– both a promise to rethink difference, life, and our relations to each other and a warning that we will not. (Halpern 2014, 238)

2 Immer diese Widersprüche. Oder was es bedeutet, wenn Wissenschaftler*innen herausfinden wollen, warum die Patient*innen etwas anderes sagen als ihre Daten

Freedom is a possibility only if your're able to say no – *The Whitest Boy Alive*

Welches Wissen über den menschlichen Denkapparat kann mithilfe von Neuronalen Netzwerken überhaupt generiert werden? Welche Dimensionen menschlichen Denkens rücken in den Konzepten Neuronaler Netzwerke in den Blick, welche werden mithilfe dieser machtvollen Metapher nicht sichtbar, nicht denkbar gemacht? Der Verweis auf die metaphorische Ebene von Neuronenmodellen kommt hier nicht zufällig, soll damit doch zweierlei deutlich gemacht werden. Zum einen besteht ein Zusammenhang zu epistemischen Deutungsmustern, mit deren Hilfe ein Zugang zum Untersuchungsgegenstand hergestellt und im Anschluss daran Teile davon sichtbar, greifbar und dadurch verständlich gemacht werden können. Die Modelle, die Metaphern, leiten das Erkenntnisinteresse und das, was wir über einen Gegenstand wissen können. Zum anderen verweist der Rekurs auf die metaphorische, also die bildhafte Ebene, auf eine der Theorie immanente Problematik, die sich in der Art der Wissensproduktion, in der Sichtbarmachung von Wissen und Korrelationen durch mathematische Formalisierung versteckt: die Verknüpfung von Sehen, Wahrnehmen und Erkennen. Westliche Epistemologie knüpft den Vorgang des Erkennens und Wahrnehmens an das Auge. Über das Wahrnehmen mit dem Auge erzeugen und interpretieren wir Wissen, Wahrheit und Wirklichkeit. Bereits bei Platon bekommt das Sehen einen Sonderstatus, ist verknüpft mit dem Intellekt, im Unterschied zu den anderen Sinnen. Die Wissenschaft schafft Objektivität über das erkennende Auge, dass die Verallgemeinerung sucht und zu einer »vision from everywhere and nowhere« (Haraway 1988, 584) wird. Das einzelne Auge wird zum Zentrum der sichtbaren Welt. »Everything converges on to the eye as to the vanishing point of infinity. The visible world is arranged for the spectator as the universe was once thought to be arranged for God. In time,

the modern individual (the ›I‹) came to be centred on, if not abbreviated to, the eye (›I‹ equals eye).« (Kavanagh 2004, 448) Diese Vorherrschaft des Auges im erkenntnisleitenden Prozess wird heute stark kritisiert. Stimmen aus den Disability Studies (Whitburn/Michalko 2019) etwa, sehen in der direkten Verknüpfung von Sehen und Wissen, nicht nur eine harmlose, traditionsreiche Beziehung zwischen Sein, Wissen und Sehen, sondern eine voraussetzungsvolle epistemische Setzung:

> Conventional empiricism, which dominates Western philosophical approaches to knowledge, privileges a particular fixed way of conceptualising the world through the senses, foremost among them sightedness. Empiricism follows a single guiding principle: nihil in intellectu nisi prius in sensu (›nothing in the intellect unless first in sense‹), a radical concept for its time that diverges from rationalism. (Ebd., 220)

Die Kognitionswissenschaften binden diese uralte erkenntnistheoretische Tradition in ihre Wahrnehmungsmodelle ein. Und die Neuronenmodelle der Computational Neurosciences modellieren die Verarbeitung von meist über das Auge wahrgenommenen Informationen. Sie verstärken damit die Verknüpfung von Ereignis, Sehen, Wissen und Bedeutung.

Die Neuronenmodelle sind körperlos, allein das abstrakte Wahrnehmen über den Sehsinn findet Eingang in die Konzeptionalisierung des vorhersagenden informationsverarbeitenden Organs. Es ist die Logik des vermeintlich objektiven Wissenschaftlers, der den *fremden Kontinent Frau* (Freud) oder fremde Kolonien ›entdeckt‹, erforscht und verobjektiviert, der mithilfe seiner Vermessungstechnologien universal angelegtes Wissen über Subjekte generiert und dabei die eigenen Erfahrungen und Wissenskontexte nicht mehr in Betracht zieht. Die Verdatung des Unbewussten, das mathematische Modellieren des Unvorhersehbaren und Unverfügbaren führt zu einer neuen Subjektivität, zu einer neuen Wahrnehmung des Selbst. Die Einlagerung einer als besonders rational und objektiv angesehenen Mathematischen Logik, von Information und subjektloser Informationsübertragung in das Konzept Neuronaler Netzwerke, aber auch allgemein in Kommunikationstechnologien und in Konzepte der künstlichen Intelligenz ist so gesehen auch ein koloniales Projekt:

> While thinking and reason are identified with the male and Western subject, emotions and the body are associated with femininity and the racial other. This projection of emotions onto the bodies of others serves not

> only to exclude others from the realm of thought and rationality, but also to hide the emotional and bodily aspects of thought and reason. (Ahmed 170)

Die Wahrscheinlichkeitstheorie verhalf der Berechnung zufälliger Verteilungen zu ihrem Durchbruch und damit der Stochastik zum breiten Einsatz in den Modellen komplexer Systeme, auch in den Computational Neurosciences. Die Physikerin Sabine Hossenfelder erinnert daran, dass es sich bei den zugrunde liegenden Verteilungswerten keineswegs um eine mathematisch begründete Richtlinie handelt:

> Die Annahme einer gleichmäßigen Verteilung beruht auf dem Eindruck, dass es sich intuitiv um eine naheliegende Wahl handelt. Aber es existiert kein mathematisches Kriterium, aus dem sich diese Wahrscheinlichkeitsverteilung ergibt. Ja, jeder Versuch, sie abzuleiten, führt lediglich zu der Annahme zurück, dass am Anfang wiederum eine Wahrscheinlichkeitsverteilung vorzuziehen ist. Der einzige Weg, diesen Kreis zu durchbrechen, besteht darin, einfach eine Wahl zu treffen. (2019, 311)

Und so wird auch der Zufall, eingehegt durch die Annahme gleichmäßiger Verteilungswerte, zu einer mathematischen Konstante, die aus ›Chancen‹ berechenbare Wahrscheinlichkeitsverteilungen werden lässt. Aus den verschiedenen möglichen Ereignissen eines Wahrscheinlichkeitsspektrums wird das wahrscheinlichste Ereignis ausgewählt – je nachdem, welches Konzept von Wahrscheinlichkeit zugrunde gelegt wird. Durch die Projektion dieser Auswahl in die Zukunft gewinnt sie an Faktizität. In diesen Algorithmen sind die neuronalen Berechnungseinheiten Entscheidungssysteme, die darüber urteilen, welche Daten sich im Spektrum von richtig und falsch befinden, und die über diese Zufälligkeitsverteilungen aus Wahrscheinlichkeit Schicksal werden lassen. Denn die Berechnung der möglichsten Möglichkeit, also dessen, was höchstwahrscheinlich eintritt, legt die Grundlage für die weiteren Berechnungen.

Auch wenn es verlockend ist, kann man Denkprozesse nicht von ihrem Ende her, das heißt von der Potenzialität des Eintretens eines möglichen Ereignisses begreifen. Die Stochastik mit ihren Berechnungen wahrscheinlicher Zukunftsereignisse, basierend auf Daten aus dem Labor, folgt ihrer eigenen Logik und bietet vermeintlich objektive Lösungsvorschläge. Es handelt sich dabei aber um die wissenschaftliche Etablierung einer am wahrscheinlichsten anzunehmenden Modellierung der Zukunft. Sie macht uns als Individuen

und als Gesellschaft abhängig von ›unserer Natur‹ und durch die Aufhebung der Möglichkeit eines freien Willens auch handlungsunfähig. Die Logik der vorhersagenden Entscheidungsmechanismen übersetzt Ratio einmal mehr in den Bereich der mathematischen Aussagenlogik und führt zu der Idee, dass die mustererkennende und vorhersagende Maschine es potenziell besser weiß als der Mensch selbst.

Mustererkennung

Künstliche Neuronale Netzwerke werden mitunter auch in anderen neurowissenschaftlichen Methoden eingesetzt, insbesondere für die Mustererkennung und bei unvollständigen Datensätzen. Ein Beispiel hierfür ist die multivariate Mustererkennung. Die Methode beruht auf der funktionellen Magnetresonanztomografie (fMRT), einem bildgebenden Verfahren, das sich bei der Auswertung von Daten künstlicher Neuronaler Netzwerkalgorithmen bedient, um Aktivitätsmuster aus den Daten herauszulesen. Obwohl die fMRT nicht zum Bereich der Computational Neurosciences gehört, möchte ich hier dennoch kurz näher auf die aktuellen Forschungsergebnisse des Teams am Max-Planck-Institut für Human Cognitive and Brain Science in Berlin unter der Leitung von John Dylan Haynes eingehen. Grund hierfür ist, dass sie in ihrer Publikation die Mathematisierung von Wahrnehmung durchaus am Beispiel der Computational Neurosciences entwickeln, ihr Anspruch aber ist, darüber hinaus zu gehen und Mathematisierung in der Wissensproduktion allgemein zu zeigen. John Dylan Haynes und sein Team haben das Verfahren der multivariaten Mustererkennung so weit professionalisiert, dass es heute als »Gedankenlesen« (Haynes 2013), als »Mind Reading« oder als »Brain Reading« bekannt ist. »Beim Brain Reading […] werden Gedanken aus den Hirnaktivitätsmustern dechiffriert.« (Haynes/Eckholdt 2021, 85) Das Verfahren ›sucht‹ mithilfe künstlich neuronaler Algorithmen Muster in den Hirnscans funktioneller Magnetresonanztomografie. Denn

> [d]ie Information, die in den fMRT Bildern relevant ist, lässt sich nicht mit bloßem Auge erkennen. Wir müssen dafür Computern beibringen, nach dem Gedankencode im Gehirn zu suchen. […] Das Hirn erzeugt bei jedem Auftreten eines Gedankens jeweils ein präzises und wiederholbares Aktivitätsmuster. (Ebd., 75)

Das Verfahren trainiert Gehirne wie Algorithmen in mehrfach durchgeführten Trainingsdurchgängen. In dem Verfahren werden den Proband*innen

nacheinander verschiedene Gegenstände gezeigt, wie ein Hammer, eine Badewanne, Nägel etc., währenddessen werden individuelle Hirnmuster aufgezeichnet. Die funktionelle Magnetresonanztomografie scannt einzelne Schichten des Gehirns und speichert potenzielle Aktivität (also erhöhten Blutmagnetismus) als Grauwerte in dreidimensionalen Pixeln (Voxeln) ab. In diesen dadurch entstehenden Hirnkarten, so die Hoffnung, lassen sich spezifische Aktivitätsmuster finden, die das Sehen eines Hammers, einer Badewanne und der Nägel repräsentieren. Die Algorithmen scannen nun also die Hirnbilder und suchen nach repräsentativen Anordnungen von Grauwerten in den Schichten der Hirnscans, um potenzielle Muster auszumachen, die spezifisch für das Sehen eines Hammers sind. Unter der Überschrift »Prinzipien der Klassifikation« fassen die beiden Autoren der Studie zusammen:

> Nehmen wir vereinfacht an, wir wollten aus der Hirnaktivität erkennen, ob jemand an einen Hund oder an das Brandenburger Tor denkt, und wir hätten schon einige Beispiele der Hirnaktivität für diese beiden Gedanken gemessen. […] Nun kann man die gemessenen Aktivitäten in ein Koordinatensystem eintragen: den ersten Wert auf der x- und den zweiten Wert auf der y-Achse. Für den Gedanken an den Hund im Kopf des Probanden wären es in diesem schematischen Beispiel die Werte 2 und 8, für das Brandenburger Tor 7 und 3. Das wiederholen wir ein paar Mal und tragen alle Messungen der Hirnaktivität in das Koordinatensystem ein. Die Messpunkte für den Hund markieren wir als Kreise, die für das Brandenburger Tor als Kreuze. Wenn alle Voxel ausgewertet sind, schauen wir uns das Koordinatensystem erneut an. Im Idealfall sehen wir eine klare Trennung zwischen den Kreisen, die für »Hunde« und den Kreuzen, die für »Brandenburger Tor« stehen. Der Algorithmus der Mustererkennung lernt nun die Linie, mit der man diese beiden Punktwolken optimal trennen kann […]. Selbst wenn die Punktwolken sich etwas überlagern, findet der Algorithmus eine Trennungslinie, mit deren Hilfe er die Muster sehr effizient auseinanderhalten kann. (Ebd., 93)

Fehlerquoten werden bereitwillig eingeräumt, mehr noch, Fehler sind notwendig damit Mustererkennungsalgorithmen daraus lernen können. Im maschinellen Lernen sind Fehler gescheiterte Kommunikationsversuche und neben dem Einspeisen von Trainingsdaten, ein wichtiges Mittel um einen Algorithmus zu schulen. In der Einführung zur methodischen Herangehensweise wird darauf verwiesen, dass eine 100-prozentige Trefferquote nicht angestrebt wird und dass die statistische Validität bereits mit Trefferquoten, die

über 50 Prozent liegen, erreicht sei: »Wenn man beim Auslesen einer Hirnregion eine Trefferquote von 60 Prozent erreicht, dann heißt das zumindest, dass irgendwelche Informationen über die Gedanken in dieser Region stecken.« (Ebd., 95) Auch wird darauf verwiesen, dass es sich bei der Methode weniger um ein Mind Reading als um einen Wiedererkennungseffekt handelt:

> Das Prinzip, erst die Aktivitätsmuster zu messen und sie dann einem Computer beizubringen, funktioniert sehr gut. Allerdings können wir damit keine beliebigen Gedanken lesen, sondern nur jene, die der Computer zuvor gelernt hat. Bei Lichte besehen geht es hier also eher um ein Wiedererkennen. (Ebd., 94)

Die Methode der multivariaten Mustererkennung zeigt exemplarisch, wie Neuronale Netzwerkmodelle als operative Entscheidungssysteme eingesetzt werden, um Aktivitätsmuster in Hirnscans ausfindig zu machen. Neben der Vorstellung der Mind-Reading-Methode widmet sich das 2021 erschienene Buch von Haynes und Eckholdt den Themen freier Wille, Lügendetektoren und Neuromarketing. Diese Zusammenstellung ist nicht zufällig: Sie baut auf dem Argument auf, dass der Mensch nicht über einen freien Willen verfügt, dass die Mind-Reading-Methode dies beweise und dass die Maschine dadurch, dass sie in den Daten versteckte Gedanken findet, als Lügendetektor und für entsprechendes Neuromarketing eingesetzt werden könne.

Mit dem Beispiel möchte ich zweierlei zeigen. Erstens kommen in der multivariaten Mustererkennung verschiedene Herangehensweisen zusammen: die im MRT generierten funktionellen Daten, die mithilfe stochastischer Berechnungen, heute Neuronalen Netzen, analysiert werden, um den Daten wiederkehrende Muster abzuringen. Zweitens wird hier einmal mehr die Deutungshoheit neurowissenschaftlicher Erklärungen aufgewertet durch Auswertungsverfahren, die allein mit rechenstarken Computern getätigt werden können, die aus einem Istzustand der Gehirnaktivität immerwährende und in die Zukunft gerichtete Verhaltensweisen herauslesen sollen. Das sogenannte Mind Reading geht konkret davon aus, dass die aus dem Gehirn ausgelesenen Daten vor allem die weniger bewussten Vorgänge eines Menschen offenlegen und damit einen Zugang zum Unbewussten ermöglichen.

Über diesen vermeintlich direkten Zugang zum Unbewussten stellt das Mind Reading den freien Willen infrage und postuliert die Annahme, dass die gewonnen Hirndaten mehr über den Menschen preisgeben, als diesem selbst bewusst sei: »Diese Experimente könnten ein Indiz dafür sein, dass der Hirn-

scanner unter bestimmten Bedingungen mehr Informationen bereitstellt als die Beobachtung des Verhaltens.« (Haynes/Eckholdt 2021, 219)

Die Annahmen der Brain-Reading-Methode führen auch dazu, Algorithmen als funktionsgleiches Pendant zu morphologisch neuronalen Netzwerken zu stilisieren. Das heißt, der Einsatz artifizieller Netzwerke in der multivariaten Mustererkennung suggeriert eine funktionelle Ähnlichkeit und leistet damit einer Essenzialisierung der mathematischen Modelle Vorschub. Besonders eindrücklich zeigt dies eine Brain-Reading-Studie, die Shinji Nishimoto und Kolleg*innen aus dem Labor von Jack Gallant an der University of California in Berkeley mit Bewegtbildern durchführten. Die Methode, mit der die Aktivitätsmuster statischer Bilder aus Hirnscans ›ausgelesen‹ werden können, lässt sich bis zu einem gewissen Grad auch auf bewegte Bilder anwenden:

> Dafür muss der Computer nach einem ähnlichen Prinzip wie bei statischen Bildern lernen, wie sich bewegte Bilder im Gehirn abbilden. Ein Film besteht aus der einer Abfolge rasch hintereinander gezeigter Einzelbilder. Jedes Bild bleibt für ca. 16 bis 50 Millisekunden stehen, dann kommt das nächste. […] Entsprechend verarbeitet das Gehirn solche Filme als Bewegungsmuster. Um Filme zu decodieren, muss man dem Computer also beibringen, viele einfache Bewegungsformen zu sehen, ähnlich den einzelnen Bildpunkten im vorherigen Beispiel. (Haynes/Eckholdt 2021, 138)

Im Rahmen der Studie wurden den Probanden einige YouTube-Videos gezeigt und ihre Hirnaktivität mittels eines Magnetresonanztomografen gemessen. Danach wurde ein Algorithmus programmiert, der in den Hirnscans nach Mustern einfacher Bewegungsfolgen suchte.

> Im nächsten Schritt lernte der Computer eine Reihe von YouTube-Filmen zu decodieren. Es wurde getestet, ob er in der Lage war, auch neue Filme – also welche, deren Aktivitätsmuster nicht zuvor anhand der Probanden gespeichert wurden – richtig auszulesen. Dabei verwendeten sie einen spannenden und innovativen »Mischungsansatz«: Der Computer versuchte, die Bewegungen zu rekonstruieren, indem er diejenigen Videos zusammenmischte, die zur Hirnaktivität der Probanden am besten passten. (Ebd., 138)

In der Studie hatten Wissenschaftler*innen die Daten der funktionellen Hirnscans mit multivariater Mustererkennung nach Mustern davon durchsucht, was sie den Proband*innen im MRT gezeigt hatten: Das heißt, aus den Mus-

tern in den Hirnscandaten wurden Bilder herausgelesen, die dann wiederum, stark verrauscht, zu einem Video zusammengeschnitten wurden.

Insbesondere das letzte Beispiel offenbart, was hier für die Analyse zusammengeführt wird: Daten aus funktioneller Magnetresonanztomografie mit Deep-Learning-Algorithmen, die nach Mustern suchen. Und hier zeigt sich auch ein weiteres Feld, in dem Deep-Learning-Algorithmen (oder auch künstliche Neuronale Netzwerke) Anwendung und Verbreitung erfahren. Der gesamte grafische und filmische Sektor hat fundamental von der Bereitstellung dieser neuen ›selbstlernenden‹ Algorithmen profitiert: von der Pinselspitze bei Photoshop, die über die zufällige Verteilung des Aufzutragenden bestimmt (wie etwa die Verteilung der Tomaten, die sich auf dem Cover dieses Buches befinden), über Programmierungen beziehungsweise Animationen von Filmsequenzen bis hin zur Manipulation ganzer Bewegtbildsequenzen, auch als *deep fakes*[1] bezeichnet. Die Vermittlungsschritte eines Unbewussten existieren in einer Welt des Mind Reading, der Mustererkennung und Vorhersage kaum mehr, denn mit den neuen Technologien und Modellen lässt sich angeblich direkt erkennen, was Menschen sich wünschen, was sie wollen und denken. Das Ableiten von Hirndaten stellt demnach eine Abkürzung zu den vermeintlichen Gedanken und Wünschen des Menschen dar und damit eine klare Positionierung gegen die Möglichkeit eines ›freien Willens‹.

Die vorhersagende Maschine behauptet, berechnen beziehungsweise auslesen zu können, was der Mensch ›wirklich‹ denkt. Durch diese Reduktion menschlicher Denkprozesse, selbst der heimlichsten Wünsche, die hier äußerst wirksam mit *unbewussten* Wünschen gleichgesetzt werden, auf wenige Muster im Gehirn, wird das zu Wissende und zu Fühlende in die Welt der Mathematischen Logik und Berechenbarkeit verlagert. Was aber macht diese Überwältigungsthese, die »Gehirne wüssten es besser als der Mensch selbst«, mit unserer Vorstellung von mündigen Subjekten?

1 Mithilfe Neuronaler Netzwerkalgorithmen können als *deep fakes* bezeichnete Videos hergestellt werden: Videofrequenzen, in denen der Mund und die Stimme von Menschen so graphisch animiert werden, dass sie Aussagen treffen, die sie nie gesagt haben.