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 innerzellularen 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

Netzwerke nimmt erst durch das operative Verständnis der Kybernetik richtig Fahrt auf und findet dadurch Eingang in die sich formierende Informatik (vgl. Breidbach 1997, 23). Auch wenn sich in den Vorläuferwissenschaften der Computational Neurosciences von Beginn an bereits komplexere Modelle von Neuronenverbänden finden lassen und detailliertere Verhaltensweisen von spezifizierten Neuronen und Synapsen beziffert und in die Computermodelle integriert wurden, zeigt die Geschichte der Physiologie, als Ausgangspunkt der Computational Neurosciences, dass anfänglich die komplexen Strukturen und Abläufe zerebraler Prozesse auf sehr vereinfachte Netzwerk Konstrukte und ein abstraktes Verständnis der Reizverarbeitung im Gehirn runtergebrochen wurde.

3.1 Künstliche Neuronale Netzwerkmodelle

In den gegenwärtigen Computational Neurosciences steht nicht mehr der Computer Pate für die Funktionsweise des Gehirns, denn nach jahrzehntelanger Gleichsetzung von Gehirn und Computer wird die Algorithmizität des Gehirns nicht mehr infrage gestellt, sondern als gegeben angenommen. Neuronale Vernetzungen stellen nun das Vorbild für die Funktionsweise des Computers dar. Zwei Funktionsweisen des menschlichen Gehirns werden in diesem Ansatz hervorgehoben: zum einen seine Fähigkeit, Informationen zu verarbeiten, und zum anderen seine Fähigkeit, aus Beispielen zu lernen. Die künstlichen Neuronalen Netzwerke, die vermeintlich die Neuronalen Netzwerke des Gehirns nachbilden, sollen als Vorbild dienen, den Computer zum Lernen zu bringen. Die verschiedenen Vordenker Neuronaler Netzwerke, so Breidbach (1997), »gewannen ihre Inspiration zu einem entsprechenden Vorgehen nicht etwa aus einem unverbauten Blick auf die Realitäten natürlicher neuronaler Netze« (23). Dies habe ich anhand der unterschiedlichen Ansätze bereits weiter oben näher ausgeführt. Die Inter- und Transdisziplinarität der Theorien wie auch des Bereiches, in denen das Modell angewendet wird, gilt bis heute: »Die derzeitig verfolgte Hypothese, über das Modell der neuronalen Netze mehr und Neues über die prinzipiellen Verrechnungseigenschaften des Hirngewebes zu erfahren, erwuchs aus einem kompliziert ineinandergreifenden Dialog verschiedener Disziplinen.« (Ebd.)

Im Folgenden werde ich die Theorien artifizieller Netzwerke seit den 1990er-Jahren bis heute vorstellen. Neuronale Netzwerke werden in der Hirnforschung wie in der künstlichen Intelligenz und der Informatik zum wichtigsten konnektionistischen Modell, zu einer alles fundierenden Meta-

pher, die jegliche logische Annahmen und Verarbeitungsprozesse im Gehirn wie im Computer ordnet. Neuronale Netze und künstliche Neuronale Netze sind originärer Forschungsgegenstand der Neuroinformatik und stellen heute einen wichtigen Untersuchungsgegenstand und ein zentrales Einsatzgebiet der künstlichen Intelligenz dar. Auch wenn biologische Neuronale Netze als Vorbild artifizieller Netze dienen, sind künstliche Neuronale Netzwerke Abstraktionen und Modelle, das Nachbauen biologischer menschlicher Neuronennetzwerke ist der Forschungsgegenstand der Computational Neurosciences.

Den Anfang machten McCulloch und Pitts mit ihrer »Arbeit zur Theorie neuronaler Netze, in der sie zeigen konnten, daß jede aussagenlogische Funktion vermittels eines neuronalen Netzwerkes von einfachen binären Schwellenwerten simuliert werden kann« (Sichtweisen der Informatik, 78). Bereits 1947 erkennen McCulloch und Pitts, dass solcherart modellierte Netze unter anderem für die räumlichen Mustererkennungen eingesetzt werden können. Mit der hebbschen Lernregel stellte Donald Hebb 1949 eine allgemeine Formel auf, die bis heute in die meisten der künstlichen neuronalen Lernverfahren integriert wurde. Auf dem Gebiet künstlicher Neuronaler Netze erzielte Frank Rosenblatt im Jahr 1958 einen wichtigen Durchbruch, indem er den Netzwerken ihre Form gab mit einer bestimmten Anzahl *n* an Input-Neuronen, einer der Anzahl der Input-Neuronen angepassten Anzahl an Hidden-Neuronen und einem am Ende des Netzwerks stehenden Output-Neuron. »Mit dem Perzeptron hatte er ein Modell entwickelt, welches sich teilweise selbst organisieren und formales Lernen realisieren, einfache Muster erkennen und klassifizieren konnte.« (Fuchs-Kittowski 1992, 78) Der Erfolg dieser ersten Generation artifizieller Neuronaler Netze kam zu einem vorläufigen Ende durch die von Minsky und Papert geäußerte Kritik und deren Veröffentlichung in dem Buch *Perceptrons* (1969). Das von Marvin Minsky und Seymour Papert publizierte Buch untersucht die prinzipielle Leistungsfähigkeit zweischichtiger neuronaler Feedforward-Netze und zeigt darin ihre funktionale Beschränktheit auf.

> Erst 1985 wurde von Rumelhart und Hinton ein leistungsfähiger Lernalgorithmus entwickelt, der es ermöglicht, ein Fehlerintegral auch für die Neuronen verdeckter Schichten zu definieren. Mit diesem Algorithmus der Backpropagation wird es nun möglich, Netzwerke mit mehreren verdeckten Schichten zu entwickeln und neuronale Netzwerke wieder verstärkt zu untersuchen. (Fuchs-Kittowski 1992, 78)

All diese Konzepte sehen das Gehirn als Rechenmaschine, das nach festen, vorgegebenen Regeln rechnet. ›Intelligenz‹ beruft sich hierbei auf die Logik der Programmierung von Computern. Der Ansatz, alle denkbaren Regeln für bestimmtes Verhalten von oben nach unten (top down) zu bestimmen – also sie zu codieren –, kommt schnell an seine Grenzen. Erst mit der Rückkehr zu einem konkurrierenden Paradigma Neuronaler Netze, das aus dem Konnektionismus stammt, und der Nachahmung des Bottom-up-Verfahrens (von unten nach oben) versuchte man, Neuronale Netze aus parallelen, in sich selbst nicht intelligenten Prozessoren aufzubauen.

Im Anschluss an diese Nachahmung des Bottom-up-Verfahrens beruhen künstliche Neuronale Netzwerke auf neuronalen *circuits*, wie wir sie aus der Elektrotechnik kennen, und wurden zunächst als Einheiten getrennter Systeme konzipiert, die sich mathematisch gut berechnen ließen. Erst später wurden aus den so beschriebenen Kreisläufen selbstreferenzielle, sich selbst organisierende, als kleine Schaltpläne imaginierte Einheiten (*circuits*) dynamisch miteinander verwobener Netzwerke. Die Kybernetiker Wiener und Minsky arbeiteten eine Theorie aus, die das Gehirn als Schaltplan plausibilisiert. Gleichzeitig und darauf aufbauend entwickelte sich der Computer als Verschaltung boolescher Funktionen. In den Jahren nach 1943 ergänzten Neurowissenschaftler*innen die Idee Neuronaler Netze mit Konzepten und Techniken aus der mathematischen Physik, der Kontrolltheorie, der Wahrscheinlichkeitsrechnung und Statistik sowie der Informationstheorie. Trotz dieser konzeptionellen Erweiterungen ist das Erbe von McCulloch und Pitts in der aktuellen theoretischen Neurowissenschaft immer noch präsent, zumindest in der terminologischen Wahl, neuronale Aktivität als Berechnung zu beschreiben. Auch wenn aktuelle mathematische Modelle nicht mehr auf Logik oder Berechenbarkeitstheorie zurückgreifen, um neuronale Systeme zu beschreiben, verwenden doch viele theoretische Neurowissenschaftler*innen das Bild, dass Neuronale Netze, Neuronen, aber auch subneuronale Strukturen wie Dendriten und Synapsen ›Berechnungen‹ durchführen. Ab den 1990er-Jahren und aufgrund sich abzeichnender Mängel des kognitiven Modells,

> nämlich daß Symbolverarbeitung auf sequentiellen Regeln beruht und lokalisiert erfolgt und daß es aus der Sicht der Neurophysiologie einer synthetischen Vorgehensweise bedarf, entwickelt man heute wieder verstärkt auf der Grundlage des Prinzips der Selbstorganisation konnektionistischer

> Modelle. Sie sollen wichtige kognitive Fähigkeiten wie Wiedererkennen und assoziatives Gedächtnis realisieren. (Sichtweisen der Informatik, 79)

Der Kognitivismus wird durch konnektionistische Modelle abgelöst beziehungsweise werden beide Bereiche stärker miteinander verknüpft. Nicht mehr die Frage der Verarbeitung von zeichenhaften Repräsentationen steht von nun an im Fokus, sondern die Selbstorganisation und Lernfähigkeit von Systemen auf neuronaler Ebene.

Der Wunsch, komplexere Zusammenhänge untersuchen zu können, führt dazu, dass in die mathematischen Gleichungen immer häufiger Stochastik miteinfließt, die Komplexität und Vorhersagbarkeit verspricht und somit vermehrt zu Modellierungen und Simulationen neuronaler Prozessverarbeitung führt.

> Nach dem Paradigma des Konnektionismus ist die Semantik nicht in bestimmten Symbolen lokalisiert, sondern eine Funktion des Gesamtzustandes des Systems. Sie ergibt sich aus dem Funktionieren z.B. der Wiedererkennung oder des Lernens. Der Gesamtzustand entsteht aus einem Netzwerk von Einheiten – oft als »sub-symbolische Ebene« bezeichnet. Indem die Bedeutungen nicht in diesen Bestandteilen, sondern in den sich aus der Interaktion der Bestandteile ergebenden komplexen Aktivitätsmustern existieren, gibt es hier eine deutlich andere Ebene für die Semantik. (Fuchs-Kittowski 1992, 79)

Nachdem Syntax und Semantik in der Nachrichtentechnologie durch den Begriff der Information verkürzt und ersetzt worden waren, entsteht nun aus dem Zusammenspiel syntaktischer Strukturen eine Semantik, die nicht mehr nur formale operative Logiken verstehen will, sondern ein Ganzes in den Blick zu nehmen meint.

Auch Hebbs und Marrs aus der Psychologie inspirierte Formalisierungsschritte sind für das Verständnis wichtig: Die Gleichsetzung von Neuronalen Netzen mit der Grundlage von Gedanken, dass sie die anatomische Form einer komplexen Gedankenwelt vorgeben und damit Denken als Abarbeiten algorithmischer Schritte wahrgenommen wird, mit ihren Reflexionsschleifen und ihrer Vorstellung von Komplexität, ist fundamental wichtig für aktuelle Debatten im Rahmen der Philosophie des Geistes, Fragen zu freiem Willen *versus* ihrer Determinierung durch die Netze im Gehirn und zu den KI-Debatten (mehr dazu in Kap. 6).

3.2 Rekurrente und Feedforward-Netzwerkmodelle – Hinton

Heute bestehen hauptsächlich zwei Neuronenmodelle nebeneinander, beide spielen eine besondere Rolle für die Konzeptualisierung künstlicher Neuronaler Netzwerke. Das ›vorwärtsgerichtete‹ Feedforward-Neuronenmodell ist die Grundlage für die neue, deutlich schnellere Generation selbstlernender (Deep-Learning-)Algorithmen. Rekurrente Netzwerke basieren auf den weiter oben beschriebenen Lernalgorithmen und zeichnen sich durch reziproke Verbindungen (Rückkopplungen) zwischen den Einheiten aus, das heißt, alle Einheiten sind wechselseitig miteinander verbunden. Die künstlichen Feedforward-Netzwerke fußen auf einer hierarchischen Architektur von Schichten, die wiederum aus kleinen Berechnungseinheiten bestehen. Für den Aufbau der Verbindungen zwischen den Neuronen wurde der heute weit verbreitete Backpropagation-Algorithmus (Rumelhart et al. 1986a) verwendet. Feedforward-Modelle weisen keine Rückkopplungen auf, das heißt, die Signalweitergabe läuft immer nur in eine Richtung, von der Eingabeschicht mit vielen Einheiten über versteckte (hidden) Schichten mit weniger Einheiten bis hin zur Ausgabeschicht, die meist nur noch über wenige oder sogar nur eine Einheit verfügt. Alle Einheiten aus der Eingabeschicht sind mit der nächsten Schicht verbunden, nicht aber untereinander.

> In this field the primary emphasis is on designing networks containing many nerve-cell-like elements that carry out useful tasks, such as pattern recognition. Feedforward networks which are made up of input neurons and output neurons and the addition of intermediate, so-called hidden, neurons increase their power and applicability. (Sterratt 2014, 241)

Am Anfang seiner Karriere steht für Geoffrey Hinton, einen der ›Väter‹ von Deep-Learning- Algorithmen, der Gedanke, dass Modelle der Informationsverarbeitung im Computer durch parallel stattfindende Prozesse repräsentiert sein müssen. Zu seiner Zeit war der Computer noch eine adäquate Analogie für den Informatiker und Kognitionspsychologen: »The brain is a remarkable computer« – befindet der bereits in der Einleitung vorgestellte Geoffrey Hinton in einem Artikel von 1992 (145). Hinton ging es zunächst um die Übersetzung von Repräsentationen, wie etwa der von Gesichtern, und darum, den Neuronalen Netzwerkalgorithmen beizubringen, Muster zu erkennen. Die Neurowissenschaften aber verabschiedeten sich in den 1990er-Jahren sukzessive von dem Wunsch, Verrechnungsfunktionen innerhalb eines parallel geschalteten Gefüges von Nervenzellen abbilden zu können, und

> folgten der Idee, in einer Analyse von Einzelzellaktivitäten prinzipielle Funktionseigenschaften des Hirns zu entschlüsseln. Bei diesem Vorgehen war aber im Laufe der siebziger Jahre klar geworden, daß die hierarchisch geordneten logischen Prozessoren, die die klassischen, das heißt nichtparallelen Rechnerarchitekturen auszeichnen, ein nur sehr unzureichendes Modell für die Organisation interneuronaler Verrechnungsprozesse darstellen. Zudem hatte die Forschung um die künstliche Intelligenz im Laufe der achtziger Jahre leistungsfähige Rechner entworfen, die […] Momente realer biologischer Systeme nachzeichneten« (Breidbach 1997, 25)

Was auch nach der Aufgabe der ersten Generation Neuronaler Netze bleibt, ist die »konstatierte Analogie zwischen realem und artifiziellem System« (ebd., 26).

Für die neue Generation selbstlernender Algorithmen bestimmt die Architektur der Neuronalen Netze ihre Leistungsfähigkeit, sodass es nicht nur darum gehen kann, leistungsfähigere Computer zu entwickeln, um eine Angleichung menschlicher und computationaler Performances zu erreichen, sondern neue Konzepte der Vernetzungsarchitektur zu entwickeln, die uns ein reales Verständnis von Lernprozessen vermitteln. Der heute als Deep Learning bezeichnete Ansatz neuer selbstlernender Algorithmen orientiert sich an zwei Schlüsseleigenschaften, die mit der Funktionsweise des menschlichen Gehirns assoziiert werden: die neuroarchitektonische Voraussetzung, Informationen parallel über mehrere, miteinander verbundene Gehirnzellen zu verarbeiten, und die Fähigkeit, aus Beispielen zu lernen:

> What makes people smarter than machines? They certainly are not quicker or more precise. Yet people are far better at perceiving objects in natural scenes and noting their relations, at understanding language and retrieving contextually appropriate information from memory, at making plans and carrying out contextually appropriate actions, and at a wide range of other natural cognitive tasks. People are also far better at learning to do these things more accurately and fluently through processing experience. (McClelland/Rumelhart/Hinton 1988, 3)

Was also macht Menschen, im Sinne eines ›Durchschnittsmenschen‹ – ordinary people –, intelligenter als Maschinen? Schaut man in die künstliche Intelligenzforschung, ergibt sich eine klare Antwort, wie im Interview weiter oben bereits angedeutet: Der Mensch kann all das, was der Maschine erst durch sogenanntes Lernen mühsam beigebracht werden muss, und dies ohne

Einwirkung eines ›äußeren Lehrers‹, der das Wahrgenommene in einen Kausalzusammenhang stellen muss: »A teacher, who knows what the response of each output unit should be for that particular input, indicates to each unit the size and sign of its error. For a theoretical model, the teacher is usually the person designing the net. In the brain the teacher is presumed to be another part of the brain.« (Crick 1989, 130) Aber nicht nur das, der Mensch kann auch Dinge erkennen oder Sätze verstehen, die er oder sie noch nie vorher gehört oder gesehen hat. Die Aufgabe, die sich daraus für die künstliche Intelligenzforschung ergibt, ist es, selbstlernende Algorithmen zu programmieren, die kausale Zusammenhänge erkennen können und keine äußere Instanz mehr brauchen, um zu entscheiden, ob die Ergebnisse in den algorithmischen Prozessen richtig oder falsch sind, sondern in die algorithmischen Einschätzungen als intrinsische Bewertungsskala in den Neuronalen Netzwerken verankert werden können. Eine Möglichkeit, um diese Instanz in algorithmische Systeme zu implementieren, stellt die Fehlerrückführung dar: die *backward propagation of errors*, kurz Backpropagation. Nach der Idee, Neuronale Netze als selbstorganisierte Systeme zu bauen, kam also ein weiterer Aspekt hinzu, der einen immensen Einfluss auf die weitere Entwicklung künstlicher Intelligenz haben sollte:

> The full name of the algorithm is »the back propagation of errors« but it is often called back prop for short. It can be applied to any number of layers, although only three layers are usually used: an input layer, a middle layer (referred to as the hidden units) and an output layer. A unit in each of the first two layers connects to all units in the layer immediately above. There are no reverse connections or sideways connections or sideways connections-a simple net indeed. Each unit forms the usual weighted sum of its inputs and emits a graded output. (Ebd.)

Der Grundgedanke basiert auf verschiedenen Schichten, zu Cricks Zeiten waren es meist drei, heute bestehen Neuronale Netze in der Regel aus mehr Schichten, was wiederum ihre Geschwindigkeit deutlich erhöht – allerdings auch eine deutlich schnellere Rechenleistung erfordert. Jede Schicht besteht aus kleinen Verarbeitungseinheiten, den Knotenpunkten (wie auf dem Buchcover zu sehen). Jeder Knotenpunkt einer Schicht ist mit jedem Knotenpunkt der nächsten Schicht verknüpft, Verbindungen rückwärts oder seitwärts gibt es nicht. Durch Zugabe eines Eingabewerts werden zunächst ›zufällige‹ Verbindungen aufgebaut, die Aktivität in der Ausgabeschicht erzeugen. Wird das erwünschte Ziel nicht erreicht, heißt, der Eingabewert wurde nicht erfolg-

reich übertragen, werden die Fehlersignale verwendet, um sie als Informationen an die versteckten Einheiten in der mittleren Schicht ›rückwärts zu übertragen‹. Diese nutzen die Fehlermeldung, um die Informationsverarbeitung in jedem dieser Knotenpunkte anzupassen. Die Schichten haben dabei unterschiedliche Funktionen. Geht man eine Schicht nach unten, werden Details über eine bestimmte Struktur und ihre Funktionen hinzugefügt; wenn man eine Ebene nach oben geht, abstrahiert man von den Details der unteren Ebene und fügt eine Struktur in ihren mechanistischen Kontext ein.

> That level of explanation can then be combined with other levels by showing how each structure perform its function in virtue of its lower level organization as well as how each structure fits within a larger containing mechanism. Going down one level involves adding details about a given structure and how it performs its functions; going up one level involves abstracting away from lower level details and fitting a structure into its mechanistic context. (Turner/De Haan 2014, 191)

Diese Deep-Learning-Algorithmen basieren auf den sogenannten Lernalgorithmen. Im Fall der Neuronalen Netze erfolgt das Lernen durch »Backpropagation«, bei der aus der Differenz zwischen der aktuellen Ausgabe und der gewünschten Ausgabe ein Signal abgeleitet wird, das an die dazwischenliegende Schicht zurückgegeben wird. Dadurch werden die Gewichte der Verbindungen zwischen diesen Zwischenschichten verändert und die Fähigkeit des Neuronalen Netzwerks, die Trainingsdaten zu reproduzieren, iterativ verbessert (vgl. McQuillan 2016, 6f.). Nach der Konstruktion eines künstlichen Netzes, eines Algorithmus, folgt die Trainingsphase, in der das Netz ›lernt‹. Die Idee hinter Deep Learning ist, der Maschine viele Beispiele für Eingaben sowie gewünschte Ausgaben zu präsentieren. Die jeweiligen Algorithmen sollen anhand der eingespeisten Trainingsdaten Muster erkennen. Ob erfolgreich Muster erkannt wurden, wird dem Algorithmus durch die Gegenüberstellung einer gewünschten Ausgabe als richtig oder falsch gespiegelt, sodass der Algorithmus, wenn er falsch liegt, die Ähnlichkeit zwischen zwei Mustern nicht richtig erkannt hat und erneut einen Suchprozess durchlaufen muss. Anhand dieser als ›richtig‹ oder ›falsch‹ erkannten Ähnlichkeit ändern sich die Verbindungsstärken des (künstlichen) Neuronalen Netzwerks auf lange Sicht so, dass Algorithmen die Ähnlichkeit von Mustern immer besser erkennen können. Deep-Learning-Netzwerke können durch folgende Methoden lernen: Entwicklung neuer Verbindungen; Löschen existierender Verbindungen; Ändern der Gewichtung von Neuronen; Anpassen der Schwellen-

werte der Neuronen, sofern diese Schwellenwerte besitzen; Hinzufügen oder Löschen von Neuronen; Modifikation von Aktivierungs-, Propagierungs- oder Ausgabefunktion. Diese Backpropagations-Netzwerke befördern eine Fehlerrückführung, wenn das eigentlich zu Erkennende vom Algorithmus nicht erkannt wurde. Für das Erkennen eines Bildes etwa verarbeiten die selbstlernenden Algorithmen schichtweise und Schritt für Schritt Informationen. Um beispielsweise ein Foto zu erkennen, registriert der Algorithmus der ersten Schicht nur Schwarz und Weiß, im zweiten Schritt ein paar grob gesetzte Merkmalsmarkierungen, sodass nach dem Durchlaufen vieler Schichten nach und nach ein Gesicht erkannt wird. Im Fall von künstlichen Neuronalen Netzen wird die Verbindung dadurch gestärkt oder geschwächt, dass die übertragenen Informationen anhand von Tausenden von Beispielen, die der Maschine zur Verfügung gestellt wurden, als richtig oder falsch erkannt werden.

Künstliche Neuronale Netzwerke sind schnell und können vielseitig eingesetzt werden: Sie kommen in Übersetzungsmaschinen, in Suchmaschinen und deren Vorschlägefunktion vor. Ein enormes Potenzial haben künstliche Neuronale Netzwerke im grafischen Bereich, in der Animation, im Film generell, im Berechnen fehlender Sequenzen, in der Berechnung von Verläufen, allgemein in der Mustererkennung und vielem, vielem mehr. In den Computational Neurosciences ist es eines von mehreren Neuronenmodellen, durch seine Übertragung und Nutzung in der Informatik definitiv das erfolgreichste. Trotz des Erfolgs dieser Feedforward-Neuronenmodelle gibt es eine breit geführte Diskussion, auch innerhalb des Feldes der Computational Neurosciences und der Neuroinformatik, über die Möglichkeit und Gefahren der Übertragbarkeit auf die Funktionsweise organischer Neuronaler Netze.

3.3 Kritik am Netzwerkmodell

Seit dem 20. Jahrhundert wird das Gehirn verstärkt über Neuronenmodelle erklärt und theoretisiert. Diese sind in den Computational Neurosciences immer auch mathematische Modelle. Den Startpunkt hierfür legten sicherlich die, anfänglich kriegsbezogenen, Arbeiten von Alan Turing, Claude Shannon, Norbert Wiener und anderen. Der Medizinhistoriker Cornelius Borck beschreibt die Phase in der die Entdeckung des Computers fällt, als spannende Zeit für die Neurophysiologie.

> Exciting times, indeed: during these decades, neurophysiology discovered the nervous system's operating mode to be a universal code of digital communication, […]. The computer was a materialization of just such electrical thinking and was thus a brain model of a new kind. In contrast to Sherrington's metaphor of the enchanted loom and brain models like the musical instrument or the telephone exchange that explained particular aspect of the brain, the computer was a real machine doing brain work. (2014, 13)

Aber ein direkter Zugang zu den funktionellen Vorgängen des Gehirns ist nach wie vor nicht gegeben. Physiologische Methoden konzentrierten sich auf die Registrierung funktionaler Veränderungen der Zellen über die Zeit, liefern aber, besonders im Falle des menschlichen Nervensystems, kaum verwertbare morphologische Informationen. »Over more than a century, brain research engaged its own variant of Heisenberg's uncertainty principle, the strict either/or in complementary data, as information related to either form or function.« (Ebd., 11)

Kritik an den neueren selbstlernenden Algorithmen und Modellen künstlicher Neuronaler Netzwerke wird da laut, wo deren Funktionsweise auf eine biologische Wetware übertragen werden soll. Wetware ist der Ausdruck für organische Lebensformen und steht im Zusammenhang mit den aus der Informatik bekannten Begriffen der Hard- und Software. Backpropagation-Netzwerke sind aus verschiedenen Gründen keine passenden Modelle für die menschlichen Wahrnehmungssysteme, und »[s]chnell konnte gezeigt werden, daß der Backpropagation-Algorithmus den biologischen Bedingungen nur eingeschränkt entspricht« (Fuchs-Kittowski 1992, 79). Das hatte auch Francis Crick erkannt, der bereits drei Jahre früher nüchtern schrieb:

> But is this what the brain actually does? Alas, the back-drop nets are unrealistic in almost every respect, as indeed some of their investors have admitted. They usually violate the rule that the outputs of a single neuron, at least in the neocortex, are either excitatory synapses or inhibitory ones, but not both. It is also extremely difficult to see how neurons would implement the back-prop algorithm. Taken at its face value this seems to require the rapid transmission of information backwards along the axon, that is, antidromically from each of its synapses. It seems highly unlikely that this actually happens in the brain. (1989, 130f.)

In einem der von mir geführten Interviews wurde dezidiert auf die Verwirrung stiftende Bezeichnung von künstlicher Intelligenz beziehungsweise *ma-*

chine learning als Neuronale Netzwerke hingewiesen: »In cognitive computational neuroscience, neural networks are the dominant modeling type. So the formalism of neural networks. Sometimes people call it machine learning. But it is usually ›neural networks‹.« (Interview 6, 30 Min.) In einem zweiten Satz wird das dominante Neuronenmodell der künstlichen Neuronalen Netzwerke kritisiert:

> [Neuronal networks] can not explain in principle, that 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. […] Sure these neural networks they're very powerful, because now you use very powerful computers, but the fundamental cognitive capacities that humans have, can not be explained by these kinds of algorithms. And we knew that in the 80ies, because we had those discussions already. (Ebd., Min. 31f.)

Neuronenmodelle sind seit der Kybernetik nicht mehr allein auf die Beschreibung organisch- morphologischer Neuronaler Netzwerke festgelegt. Sie können heute allgemein vielmehr als Modelle der Entscheidungsfindung beschrieben werden, deren Annahmen auf die neuronalen Strukturen des Gehirns rückübertragen werden. Die synaptischen Einheiten der künstlichen Netzwerke treffen mithilfe stochastischer Berechnungen Vorhersagen und können so darüber ›entscheiden‹, ob etwa ein Muster sich ähnelt oder wie Netzwerke untereinander kommunizieren.

Neuronenmodelle und mit ihr die mathematisierte und computerbetriebene Forschung haben konkret einen Perspektivenwechsel in den Neurowissenschaften hervorgebracht beziehungsweise zu einer Verengung dessen geführt, was unter Begriffen wie Lernen, Denken und Wahrnehmen aufgefasst und untersucht wird. Denken etwa wird unter den Schlagwörtern »Argumentieren«, »Problemlösen«, »Planen« zusammengefasst, »Wahrnehmen« in den Cognitive Sciences wird mit dem Auge gleichgesetzt, mit Vision, Wahrnehmung von Bildern, Bildsprachen und visuellen Repräsentationen. Insbesondere die Feedforward-Netzwerke haben sich zuletzt eher im Bereich der künstlichen Intelligenz hervorgetan, als einen Beitrag für die Computational Neurosciences zu bieten.

Im nächsten Kapitel sollen die Effekte, die sich aus den in Kapitel 1 und 2 beschriebenen Entwicklungen ergeben, mit Blick auf die Modelle der Hirnforschung beschrieben werden. Ausgehend von den Modellen der Kybernetik, bilden sich in den 1980er- und 1990er-Jahren stochastische Neuronenmo-

delle heraus. Auf der Grundlage des Computers als neuen Arbeitswerkzeugs und einer auf Wahrscheinlichkeitsrechnungen beruhenden Statistik differenzieren sich die mathematischen Modelle komplexer Netzwerke immer weiter aus. Durch die Verknüpfung von Statistik und der Logik der Mathematik entsteht die Stochastik und Probabilistik, und deren Anwendung führt zu einem epistemischen Wandel von Begriffen wie Komplexität, Kausalität und Zufall und von Zeitlichkeit. Dieser beschriebene Bedeutungswandel in der mathematisch-technisch operationalisierten Erkenntnisproduktion wird dann in einem zweiten Schritt als instrumentelle Vernunft ausgewiesen und kritisiert.