der Sinnesorgane und das Verfrachten der etwaigen primären Effekte von diesen peripheren Bereichen des Nervengewebes in untergeordnete Hirnareale, folgen einem festgelegten Programm. Eine Erregung der Sinneszellen im Auge führt zu einer Erregung im Sehnerv, die zu einer genau festgelegten Region im Hirngewebe transportiert wird. Entsprechende Reaktionen sind fest »programmiert«, sie vollziehen sich nach Art eines Reflexes. Was aber passiert daraufhin? Wie erkläre ich, daß ich zwei verschiedene Vasen als Vasen erkenne? Das Sehsystem reagiert hierbei doch auf unterschiedliche Reizkonfigurationen; es kommt aber zu einer Identifikation, in der die beiden unterschiedlichen Objekte beide als Vasen benannt werden. Erst dann werden funktionelle Details benennbar, die diese einander ähnlichen Objekte wieder voneinander abzugrenzen erlauben. Die entsprechenden Programme im Wahrnehmungsgeschehen können hierbei – erfahrungsabhängig – variieren. (Breidbach 1997, 24)

Auf dieser Beschreibung von Assoziationsmodulen aufbauend, formulierte Hebb ein Modell, das plastische Verhaltensmuster als hirnphysiologische Ereignisse erklärbar macht, und gilt damit als Entdecker einer synaptischen Plastizität, welche bis heute die neurophysiologische Grundlage von Lernen und Gedächtnis darstellt.

Diese sogenannte Hebb'sche Regel formuliert einen lokalen Mechanismus, über den die Textur des Nervengewebes – und damit auch seine Funktionsarchitektur – sich aufgrund externer Reizeingaben derart ändert, daß diese neuronale Textur komplexe Erregungsmuster speichern kann. Hebb gewinnt damit das Konzept eines neuronalen Speichers, das den grundlegenden Mechanismus benennt, über den auch Lernprozesse in eine künstliche – parallelverarbeitende – Computerarchitektur eingebunden werden können. (Ebd.)

Neu in Hebb's Verständnis neuroanatomischer Netze war die Einhegung der bereits aus der Lerntheorie bekannten Rückkopplung, auch Feedback genannt. Die Annahme von Rückkopplungsschleifen in der neuronalen Prozessverarbeitung verlangt nach ganz neuen Modellen, die über die sonst in der Physiologie angenommene Einbahnstraße organischer Verarbeitung hinausreichen. Rückkopplung bedeutete nun, dass Prozesse nicht nur eine Richtung kennen, sondern dass bereits getätigte Reaktion auf einen Eingangsreiz auf noch unbestimmte Weise auf die sensorische Information des Inputs zurückwirkt und damit selbst wieder zum Auslöser einer neuen Reaktion wird. Dank des Modells von einer Vielzahl miteinander verbundener, aber voneinander abgegrenzter neuronaler Feedbackschleifen, sogenannter Zell-Ensembles, die für die Verarbeitung spezifischer Assoziationen verantwortlich sind, lassen sich Modelle der Außenwelt konstruieren, mit denen man die Rückwirkung möglicher Reaktionen vorausberechnen kann. Die Zell-Ensembles erklären auch, wie Neuronale Netzwerke entstehen. Denn Hebb's Grundgedanke ist, dass die Aktivierung nachgeschalteter Zellen nicht durch das Feuern einzelner Synapsen zu bewerkstelligen ist, sondern erst wenn mehrere aktive Nervenzellen sie gleichzeitig stimulieren. Durch die Informationsweitergabe über Netzketten, in denen viele unterschiedliche synaptische Erregungen zusammenlaufen, können gezielt bestimmte Gruppen von Nervenzellen in der sensorischen Hirnrinde durch einen äußeren Reiz aktiviert werden. Die so stimulierten Netzverbindungen verfestigen sich mit jedem erneuten Auftreten des Reizes, bis die Verbindungen zwischen den gleichzeitig feuерnden Nervenzellen des Netzes schließlich so stark sind, dass sie einander selbst erregen können, wenn nur ein Teil von ihnen aktiviert wurde. So miteinander verbundene Nervennetze nennt Hebb Zell-Ensembles, die der Erzeugung einer internen Repräsentation eines von außen zugefügten Stimulus dienen.

Wie aufgezeigt, entwickelte Hebb seine Theorie anhand der Verarbeitung visueller Reize im Gehirn. Das ist kein Zufall, sondern dem Umstand experimenteller Settings und der Verknüpfung von Wahrnehmung mit Sehen geschuldet und wird seitdem bis heute zum Haupteinsatzgebiet kognitiver und neurowissenschaftlicher Untersuchungen. Mit dem Fokus auf visuelle, auf Symbolen und Bildern basierende Reizverarbeitung geht die Neurophysiologie eine unbewusste, aber effektive Kooperation mit der Informatik ein. In der Theorie Hebbs liegt der Fokus also auf der Verarbeitung visueller Reize, ein Untersuchungsgegenstand, der die Erklärung, dass assoziative Gedanken zu Repräsentationen im Gehirn führen, naheliegend macht. Ein neuronales Muster für die Verarbeitung von Bildern, etwa einer Vase, können wir uns alle vorstellen. Auch auf das Bild einer Vase können wir uns (fast) alle einigen. Wie aber sieht ein Repräsentationsmuster höherer kognitiver Denkleistungen abstrakter Begrifflichkeiten aus? Wie manifestieren sich Emotionen, Erinnerungen, komplexe Gedanken? Hebb zeigt über die Gestaltpsychologie auf, dass Gedanken/Assoziationen zumindest im Prinzip eine ebenso solide physiologische Grundlage haben können wie Muskelbewegungen. Sie manifestieren sich im Gehirn als erlernte neuronale Aktivitätsmuster, die zunächst durch sensorische Eingangsreize erzeugt werden, dann aber autonomen Status erlangen. Die Effekte dieser engen Verknüpfung von Reizverarbeitung visueller ›Assoziationen‹ und unsere Vorstellung vom Denken als solches zeigen sich in vielerlei Hinsicht und werden auch in den folgenden Kapiteln noch Thema sein. Wichtig für die weitere Erkundung über das Konstrukt der künstlichen Neuronalen Netzwerke ist aber zunächst, dass der Grundgedanke der Rückkopplung geboren war und mit ihr ein Modell von Wachstum und Vernetzungsmodalitäten. Was jetzt noch fehlte, war eine Systematik, die die Ausgewogenheit wiederherstellen konnte und dem Wachstum auch hemmende Einflüsse an die Seite stellte.

Neben der Physiologie waren auch die Psychologie und die Philosophie des Geistes nach Hebbs Modell nicht mehr dieselben. Durch das Aufgreifen von Ansätzen aus der Assoziationspsychologie wird Assoziation nunmehr als Aktivierung von Nervenbahnen beschrieben:

Die Assoziationsregeln fassen sich in der Analyse der Verknüpfungen dieser Nervenbahnen. Erfahrungsbedingte Variationen dieser Bahnen entsprechen den assoziativen Gedächtnisfunktionen. Die Phänomenologie des Denkens entspricht der Mechanik des neuronalen Apparates. Diese Idee bildet denn auch den Kern des Entwurfes der modernen Neurophilosophie.

sophie. Denkprozesse werden als physiologische Ereignisse beschrieben. (Breidbach 1997, 34)

Damit steht Hebb in der Tradition einer Gestaltpsychologie, entwickelte aber aus ihr heraus eine neurophysiologische Perspektive, die dann über die Informatik zurück in die Neurowissenschaften Einzug hielt.

2.5 Cognitive Turn und Konnektionismus

Das McCulloch-Pitts-Projekt beförderte den Konnektionismus und den *cognitive turn* der 1940er-Jahre (vgl. Kay 2001; Breidbach 1997; König/Engel 1998). Ein wichtiger Aspekt dieser kognitiven Wende liegt darin, dass es unter der Hand die Psyche zurück ins Labor und in weitere Disziplinen brachte: in die Neurophysiologie, die den Geist/die Psyche in ihrer empirischen Forschung fast völlig ausblendete; in die Psychologie, die zu dieser Zeit von den Behaviorist*innen dominiert war, welche das Verhalten beobachteten und den Bereich des Mentalen negierten; aber auch in die Psychiatrie, die lange von qualitativen Darstellungen geleitet wurde.

Der Konnektionismus ist die Lehre von Verbindungen und hatte sich zum Ziel gesetzt, aus den Anleihen der Netzidee die Art und Weise zu ergründen, mit der sich darauf aufbauende höhere Kognitionen darstellen und erklären lassen. Das bedeutet, dass die wichtigste Annahme des Konnektionismus darauf beruht, das Verhalten von Systemen als Netzwerk zu verstehen. Er bietet somit Modelle künstlicher Neuronaler Netzwerke zum Verständnis aller komplexeren Systeme wie organische, neuronale, mentale, emotionale und künstliche Intelligenz an. Durch die Behauptung, alles sei Netz, wurde nicht nur das Gehirn in Analogie zum Computer gesetzt, sondern auch der Geist, was wiederum »für die Forschung zu künstlicher Intelligenz und die entstehenden Kognitionswissenschaften zentral wurde« (Lettow 2011, 195). Die Etablierung einer ›Mechanik des Geistes‹ barg neue Implikationen für die immer wiederkehrende Frage nach dem Verhältnis von Körper und Geist:

Within McCulloch's and Pitts' »experimental epistemology,« the mind became a legitimate, even esteemed object of study in the 1940s, '50s, and '60s, an object of quantitative representations that bore novel implications for the age-old queries about the relations between mind and body, the natural vs. the artificial, and for automated military technologies of the postwar era. (Kay 2001, 592)

Die konnektionistische Idee unterscheidet sich von vorherigen Modellen vor allem dadurch, dass nicht mehr Einzelneuronen für die Verarbeitung von Information oder Reizen verantwortlich gemacht werden, sondern Neuronennetzwerke. Diese Netzwerke sind selbstorganisiert, verarbeiten Reize parallel und verteilt – nicht streng hierarchisch –, verändern sich durch Erfahrungen, sind also plastizität (vgl. Engel/König 1998, 164f.). Die Dynamik Neuronaler Netzwerke wird als ihre grundlegende Eigenschaft ausgemacht. Aus den Überlegungen der Verbindungslehre geht die Kognitionswissenschaft hervor. Allgemein gesprochen basiert die Kognitionswissenschaft auf der zentralen Annahme, dass das Gehirn ein informationsverarbeitendes System ist, und wird daher auch als »Computermodell des Geistes« bezeichnet. Grob lassen sich zwei Formalisierungsweisen kognitiver Prozesse im Konnektionismus ausmachen: die symbolische und die konnektionistische künstliche Intelligenz.

Die symbolische künstliche Intelligenz war vor allem von Mitte der 1950er bis Ende der 1980er-Jahre wegweisend und beschreibt alle Methoden, die auf symbolischen, für den Menschen »lesbaren« Darstellungen von Problemen, von Logik und von Suchmechanismen basieren. Sie findet nach wie vor Anwendung, aktuell wieder verstärkt in Bereichen, in denen das Verhalten eines Algorithmus, also die Ergebnisse, überprüfbar und erklärbar sein müssen. Dies ist mit artifiziellen, Neuronalen Netzwerken, sprich mit konnektionistischer KI, meist nicht zu erreichen. Konnektionistische KI ist mit künstlichen Neuronalen Netzwerken – den Algorithmen – umrisse und verlangt selbstorganisierte, selbstlernende Algorithmen, die iterativ vorgehen und etwaige Fehlermeldungen rekursiv rückspeisen, um aus diesen »zu lernen«.

Was in der künstlichen Intelligenzforschung noch kein wirkliches Problem darstellt, da je nach Problem auf die jeweilige Methode zur Gestaltung algorithmischer Lösungsansätze zurückgegriffen werden kann, gestaltet sich für den Konnektionismus und die Kognitionswissenschaften schon etwas schwieriger, da sich die beiden Auffassungen im »Spannungsfeld zwischen formale[m] Modell und nichtformaler Welt« (Kittowski 1992, 71) befinden und einander ausschließen. Das Pendant zu den beiden in der KI unterschiedenen Auffassungen in den Kognitionswissenschaften wird durch das symbolische und das assoziative Konzept beschrieben. Da vor allem die Plastizität des Gehirns und seiner Neuronalen Netzwerke mit dem (formal agierenden und auf Symbolverarbeitung basierenden) Kognitivismus nicht eingefangen werden kann, befördert der nichtformal argumentierende Konnektionismus, der Wahrnehmung als Resultat einer spezifischen neuronalen Architektur

mit sich ändernden Verbindungsstärken zwischen den Neuronalen Netzen versteht, neue Hoffnungen.

Gemeinsam aber ist den theoretischen Ansätzen der Kognitionswissenschaften, des Konnektionismus und der K.I., dass der Begriff der Wahrnehmung insgesamt sehr vage bleibt:

In attempting to formulate a vocabulary of teleological functions, it is important to recognize that the term »cognition« is inherently vague. As was noted, many researchers take this word to mean the processes internal to a brain that culminate in the encoding of memories, planning of action, or directly as immediate actions. However, as far as the electrochemical activity of individual neurons is concerned, the terms »cognition,« »action,« »perception« »volition,« and »emotion« have no distinctive meanings. It is likely that there are no specific markers, in regard to spike trains or patterns of membrane polarization, which discriminates the type of cognitive activity taking place in any particular neuronal assembly. (Turner/De Haan 2014, 191)

Die Kybernetik selbst konnte sich nach den 1960er-Jahren nicht mehr als eigenständige wissenschaftliche Disziplin durchsetzen. Ihre Spielart Konnektionismus wiederum kehrt in verschiedenen Varianten zurück. So wurde er etwa für »Kognitivismus und KI [...] in den 60er- und 70er-Jahren für die Philosophie des Geistes maßgeblich« (Lettow 2011, 196). Als künstliche Intelligenz und die Kognitionswissenschaft Ende der 1980er-Jahre in die Krise gerieten, bekam der Konnektionismus neuen Aufschwung und galt, als falsch verstandene Metapher, als eine »Rückkehr zur Biologie [...] das heißt zu kybernetischen Modellen und zur Theorie der neuronalen Netze« (ebd.). Das aus der Orientierung an der Hirnphysiologie gewonnene Konzept der Neuronalen Netze wurde dann Anfang der 1990er-Jahre in den Neurowissenschaften »wieder hoffähig, schien hier doch eine Theorie gefunden zu sein, die es erlaubte, die funktionelle Plastizität der Einzelzelle und die funktionelle Determiniertheit bestimmter Hirnregionen miteinander zu verbinden« (Breidbach 2001, 39).

Im Jahr 1949 waren Elektroden erstmals in Nervenzellen eingeführt worden, um deren Feuerungsrate abzuleiten. Das bedeutet, dass ab den 1950er-Jahren auch die Neurophysiologie über ein neues Experimentalinstrumentarium verfügte, mit dem die Aktivität einzelner Nervenzellen gemessen und nicht mehr nur vorhergesagt werden konnte. Dem Neurowissenschaftler John Eccles gelangen 1953 die ersten intrazellulären Ableitungen, in dem Fall von

den Motoneuronen des Rückenmarks. »Damit war nun schließlich auch das Instrumentarium des Physiologen in die Dimension hineingewachsen, die dem Neuroanatomen seit dem Ende des 19. Jahrhunderts greifbar war: die Dimension der Nervenzelle.« (Breidbach 1997, 272) Die Möglichkeit, Elektroden in Nervenzellen einzuführen und ganz punktuell Nervenaktivität messen zu können, ist nicht nur wegweisend für die weitere Entwicklung der Neurophysiologie. Die Methoden des Ableitens stellen die grundlegende Messmethode der Computational Neurosciences dar, die aber erst mit dem Zusammenwachsen der Möglichkeiten aus der Kognition und der Informatik zu dem werden konnten, was sie heute sind.

2.6 Neuronenmodell nach Hodgkin, Huxley und Eccles

Die Neuronendoktrin des 20. Jahrhunderts beruhte, trotz Exner und Cajals Exkursen in die Welt der Neuronalen Netzwerke, zunächst auf der Beschäftigung mit einzelnen Neuronen und darin, diese zu charakterisieren, Funktionen zuzuschreiben und zu modellieren.

Die beiden Physiologen Alan Lloyd Hodgkin (1914–1998) und Andrew Fielding Huxley (1917–2012) erforschten die Riesenaxone von Tintenfischen, deren Aktionspotenziale sich dank ihrer Größe von bis zu 1 mm Durchmesser einzeln ableiten ließen. Nachdem Hodgkin und Huxley 1939 die erste intrazelluläre Messung eines Aktionspotenzials vornehmen konnten, arbeiteten sie zwischen 1946 und 1952 intensiv an der Weiterentwicklung ihrer Messmethoden. Für ihre Messungen einzelner Neuronen entwickelte Huxley eine Spiegelvorrichtung, mit der das Axon gleichzeitig von vorne und von der Seite betrachtet werden konnte. Dies ermöglichte die Einführung einer Elektrode, ohne die Zellmembran des Neurons zu beschädigen, sodass Messungen der Ionenströme und des Membranpotenzials durchgeführt werden konnten.

Ihren Durchbruch erzielten Hodgkin und Huxley im Jahr 1952 mit dem von ihnen entwickelten Modell zur Simulation von Neuronen. Das Hodgkin-Huxley-Modell ist das erste quantitative und bis heute meist verwendete Modell zur Beschreibung, wie ein Aktionspotenzial in einem Neuron generiert wird. Es ist vermutlich kein Zufall, dass sich der Erfolg des Modells daran bemisst, dass die Übersetzung der biochemischen Prozesse in die Sprache der Mathematik als besonders gelungen gilt.

Mit dem Hodgkin-Huxley-Modell lässt sich erstmals der Mechanismus der Aktionspotenziale quantitativ beschreiben. Zunächst benötigte es hierfür die Feststellung verschiedener Gegebenheiten und Zustände des Neurons:

zum einen die Anwesenheit von Flüssigkeiten, intra-zellulär wie extrazellulär, und zum anderen, dass in diesen Flüssigkeiten Ionen enthalten sind, die unterschiedlich elektrisch aufgeladen sind. Das Modell nimmt ein spezifisches Ruhepotenzial an, quasi einen Zustand 0, und beschreibt die Membranbeschaffenheit, seine Durchlässigkeit und geht davon aus, dass diese Membrandurchlässigkeit nicht nur gemessen werden kann, sondern auch Auskunft darüber gibt, wie aktiv ein Neuron ist beziehungsweise wie hoch ein Spike in einem Neuron ausfällt. Das Hodgkin-Huxley-Modell beruht auf der Modellierung des Membranpotenzials und dem daraus sich ergebenden Ionenstrom in der Synapse. Ausgehend von dem Wert des Ruhezustands eines Neurons $V_m = -65$ mV, lässt sich anhand gemessener Schwankungen die Reizweitergabe von Synapse zu Synapse modellieren. Durch einen Reiz, der beispielsweise das elektrische Signal eines anderen Neurons oder auch künstlich angelegter Strom sein kann, steigt das Membranpotenzial an. Erreicht der Reiz einen bestimmten Schwellenwert, findet eine abrupte Umkehrung des Membranpotenzials auf ungefähr +40 mV statt. Diese Phase wird als Depolarisation bezeichnet. Während der anschließenden Repolarisation fällt das Membranpotenzial schnell ab, bis es negativer ist als das ursprüngliche Ruhepotenzial. Diese Phase des Aktionspotenzials wird als Hyperpolarisation bezeichnet. Anschließend stellt sich das Ruhepotenzial $V_m = -65$ mV wieder ein (vgl. Sterratt 2014, 13).

Die Arbeit der beiden Physiologen verlief in drei Hauptphasen: Sie zeichneten intrazellulär Aktivität am Riesenaxon eines Tintenfisches auf. Sie passierten diese Ergebnisse an ein mathematisches Modell an (Modellierung) und sie lösten die Gleichungen, die das Modell definieren, um das Verhalten des Membranpotenzials unter verschiedenen Bedingungen zu beschreiben (Simulation, vgl. Sterratt 2014, 48). Ihr eigentliches Vorhaben, die molekularen Mechanismen, die den Eigenschaften der Membran zugrunde liegen, herauszufinden, konnten sie nicht erfolgreich durchführen.

Das Hodgkin-Huxley-Modell wurde zur Untersuchung einzelner Neuronen entwickelt und eignet sich aufgrund seiner vielen Parameter nur schlecht zur Modellierung großer Netzwerke. Hierfür bieten sich einfachere Modelle wie das Fitz-Hugh-Nagumo-Modell an (vgl. ebd., 197). Auch wenn das Hodgkin-Huxley-Modell für die Modellierung großer Netzwerke zu aufwendig ist, ist es an anderer Stelle zu verkürzt: So finden sich mittlerweile Beschreibungen synaptischer Ströme, in denen die Übertragung durch verschiedene Ionen übermittelt wird. Diese Unterscheidung wird im Hodgkin-Huxley-Modell nicht weiter berücksichtigt.

Gemeinsam mit John Carew Eccles (1903–1997), der in etwa zur selben Zeit die Signalweiterleitung von Nervenzellen untersuchte, erhielten Hodgkin und Huxley 1963 den Nobelpreis in Physiologie und Medizin für ihr Verdienst, den Ionenmechanismus bei Erregung und Hemmung in der Nervenzellmembran entdeckt zu haben. Als Schüler Sherringtons knüpfte Eccles an dessen Arbeiten zur Analyse der Reflexorganisation und der interneuronalen Verrechnung an und identifizierte im Jahr 1964 »den Mechanismus der synaptischen Erregungsübertragung«. (Breidbach 1997, 271) Eccles bewies damit die von Sherrington zunächst nur postulierte physiologische Grundschalt-einheit im Nervensystem und charakterisierte sie gleichzeitig, in dem er die synaptische und postsynaptische Inhibition als Kommunikation der Zellen beschreibt: »Communication between nerve cells occurs at these numerous areas of close contact or synapses, the name first applied to them by Sherrington, who laid the foundations of what is often called synaptology.« (1963, 1. Nobelpreisrede) Er widmete seine Forschung der Frage, wie diese Kommunikation genau vorstatten geht, also wie Aktionspotenziale über den synaptischen Spalt hinweg weitergeleitet werden. Um diesen Sprung erklären zu können, wurde zunächst auf zwei unterschiedliche Erklärungswegen zurückgegriffen: Während die eine, unter anderem angeregt durch Sherrington, davon ausging, dass chemische Botenstoffe an den Synapsen eine zentrale Rolle spielen, hielt die andere eine direkte elektrische Weiterleitung für wahrscheinlicher. Eccles hing lange Zeit der elektrischen Theorie an und sammelte in seinen Experimenten Daten, um diese zu unterstützen. 1949 gelangte er zu dem Schluss, dass es sich doch um eine chemische Vermittlung zwischen den Synapsen handelt. Durch Potenzialmessungen in Einzelzellen lebender Tiere konnte er 1951 an einer inhibitorischen Synapse ein Potenzial messen, das die Wirksamkeit der chemischen Übertragung nun auch für das zentrale Nervensystem bewies.

Der Physiologe Eccles stellte auch philosophische Überlegungen an. Darauf zeugt etwa das mit Karl Popper verfasste Werk *The Self and its Brain* (1977) (*Das Ich und sein Gehirn* [1982]). Darin äußert sich Eccles auch zu der Frage, wie das Leib-Seele-Problem aufgestellt ist, also ob die physiologischen und chemischen Bedingungen des Gehirns, die er ja konkret untersuchte, das Ich und seine mentalen Zustände hervorbringen oder ob es noch eine darüber oder darunter liegende Instanz der Vermittlung benötigt. Eccles glaubte an Zweiteres.

2.7 Assoziatives Lernen und Computermodellierung – Marr

Die Kognitionswissenschaft verlegte sich nun endgültig darauf, Sehen und visuelle Wahrnehmung als Ausgangspunkt für ihre Theorien neuronaler Verarbeitung anzunehmen. 1973 laden Marvin Minsky und Seymour Papert den Mathematiker, Informatiker und Psychologen David Marr (1945–1980) in das von ihnen gegründete Artificial Intelligence Laboratory am MIT ein. Marr nimmt die Einladung an und entwickelt hier seine mathematischen Berechnungen visueller Reizverarbeitung unter Berücksichtigung bereits vorhandenen anatomischen und physiologischen Wissens. Der junge Marr nutzte die Zeit im AI-Labor, um seine Überlegungen zu spezifizieren, die posthum nach seinem frühen Tod im Alter von 35 Jahren unter dem Titel *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information* (1982) veröffentlicht wurden.

Die Wichtigkeit von Marrs Arbeiten für die Computational Neurosciences zeigt sich in seinen grundsätzlichen Überlegungen über die menschliche Wahrnehmung von Objekten und den daraus resultierenden Computermodellen. Er setzte zunächst Sehen (Vision) mit dem Wahrnehmen von stark abstrahierten Repräsentationen/Bildern der Welt gleich und entwickelte ein dreistufiges Modell, mit dem sich die Verarbeitung dieser visuellen Repräsentationen formalisieren ließe. Marr schreibt:

Vision is therefore, first and foremost, an information-processing task, but we cannot think of it just as a process. For if we are capable of knowing what is where in the world, our brains must somehow be capable of representing this information – in all its profusion of color and form, beauty, motion, and detail. The study of vision must therefore include not only the study of how to extract from images the various aspects of the world that are useful to us, but also an inquiry into the nature of the internal representations by which we capture this information and thus make it available as a basis for decisions about our thoughts and actions. (2010 [1982], 3)

Zunächst merkt Marr an, dass es mehrerer methodischer Zugänge bedürfe, um sich dem Problem der Wahrnehmungsverarbeitung zu nähern. Insbesondere drei Ansätze in den Neurowissenschaften müssten in ein Zusammenspiel gebracht werden müssen, wie Rich et al. zusammenfassen: »[D]ata analysis, mathematical modelling and the one that became clearer through Marr's later work, namely computational modelling. [...] Computational modelling most interested Marr. He looked upon the brain as a physical device that

is performing computational tasks involving representing, processing and acting upon information.« (2015, 2) Diese drei Ansätze nehmen die zentrale Rolle in Marrs Buch ein und bilden die Grundlage für seine computergestützten Theorien über das menschliche Sehsystem.

Um ein allgemein gültiges Modell zu formulieren, nimmt Marr die Dualität zwischen Repräsentationen und Prozessen an. Bei Untersuchungen sowohl von Repräsentationen als auch von Prozessen werden allgemeine Probleme oft durch Alltagserfahrungen oder durch psychophysikalische oder sogar neurophysiologische Befunde ganz allgemeiner Art nahegelegt (vgl. Marr 2010, 331). Um zu einer Computertheorie der Wahrnehmung zu kommen, formalisiert Marr das Sehen von Objekten und seine räumlichen Aspekte: Er definiert Eigenschaften des zweidimensionalen Bildes, er schlägt auf Betrachter*innen zentrierte Darstellungen von Tiefe und sichtbaren Oberflächen und Konturen vor und setzt auf Diskontinuitäten der Objektgrößen; er schlägt ein 3-D-Modell vor, ein objektzentriertes Koordinatensystem, in dem die Objekte als Volumen angegeben werden, die die Organisation des von einem Objekt eingenommenen Raums und nicht nur seine sichtbaren Oberflächen explizit machen, und in dem die Objekte durch ihre verschiedenen Größen zueinander in Relation gesetzt werden (vgl. ebd., 330). Wenn eine rechnerische Theorie für einen Prozess formuliert wurde, können Algorithmen zu dessen Implementierung im Neuronalen Netzwerk, heißt in einem entsprechenden Computermodell entworfen werden. (Vgl. ebd., 331)

Diese äußere Objektwelt wird durch Wahrnehmungsprozesse, das heißt Berechnungsschritte ins Innere übertragen, so Marr:

From a philosophical point of view, the approach that I describe is an extension of what have sometimes been called representational theories of mind. On the whole, it rejects the more recent excursions into the philosophy of perception, with their arguments about sense-data, the molecules of perception, and the validity of what the senses tell us, instead, this approach looks back to an older view, according to which the senses are for the most part concerned with telling one what is there. Modern representational theories conceive of the mind as having access to systems of internal representations; mental states are characterized by asserting what the internal representations currently specify, and mental processes by how such internal representations are obtained and how they interact. (2010, 6)

Gleichzeitig widmet er sich in seinen Arbeiten dem altbekannten Problem, wie sich die Beziehung zwischen Denken und Gehirn theoretisch fassen lässt.

So beschreibt eine der von mir interviewten Wissenschaftler*innen den Stellenwert von Marr für den Bereich der Cognitive Computational Neurosciences folgendermaßen:

And they run into problems and they have to theorize of what is actually the relationship between thinking and the brain? What does it mean for the brain to compute? What does it even mean to explain thinking in brain processes? And in the foundations in cognitive science there is a classical work by David Marr who proposed, you can't explain for normative levels of distraction, what's the nature of the problem solved, what's the algorithm used for solving it, and what's the physical implementation? And this framework, I think, is now being readopted in cognitive neurosciences. They say, oh we need to think of these different levels of explanation. This is one example, but also with other things that not everything is associative learning, people have very creative processes, problem solving processes or actually a fundamental aspect of human cognition is that it is very open ended. And this we already knew, like in the 80s there were discussions about, that neural networks couldn't explain this. They cannot explain that in principle, people can understand sentences they have never heard before. They can speak sentences they have never said before. How is that possible? A neural network can only reproduce what it is trained on. (Interview 3, 20 Min.)

Marr bereitete mit seinen Arbeiten den Boden für die Verwendung von Computermodellen in den Kognitionswissenschaften als Alternative zur Konzeptualisierung Neuronaler Netzwerke. Gleichzeitig behauptet Marr nicht, mit seinem Ansatz alle Prozesse einschließen zu können, sondern zielt auf die Unterschiedlichkeit einzelner neuronaler Areale, deren Zusammenspiel in den Computermodellen eingefangen werden sollte. In seinem Vorwort zu Marrs Buch schreibt Shimon Ullman 30 Jahre später:

Because of this complexity, it is perhaps not surprising that the full understanding of the computations performed at the level of V1 and V2 is almost as elusive as it was thirty years ago. How discrete events are represented, and how they are anchored in present or future time and space to a specific context, remains poorly understood – though the notion of the hippocampus acting as a type of distributed associative memory that binds events to context has been discussed extensively in the neuroscience literature. (In Marr 2010, xx)

Die Arbeiten Marrs und insbesondere sein posthum erschienenes Buch *Vision* setzen den Startschuss für die Computational Neurosciences sowie die Neuroinformatik, die, so wie Marr selbst, sich in ihren Überlegungen aus den Teildisziplinen der Psychologie, Mathematik und Informatik bedienen. Der Unterschied zwischen den beiden Bereichen Neuroinformatik und Computational Neurosciences ist oft nur ein gradueller: Die Neuroinformatik als Unterdisziplin der Informatik und der Neurobiologie befasst sich mit der Informationsverarbeitung Neuronaler Systeme in einem stark anwendungsorientierten Bezug zur Verwendung der Erkenntnisse in technischen Systemen. Computational Neurosciences hingegen als Unterdisziplin der Neurobiologie und der Physik suchen nach einem Verständnis neuronaler Systeme mittels mathematischer Modelle. Ab den 1980er-Jahren werden – unter anderem auch im Anschluss an Marr – immer häufiger Computermodelle zur Berechnung neuronaler Verarbeitungsprozesse entwickelt. Im Unterschied zu Marrs anfänglichen Überlegungen von Computermodellen spielen dabei stochastische, also wahrscheinlichkeitstheoretische Überlegungen sowie die Vorhersage von eintretendem Verhalten eine bedeutendere Rolle und damit weniger die Berechnung, sondern die Simulation von Prozessen. Durch die rechenstarken Computer können immer mehr Variablen in die Berechnung und die Simulation aufgenommen werden, um eine mit Komplexität verwechselte Variabilität der Ausgangs- und Randbedingungen zu bestimmen.

2.8 Computermodelle und ihre Auswirkungen für die Erkenntnisproduktion

Die Arbeit der Computational Neurosciences liegt in der theoretischen Vorbereitung, der Konstruktion und Verwendung von Computermodellen. Modelle sind vereinfachte Repräsentationen von ›Wirklichkeit‹, entweder in Form konkreter Gegenstände oder als rein abstrakte Übersetzungen etwa in Computermodellen. Um ›Realität‹ in Modelle übersetzen zu können, werden Vereinfachungen und Abstraktionen vorgenommen, die meist einen Schwerpunkt in der Funktionsweise des Modells festlegen. Die Abstraktionen führen so zur Konzentration auf Teilespekte des gesamten komplexen Gegenstandes oder Vorgangs, was sich wiederum auf die theoretischen Überlegungen zurückwirkt. Theoretische Überlegungen und die Art der Modellierungen, die vorgenommen werden, bedingen somit einander. In den organischen wie artifiziellen Neuronenmodellen wurde der Fokus auf die Entscheidungsmechanismen synaptischer und neuronaler Tätigkeit gelegt.

Das Wesensmerkmal der Kybernetik waren die mathematische Behandlung von Information und die mathematische Neudeinition des Informationsbegriffes. In der Kybernetik wurden nun alle Naturvorgänge als informationsverarbeitende Vorgänge betrachtet, der Fokus lag auf den Steuerungs- und Regelungsmechanismen. Norbert Wiener, gerne als Vater der Kybernetik beschrieben, greift hierfür die boolesche Algebra auf und macht Vorschläge, wie diese in die Rechenmaschinen eingebaut werden muss und auf welchen Prämissen sie beruht.

Die Implementierung einer elektrotechnisch ausführbaren Aussagenlogik unter Zuhilfenahme boolescher Ordnungsfunktionen führt sukzessive zu den Möglichkeiten gegenwärtiger Erkenntnisproduktion mithilfe von Computern. Nicht alle am Computer getätigten wissenschaftlichen Arbeiten begründen sich auf die oben beschriebene Vereinigung rechnender Maschinen und mathematischer, statistischer Logik. Gleichzeitig beruht jede empirische, statistische Praxis des Computers auf der oben beschriebenen Implementierung einer spezifisch mathematisch formalisierten Sprache (und ihren Kategorisierungen) und hintereinander geschalteten Schritten, deren Algorithmen nach binären Antwortmöglichkeiten wie ja/nein, anwesend/abwesend, ähnlich/nicht ähnlich verlangt. Diese mathematischen Bedingungen der Erkenntnisproduktion wie der zugrunde liegenden Logik bestimmen gegenwärtig die Erkenntnisse über die Funktionsweise menschlicher Gehirne: durch ihre wissenschaftliche Praxis stochastischer Berechnungsweisen synaptischer Aktivität und Neuronaler Netze ebenso wie durch ihre algorithmische Suche nach Mustern in Big Data. Die Logik der formalen Sprache, alle physikalischen wie physischen Phänomene des Gehirns sowie die Logik sich in Konkurrenz zueinander (Ausschluss/Entweder-oder) verhaltender und linear angeordneter Prozesse bedingen demnach die Erkenntnisse über das Gehirn. Für diesen letzten Schritt von einer zunächst mathematischen, deduktiven Logik hin zu einer induktiven Logik braucht es noch die Implementierung der Errungenschaften der Wahrscheinlichkeitstheorie, die sich parallel entwickelte.

Auf diesem Weg wurden Theorien und Neuronenmodelle entwickelt, mit deren Hilfe sich Vorhersagen zukünftigen Verhaltens ableiten, wie zum Beispiel Wachstumstendenzen oder die Entscheidungstätigkeit Neuronaler Netze. Stimmen die experimentellen Ergebnisse mit den aus der Theorie abgeleiteten Vorhersagen überein, ist die Theorie bestätigt, stimmen sie nicht überein, kann die Theorie verworfen werden. Will man eine Theorie vollständig testen, um herauszufinden, ob die hieraus gewonnenen Annahmen wirklich

richtig sind, muss man auf die statistischen Prüfmethoden zurückgreifen, da die stochastischen Modelle allein in der Logik der Statistik auf ihren Wahrheitsgehalt beziehungsweise ihre Validität getestet werden können. Die Ergebnisse werden eben nicht auf wahr/falsch geprüft, sondern dahingehend, wie hoch die Wahrscheinlichkeit ist, dass die Annahme valide ist. Es geht also um Validität, nicht um Wahrheit, siehe Ian Hacking. Zum Testen braucht es statistische Schätzer wie die Bonferroni-Ungleichungen zur Abschätzung der Wahrscheinlichkeit des Durchschnitts beziehungsweise der Vereinigung von Ereignissen.

Das Fachbuch *Principles of Computational Modelling in Neuroscience* (Sterratt et al. 2014) hält fest, dass für ein erfolgreiches Modellieren Theorien zunächst in Worte gefasst oder vielleicht mit einem Diagramm dargestellt, aber in jedem Fall beschrieben werden müssen. Auch um die Vorhersagen aus einer Theorie abzuleiten, können Worte und Argumentationen oder weitere Diagramme eingesetzt werden. Sie stellen wichtige Werkzeuge für die Theoriebildung dar. Aber es ist riskant, sich allein auf sie zu verlassen, da sich mit ihnen keine komplexen Problemfelder beschreiben oder theoriegetriebene Vorhersagen vornehmen lassen: Mit Worten, also sprachlich formulierten Theorien, wird erstens nicht genau genug spezifiziert, und zweitens sind viele Theorien zu komplex für qualitatives verbales Schlussfolgern, sodass keine Vorhersagen getroffen werden können. Hinzu kommt, dass die hierfür eingesetzte Statistik knifflig und kompliziert ist. Hier kann Stochastik weiterhelfen. Die Lösung des Problems der hohen Komplexität besteht darin, die Theorie genauer zu spezifizieren, und zwar in der Sprache der Mathematik, die die Beziehung zwischen den Variablen genauer und eindeutiger herstellen kann. Die Gleichungen können dann aufgelöst werden, um eine Vorhersage zu treffen. Eine Theorie, die durch Gleichungssätze spezifiziert wurde, nennen wir ein mathematisches Modell. Jedes Modell kann dann als eine oder mehrere Differentialgleichungen ausgedrückt werden, um die Komplexität des ausgehenden Problems einzufangen und jede mögliche Antwort mathematisch zu modellieren und nachzuvollziehen (vgl. Sterratt 2014, if.).

Ein gut konstruiertes und nützliches Modell ist eines, das dazu dient, das Verständnis des Untersuchungsgegenstandes zu vergrößern, nicht um das System detaillierter zu beschreiben oder zu untersuchen, sondern um das *Verhalten* eines Systems unter den beschriebenen Bedingungen *vorherzusagen*.

Mathematical modeling involves building formal reductions based on descriptive and mechanistic models of natural phenomena associated with

the brain over the huge range of spatial and temporal scales that characterize it. These reductions have to take into account the complexity and heterogeneity of the brain's components. Mathematical methods and computer simulations are used to explore whether the mechanisms proposed are capable of accounting for the phenomena they are intended to explain. (Willshaw/Dayan/Morri 2015, 2)

Ganz in der Tradition von Marr ist für die Vorgehensweise des Computermodellings in den Computational Neurosciences die sukzessive und mittlerweile sehr detaillierte Vermessung der einzelnen Teilbereiche eines Prozesses, der untersucht werden soll, grundlegend. Für die Computational Neurosciences ist bis heute die weiter oben beschriebene Hodgkin-Huxley-Simulation der Ausbreitung eines Nervenimpulses (Aktionspotenzials) entlang eines Axons das maßgebliche Modell. Das System und die Gleichungen, die dieses System beschreiben, stehen fest, für weitere Berechnungen können die Variablen und die eingesetzten numerischen Bestimmungen geändert werden. *Computational modelling* ist die einzige Methode, mit der eine Theorie in den Computational Neurosciences auf ihre Richtigkeit geprüft werden kann.

Computermodelle dienen als Hilfsmittel, geeignete Entscheidungen zu fällen und damit die eigenen Überlegungen zu begründen. Ohne den Computer wäre dies in den Computational Neurosciences nicht möglich, da eine große Anzahl interagierender Elemente des neuronalen Systems in die Überlegungen miteinbezogen werden muss. Die für die Berechnungen verwendeten Daten sind oft nicht von ausreichend guter Qualität, da sie aus indirekten Messungen gewonnen werden. Computermodelle helfen dabei, die interessanten Variablen in den Daten abzuleiten. Computermodelle beseitigen durch ihre mathematische Formalisierung die Mehrdeutigkeit von Theorien. Die Modelle, die für einen Großteil neurobiologischer Systeme meist auf zellulärer Ebene entwickelt wurden, haben einen solchen Komplexitätsgrad erreicht, dass sie von den Wissenschaftler*innen als realistische Modelle akzeptiert werden. Die Anzahl der interagierenden Elemente, die simuliert werden, etwa von Neuronen, kann dank der Fortschritte in der Computertechnologie sehr groß und damit repräsentativ sein. Computermodelle helfen dabei, Hypothesen zu überprüfen, und können somit Experimente ergänzen. Obwohl Experimente für die Entwicklung eines Modells und die Festlegung von Anfangsparametern unerlässlich sind, können Modellierungen dazu dienen, die effektive Reichweite von Experimenten zu erweitern (vgl. Sterratt et al. 2014, 5).

Jedes mathematische Modell bedient sich als Grundlage eines anderen, auf eine bestimmte Funktion spezialisierten Typus mathematischer Modelle. Die Wahl des Modells hängt davon ab, welche Aspekte eines Systems modelliert werden sollen und wie man mit Parametern umgeht, die experimentell bisher noch nicht vermessen wurden oder die gar nicht gemessen werden können.

Im nächsten Abschnitt wird es um die Geschichte der Neuronalen Netzwerke und der künstlichen Neuronalen Netzwerke gehen. Für die Computational Neurosciences sind Neuronale Netzwerke nicht das bevorzugte Untersuchungsfeld, die von mir befragten Wissenschaftler*innen haben sich in der Beschreibung ihres Untersuchungsgegenstandes explizit von Neuronalen Netzwerken abgrenzt. Dabei haben sie deutlich gemacht, dass die Computational Neurosciences in der Tradition der physiologischen und physikalischen Beschreibung von Prozessen stehen beziehungsweise ausdrücklich an die Kognitionswissenschaften anknüpfen. Darüber hinaus waren sie sich darin einig, dass Computational Neurosciences an die drei Ebenen der Analyse (Marr) anknüpfen und es hierbei konkret um das Finden von Computermodellen zur Beschreibung neuronaler Reizverarbeitung geht. Auch wenn Computational Neurosciences nicht dezidiert an der Weiterentwicklung artifizieller Netzwerke forschen, im Gegensatz zu anderen, dem Konnektionismus und damit der Informatik näherstehenden Disziplinen, verwenden sie erstens in ihren Computermodellen neuronale Netzalgorithmen, und zweitens erhalten Neuronale Netze ab den 1990er-Jahren eine enorme Bedeutung, die im Folgenden näher beleuchtet werden soll.

Stochastik in Computermodellen

Nachdem Marr wichtige Grundlagen für die Übersetzung von Wahrnehmung in Computermodellen geschaffen hatte, sind mathematische Modelle heute aus den Computational Neurosciences nicht mehr wegzudenken. Das Modellieren von Neuronenmodellen mithilfe stochastischer Berechnungen in Computern stellt heute das Hauptbetätigungsgebiet der Computational Neurosciences dar. Jedes stochastische Modell basiert auf der Implementierung von Wahrscheinlichkeitsverteilungen und Zufallselementen. Gleichzeitig ist die Operationalisierung von Zufallsverteilungen in den Algorithmen die Hauptschwierigkeit beim Modellieren: »At the heart of creating a stochastic model is the problem of defining suitable probability distributions for model quantities that are random variables.« (Sterratt et al. 2014, 344) Das bedeutet,

dass jedes Computermodell nicht nur auf mathematischen Modellen beruht, sondern auch, dass jedes Computermodell, das angemessene Wahrscheinlichkeitsverteilungen für Zufallsvariablen festlegt, auf der Verwendung von Stochastik beruht. »Quantities whose value is in some way random, be they model variables or parameters, or experimentally measured data, are described by probability distributions which assign a probability that a quantity will have a particular value or range of values.« (Ebd., 342) Zufallsgesteuerte Prozesse in Neuronenmodellen, die mit Stochastik berechnet werden, gibt es viele. Hierzu gehört zum Beispiel das Modellieren der Membrantätigkeit bei sich öffnenden und schließenden Ionenkanälen, molekulare Wechselbeziehungen in innerzellulären Signalwegen und die Transmitterstreuung, um nur einige prominente Beispiele zu nennen. Als Beispiel eines nicht linearen Prozesses kann das Aktionspotenzial eines Neurons herangezogen werden:

[W]enn jetzt das Membranpotenzial eine bestimmte Schwelle erreicht, dann gibt es eine starke Reaktion einer Nervenzelle, das ist ein höchst nicht-lineares Phänomen. Denn wenn man zum Beispiel den Input ums Doppelte erhöht, erhöht sich jetzt nicht das Membranpotenzial ums Doppelte, sondern das ist ein superstarker nicht-linearer Prozess. Und das, mit diesen Tools, die man da in der Physik gelernt hat, die sind sehr hilfreich, um komplexe neuronale Systeme zu untersuchen. (Interview 2, Min. 7f.)

3 Ideengeschichte Neuronaler Netzwerkmodelle. Übersetzungen und das Finden einer adäquaten symbolischen Sprache komplexer Prozesse

Die in den beiden letzten Kapiteln beschriebene Ideengeschichte der Durchsetzung einer Mathematischen Logik, dem Zusammenkommen von Mathematik in Experimentaltechnologien und damit die Herausbildung der Physik und – elementar für die Neurowissenschaften – der Physiologie dient als Ausgangspunkt für die Beschreibung aktueller Entwicklungen in den folgenden Kapiteln. Diese werden von mir mit den Konzepten der *Laboratisierung der Gesellschaft* und der *Mathematisierung der Wahrnehmung* beschrieben. Das 1943 von McCulloch und Pitts vorgeschlagene und mit dem Nobelpreis ausgezeichnete Neuronenmodell findet heute zwar kaum noch Anwendung, ist aber nach wie vor wohlbekannt, eröffnete der Ansatz doch ein grundlegend neues, technisches, Verständnis neuronaler Prozesse. Das Konzept der Neuronalen