

Auf welche Weise können wir die Welt erkennen?

Ulrich Eckhardt

Inhaltsverzeichnis

1	Zur Person	1
2	Anwendungen von Modellen der visuellen Wahrnehmung	2
3	Form	4
4	Das menschliche Wahrnehmungssystem	6
5	Modelle der Wahrnehmung	13
5.1	Perzeptrons	14
5.2	Neuronale Netzwerke	18
5.2.1	Der Satz von Kolmogorow	24
	Literatur	25

Abbildungsverzeichnis

1	Das Symbol „Herz“ in moderner Sicht (rechts) und in der Sichtweise eines alten Ägypters der 21. Dynastie (links; bâ-Hieroglyphe)	6
2	Schematisches Modell der Wahrnehmung	7
3	Schematische Darstellung des visuellen Systems	9
4	Das Kanizsa-Dreieck.	10
5	Schema zur Informationsverarbeitung im Menschen	11
6	Reduktion eines Bildes auf „wesentliche“ Merkmale	12
7	Perzeptron.	14

Vortrag, gehalten am 21. April 2001 auf dem Halbjahres-Meeting der innovas GmbH, „Maritim Golf- & Sport-hotel“ Timmendorfer Strand.

8	Testbilder nach Minsky und Papert	16
9	Das Heckenlabyrinth von Chevening	19
10	Die „American Bankers Association E-13 B Fonts“.	20
11	Die Ziffern 1, 3 und 8 der American Bankers Association E-13 B Fonts.	21
12	Beispiel eines neuronalen Netzes mit zwei „verborgenen“ Schichten.	23

Tabellenverzeichnis

1	Die vier Möglichkeiten für die relevanten Sensoren	17
---	--	----

1 Zur Person

Ich sollte gleich am Anfang sagen, daß ich ein „angewandter“ Mathematiker bin, kein „reiner“. Das heißt, daß meine Forschungsgegenstände von außen, durch die Anforderungen von Anwendern bestimmt sind. Etwas drastischer ausgedrückt: Der angewandte Mathematiker verhält sich zum reinen wie der Gourmet zum Biologen: Während dieser das Reh in seinem Biotop beobachtet und dessen Verhalten studiert, ist jener nur daran interessiert, wie man das Tier zubereitet. Gelegentlich pflege ich auch zu sagen, daß der angewandte Mathematiker ein Symbiont — etwas drastischer: ein Schmarotzer — ist. Er bedarf eines Wirts zum Überleben.

Diese Situation hat Konsequenzen. Beispielsweise kann ich mich niemals „spezialisieren“. Mein Arbeitsgebiet ist offiziell „Angewandte Funktionalanalysis“, praktisch bin ich für Optimierung zuständig. Ich kann einen anrufenden Ingenieur, der mich um Rat fragt, nicht darauf verweisen, daß ich nicht zuständig sei. Ich kann ihn auch nicht an den zuständigen Kollegen verweisen, denn ein Ratsuchender schätzt es nicht, „herumgereicht“ zu werden, und dann wird in vielen Fällen der angesprochene Kollege den Ratsuchenden „abwimmeln“, denn Antworten an fragende Ingenieure kosten Zeit und sind der Karriere nicht förderlich.

Wie gesagt, ich bin nicht spezialisiert, ich bin eher ein universeller Dilettant. Diese Situation schätze ich ungemein, ich bin eben neugierig, und es gefällt mir, in verschiedenen Gebieten zu „wildern“. Dafür muß ich in Kauf nehmen, daß mich alle Kollegen für allenfalls drittklassig halten. In meinem Alter hat man auf einem Gebiet der Papst zu sein, ich bin allenfalls Wanderprediger!

Es war nie meine Absicht, Mathematiker zu werden, ursprünglich hatte ich Astronomie studiert, dann habe ich meine Diplomarbeit in Mathematik geschrieben, fünf Jahre an einem wirtschaftswissenschaftlichen Institut gearbeitet, danach acht Jahre in einer Großforschungseinrichtung, der Kernforschungsanlage Jülich, und schließlich, als der jugendliche Elan nachließ, als Mathematiker an der Universität.

Um einen kleinen Eindruck zu geben, mit welchen nichtmathematischen Themen ich mich beschäftigt habe, zähle ich einfach einmal ein paar Gebiete auf in denen ich ein wenig gearbeitet habe (beziehungsweise wissenschaftliche Zeitschriften, in denen ich publiziert habe).

- Cytologie (*Cytometry* 1981),
- Ingenieurwissenschaften (*Ingenieur-Archiv* 1982),
- Festkörperphysik (*Physical Review B* 1973).
- Urheberrecht (Gutachten „Baustatikprogramm“, OLG Frankfurt 1984),
- Wirtschaftswissenschaften (Gutachten über Verkehrswertermittlung 1997),
- Informationstechnologie (insbesondere zwei Erfindungsmeldungen),
- Softwareentwicklung (für kommerzielle Anwendungen).
- Theologie.

Damit will ich nicht demonstrieren, was für ein verflucht interdisziplinärer Kerl ich bin, man kann dies auch anders interpretieren: Ich stecke meine Nase vorwiegend in Angelegenheiten, die mich nichts angehen!

In den letzten zwei Jahrzehnten habe ich mich vorwiegend mit mathematischer Modellierung von Wahrnehmungsvorgängen befaßt. Dies hat verschiedene Gründe, einmal fand ich einen Wirtskörper am Siemens Forschungslaboratorium in München, dann gibt es auf diesem Gebiet zahlreiche äußerst spannende und neuartige mathematische Probleme und schließlich ist dieser Forschungsgegenstand philosophisch hochinteressant. Es hat nämlich mit der Grundfrage der Erkenntnistheorie zu tun: „Können wir die Welt erkennen, und wenn ja, wie?“ Seit etwa fünf Jahren arbeiten wir an einem von der Deutschen Forschungsgemeinschaft geförderten Projekt „Formrepräsentation in diskreten Strukturen“. Dieses hat gute diplomatische Beziehungen zu dem DFG-Forschungsschwerpunkt „Raumkognition“.

2 Anwendungen von Modellen der visuellen Wahrnehmung

Ich skizziere einige Anwendungen von Modellen für die visuelle Wahrnehmung. Nicht an allen Anwendungen habe ich selbst mitgearbeitet. Wie dies bei forschungsorientierten Arbeiten typisch ist, sind die meisten der Aktivitäten, an denen ich beteiligt war, nicht zur Produktreife gediehen. Man stellt beispielsweise nach zwei Jahren Forschung fest, daß die ursprünglich intendierte Anwendung nicht wie gedacht funktioniert (Die Ergebnisse der Forschung sind eben — und dies scheint man in der Forschungspolitik heute nicht mehr zu wissen — nicht vorhersehbar, sonst wäre es eben nicht Forschung, sondern Entwicklung), oder die Geschäftsleitung hat ihre Geschäftsziele geändert, weil sich der Markt geändert hat, oder die Politik gibt neue Vorgaben, die das Forschungsprojekt obsolet oder unerwünscht werden lassen, oder man stellt fest, daß man sich bezüglich der Kosten verkalkuliert hat, oder die Firma, die den Auftrag erteilt hat, stellt ihre Zahlungen ein oder

Die Anfänge der Beschäftigung mit Modellen der Wahrnehmung reichen bis in den Anfang der sechziger Jahre zurück. Damals wurden in den USA und auch in der Sowjetunion zahlreiche Forschungsarbeiten auf dem Gebiet der Zeichenerkennung (Pattern Recognition) unternommen. Das erste Buch über das Gebiet, das Buch von Frank Rosenblatt über „Principles of Neurodynamics“ erschien bereits 1962 [15]. Das sowjetische Pendant, das Buch von Arkadjew und Brawerman über *Zeichenerkennung und maschinelles Lernen*, erschien in deutscher Übersetzung im Jahre 1966 [1]. 1965 publizierte Nilsson ein Buch über lernende Maschinen [13]. In diesen Jahren — und dies geschieht ja heute auch noch — sparte man nicht mit vollmundigen Versprechungen und euphorischen Visionen. Ein Motiv für die intensiven Forschungsbemühungen in den USA war auch die Hoffnung, daß man in absehbarer Zeit automatisch russische Publikationen lesen und auch automatisch übersetzen könne. Möglicherweise hat man in der Sowjetunion ähnliche Hoffnungen gehegt. Im Jahre 1971 erschien im *Scientific American* ein Artikel von Casey und Nagy, in dem der damalige Stand dargestellt wurde [4]. Dort hieß es dann, nach einem Jahrzehnt Ernüchterung:

... progress in the field has not fulfilled many of the sanguine predictions made a decade ago. Even the simpler problems turned out to be a good deal more difficult than had been expected and the “universal” methods pattern recognition proposed at that time could not provide useful results with the computational resources available either then or now.

Es sollte noch sehr lange dauern, bis aus den ersten Prototypen für lesende und erkennende Maschinen Produkte wurden, die problemlos eingesetzt werden konnten. Auch heute noch steht die „Universalmaschine“ aus, man hat aber eine große Anzahl von Anwendungen gefunden, in

denen sich Maschinen, die visuelle Erkennung in einem wohldefinierten Umfeld mit hinreichender Zuverlässigkeit leisten, bewährt haben. Ein weites Anwendungsgebiet fanden Maschinen, die gedruckte oder maschinengeschriebene Texte hinreichender Qualität lesen können. Man spricht von OCR-Anwendungen (**O**ptical **C**haracter **R**ecognition). Hierzu gehören Geräte, die Text mit einer ausreichenden Zuverlässigkeit in ASCII-Daten verwandeln, Systeme, die mit hoher Geschwindigkeit und hoher Präzision Spezialzeichen auf Bankschecks lesen, Maschinen, die Briefe mit maschinengeschriebenen Postleitzahlen automatisch sortieren oder Systeme, die Formulare auswerten, sind heute schon weithin im Gebrauch. Es gibt jedoch auf diesem Gebiet zahlreiche Desiderata, ich denke dabei nicht an die Auswertung der Stimmzettel bei US-Wahlen! Die Erkennung von handgeschriebener Schrift ist noch ein weithin unlösbares Problem, und wer in seinen Palmtop mit der Hand Zeichen einschreibt, freut sich schon, wenn diese auch richtig interpretiert werden, und Voraussetzung scheint zu sein, daß sich der Schreibende an den Rechner anpaßt, und nicht umgekehrt.

Ich nenne einige Problemfelder, in denen es darum geht, visuelle Informationen automatisch zu verarbeiten:

- Büroautomatisierung. Ein Engpaß ist immer noch die Schnittstelle Papier — Rechner. Angesichts der zahlreichen schwer miteinander vereinbaren Datenformate auf verschiedenen Rechnern hat man sogar gelegentlich Daten über die „Papierschnittstelle“ konvertiert.
- Hinreichend autonome Roboter müssen visuelle Informationen schnell und zuverlässig interpretieren können. Man denke etwa an Haushaltsroboter oder aber an robotische Steuerungshilfen für Kraftfahrzeuge.
- Analyse von Dokumenten. Hier gibt es zahlreiche sehr unterschiedliche Anwendungen:
 - Sortieren des Posteingangs (ein Prototypsystem, welches dieses leistet, wurde im Siemens-Forschungslabor getestet).
 - Analyse von technischen Zeichnungen. Ein großer Teil aller technischen Zeichnungen, Architekturzeichnungen usw. befindet sich heute noch auf Papier. Es wäre sehr wünschenswert, wenigstens einen Teil dieser Dokumente automatisch auf CAD-Files übertragen zu können. Ein nicht sehr euphorisches Resümee eines langjährigen Forschungsvorhabens auf diesem Gebiet gaben Hofer-Alfeis und Maderlechner im Jahre 1988 [7].
 - Ein wichtiges Thema wäre die Analyse von „alten“ Dokumenten, etwa von alttibetischen Texten, Hieroglyphen usw. Hier gibt es eine Anzahl von Anwendungen, man muß allerdings feststellen, daß diese Anwendungen zwar hochinteressant sind, daß sie aber kaum je für den „Markt“ von Interesse sein dürften.
- In der medizinischen Diagnostik besteht das typische Problem darin, in den vorhandenen Daten ein gewisses „Muster“ zu erkennen, welches auf eine Krankheit schließen läßt. Insbesondere wäre es nützlich, dem Arzt automatisch Diagnosehilfen auf der Grundlage etwa einer automatischen Analyse von Röntgenaufnahmen anzubieten. Dieses Gebiet wird uns noch auf viele Jahrzehnte beschäftigen.
- Ein wichtiges Gebiet ist die Klassifikation, seien es eingehende Postdokumente (Geschäftsbrief, interner Bericht, Formular, Werbesendung, ...) oder aber Pflanzen und Tiere.
- Für den Informatiker, aber auch für den Erkenntnistheoretiker ist es eine hochinteressante Frage, welche Erkenntnisleistungen wir mittels Computern nachahmen können — und welche nicht. Man denke hierbei etwa an den bekannten Turing-Test! [18]

- Durch das Internet ergibt sich eine neue Dimension der Verarbeitung visueller Information. Ein Thema wäre die Suche nach Formen in Bildsequenzen (in welcher Aufzeichnung von Fußballspielen berührt der Fuß des Spielers X den Ball?), aber auch die Erkennung und Verarbeitung von Formmerkmalen wird eine immer wichtige Rolle spielen. Für jedes in den letzten Jahren gefertigte Produkt — vom Plüschteddy bis zur Kaffeekanne — gibt es irgendwo ein CAD-File, welches das Produkt genau beschreibt. Solche Files werden heute schon — beispielsweise in der Werbung — genutzt. Es wäre nützlich, wenn man direkt suchend auf derartige Datensätze zugreifen könnte.

Ein wichtiger Begriff bei der visuellen Wahrnehmung ist der der Form. Es handelt sich um einen mathematisch nicht ausreichend präzisierbaren Begriff. Wir wollen uns kurz mit diesem Begriff befassen.

3 Form

Schägt man in einem Lexikon unter dem Stichwort „Form“ nach, dann ist man überwältigt von den zahlreichen Aspekten dieses Begriffs. Die Bedeutung dieses Wortes variiert, je nachdem, ob ein Philosoph, ein Künstler, ein Biologe oder ein Kind im Sandkasten das Wort „Form“ benutzt. Wenn wir uns nicht im Uferlosen verlieren wollen, müssen wir uns also einschränken. Ich denke hier — immer noch nicht mathematisch präzise — an die Form, in der sich Objekte der Umwelt unserem visuellen System darbieten. In diesem Verständnis des Begriffs arbeiten wir im Rahmen eines von der deutschen Forschungsgemeinschaft geförderten Projekts „Formrepräsentation in diskreten Strukturen“. Dieses Projekt hat Bezüge zu dem Forschungsschwerpunkt „Raumkognition“, der ebenfalls von der DFG gefördert wird.

Eine formbezogene Sprache spielt insbesondere in den Lebenswissenschaften eine große Rolle. Die Sprache der Formen ist die Sprache der Biologie, und es gibt bereits eine umfangreiche Literatur über mathematische Beschreibung von Form und deren Veränderung in lebenden Systemen. Neben dem Klassiker d’Arcy Wentworth Thompson [17], dessen Buch 1917 erschien und seitdem zahlreiche Auflagen und auch mehrere Übersetzungen ins Deutsche erlebt hat und dem Buch von Lord und Wilson [11] finden wir hier erlauchte Namen wie den des genialen Mathematikers Alan Turing, der sich mit mathematischen Modellen für die Entstehung und Veränderung von Formen befaßte [19]. Ein moderneres Buch über die statistische Theorie der Form stammt von Dryden und Mardia [5], ein sehr tiefsinniger, aber auch sehr schweriger Zugang findet sich in der „Shape Theory“ des Topologen Karol Borsuk [3] Wichtige und praktisch bedeutsame Beiträge hat die „Mathematische Morphologie“ der Pariser Schule um Jean Serra [16] geliefert.

Bekannt ist, daß auch Goethe sich mit der Theorie der biologischen Form beschäftigte. Tatsächlich hat er den Fachbegriff „Morphologie“ erstmalig benutzt. er schreibt in seinen „Vorarbeiten zur Morphologie“ (1788/89):

Bei der fortschreitenden Veränderung der Pflanzenteile wirkt eine Kraft, die ich nur uneigentlich Ausdehnung und Zusammenziehung nennen darf.

Besser wäre es, ihr ein x oder y nach algebraischer Weise zu geben, denn die Worte Ausdehnung und Zusammenziehung drücken diese Wirkung nicht in ihrem ganzen Umfange aus. Sie zieht zusammen, dehnt aus, bildet aus, bildet um, verbindet, sondert, färbt, entfärbt, verbreitet, verlängert, erweicht, verhärtet, teilt mit, entzieht und nur allein, wenn wir alle ihre verschiedenen Wirkungen in einem sehen, dann können

wir das anschaulicher kennen, was ich durch diese vielen Worte zu erklären und auseinanderzusetzen gedacht habe. Sie tut das alles so stückweise, so sachte, so unmerklich, daß sie zuletzt uns vor unseren Augen einen Körper in den andern verwandelt, ohne daß wir es gewahr werden.

Interessant ist, daß Goethe, der sich der Mathematik gegenüber durchaus zwiespältig verhielt, hier anmerkt, daß man „eigentlich“ die Morphologie „nach algebraischer Weise“ anpacken müßte, aber er wußte genau, daß Mathematik seine Sache nicht war. Es ist höchst fraglich, ob Goethe mit seinem „ x und y “ das gemeint hat, was heute von der Mathematik auf diesem Gebiet angeboten wird.

Um einen moderneren Autor zu zitieren: In einem sehr bekannten Artikel aus dem Jahre 1967 schrieb Harry Blum¹:

I have approached the problem of shape by assuming that the current mathematical tools were somehow missing essential elements of the problem. For despite more than two millenia of geometry, no formulation which appears natural for the biological problem has emerged.

Im Verlauf seines Artikels stellt man dann allerdings fest, daß Blum wieder einmal „das Rad neu erfunden hat“.

Bei der Beschreibung biologischer Form bedient man sich neben anderen „Nichtstandarddeskriptoren“ ausgiebig der Sprache der Formen. Ein Beispiel: In dem bekannten botanischen Bestimmungsbuch von Schmeil-Fitschen² findet man für eine bestimmte Art von Rosen die Beschreibung:

Stacheln meist schlank, kräftig, am Grunde nicht auffallend verdickt (Ansatzfläche etwa 3–14 mm lg. u. 2–3 mm br.); Blättch. eiförm. od. elliptisch. ziemlich dünn, kahl od. seidig behaart, Bltn. einzeln od. in Doldenrispen; Bltnstiele bis 3mal so lang wie der Bltnbecher; Bltnblätt. etwa $2-2^{1/2}$ cm lg., meist hellrosa; äußere Kelchblätt. fiederspalt., mit wenigen schmalen Fiedern. ♣ 6. Waldränder, Hecken, Weiden
Hecken-R., **R. canina L.**

In den Kriminalromanen von Agatha Christie wird Hercule Poirot wie folgt beschrieben³

„... ein kleiner, älthlicher Mann mit großem Schnurrbart und eiförmigem Schädel.“

Würde man jetzt den Schluß ziehen, daß Hercule Poirots Schädel die Form eines Heckenrosenblattes habe, dann sieht man, daß dieses Formmerkmal sich nicht transitiv überträgt. Mehr noch, Formmerkmale sind sehr stark kulturabhängig. Unter dem Begriff „herzförmig“ versteht ein moderner Europäer die in Abbildung rechts dargestellte Form. Ein alter Ägypter würde hingegen

¹Harry Blum: A transformation for extracting new descriptors of shape. In W. Wathen-Dunn, editor, *Models for the Perception of Speech and Visual Form*, pages 362–380, Cambridge: M.I.T. Press 1967. (In: J. K. Aggarwal, Richard O. Duda and Azriel Rosenfeld, editors, *Computer Methods in Image Analysis*, pages 153–171, The Institute of Electrical and Electronic Engineers, Inc., New York, 1977).

²Otto Schmeil: *Flora von Deutschland und angrenzender Länder ein Buch zum Bestimmen der wildwachsenden und häufig kultivierten Gefäßpflanzen*, 89., neubearb. und erw. Aufl. von Karlheinz Senghas und Siegmund Seybold. Quelle & Meyer, Heidelberg u.a., 1993.

³Agatha Christie, *Tod in den Wolken*, (*Goldmanns Taschen-Krimi, Band 4*), Wilhelm Goldmann Verlag München, 1959.

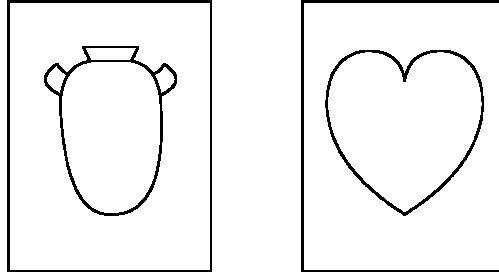


Abbildung 1: Das Symbol „Herz“ in moderner Sicht (rechts) und in der Sichtweise eines alten Ägypters der 21. Dynastie (links; bà-Hieroglyphe)

das links dargestellte Gebilde als „herzförmig“ bezeichnen, zumindest seit der 21. Dynastie taucht dieses Gebilde in dieser Bedeutung in zahlreichen Totenbuchpapyri auf. Kurios ist, daß der alte Ägypter genausogut wußte, wie dies der moderne Europäer weiß, daß das Herz selbst nicht herzförmig ist!

4 Das menschliche Wahrnehmungssystem

Bevor ich mich — im zweiten Teil meines Vortrages — mit einzelnen ausgewählten Anwendungen befasse, scheint es angebracht, sich zunächst einmal dem „Endabnehmer“ aller visuellen Information, dem menschlichen Wahrnehmungssystem, zuzuwenden.

Wir Naturwissenschaftler und Techniker lieben es, die Dinge auf eine Weise zu vereinfachen, daß sich den Lebenswissenschaftlern oder Philosophen das Nackenfell sträubt. Ganz naiv stellt man sich den Wahrnehmungsvorgang wie in Abbildung 2 vor. Der Mensch nimmt über Sensoren aus der Umwelt Signale auf, und diese werden irgendwo im Langzeitgedächtnis gespeichert. Die Summe dieser Informationen bezeichnet man als „Erfahrung“.

Ganz so einfach ist es sicher nicht! Wie schon Platon in seinem Höhlengleichnis⁴ andeutete, können wir das „Wesen“ der Dinge nicht erkennen, wir sehen allenfalls „Schatten“.

Eine noch radikalere Theorie schlug der englische Mathematiker, Philosoph und Theologe George Berkeley (1685–1753) vor. Er hat sich in drei wichtigen Punkten gegen Newton beziehungsweise gegen die Newtonianer gewandt, er wollte den Newtonschen „absoluten Raum“ nicht akzeptieren, er hielt die Fluxionenrechnung Newtons für mathematisch nicht einwandfrei begründet und schließlich wandte er sich gegen die Versuche der Newtonianer, die visuelle Wahrnehmung auf die geometrische Optik zu reduzieren. Heute würde jeder Schüler Berkeley in allen drei Punkten recht geben. Berkeleys erstes wissenschaftliches Werk war den Problemen der visuellen Wahrnehmung gewidmet [2] und erschien 1709 (Newton war zu dieser Zeit bereits 66 Jahre alt). Berkeley wies anhand zahlreicher Phänomene nach, daß das Auge eben keine „optische Maschine“ sei. Man muß sagen, daß die Hypothesen über insbesondere die Entfernungsschätzung zu dieser Zeit — in der ersten Euphorie nach der 1704 erfolgten Veröffentlichung von Newtons *Opticks* — noch reichlich naiv waren. Ein wichtiger Punkt in der Argumentation von Berkeley war die „Mondillusion“. Man meint nämlich, daß der Mond am Horizont größer sei, als im Zenit. Die Diskussion über dieses Phänomen hält heute noch an. Erst im letzten Jahre erschien wieder ein Artikel zu diesem

⁴Platon, *Der Staat (Politeia)*, Siebentes Buch, 514 ff.

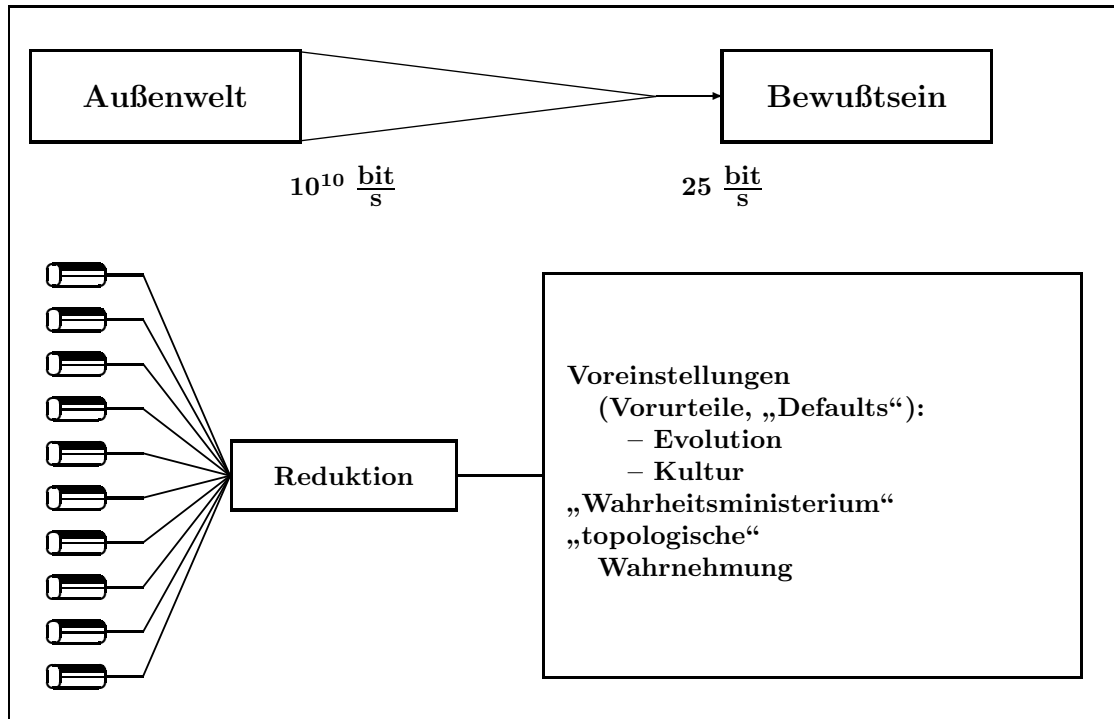


Abbildung 2: Schematisches Modell der Wahrnehmung

Die Außenwelt wird über Sensoren wahrgenommen. Über verschiedene Stationen (Kasten „Reduktion“ wird ein Bild der Außenwelt in das Bewußtsein übertragen

Thema⁵. Wie Berkeley — aus heutiger Sicht zutreffend — erkannt hat, findet der Prozeß des Wahrnehmens vorwiegend nicht im Auge statt, sondern im Gehirn, im Bewußtsein⁶. Berkeley hatte sich ausgiebig mit der Theorie der Wahrnehmung befaßt, ehe er seine philosophischen Arbeiten schrieb, in der er die „materialistische Hypothese“ schlichtweg leugnete. Nach Berkeley ist es eine überflüssige Verschwendung, wenn die Natur, die ja sonst so überaus sparsam ist, zwei Parallelwelten schafft, nämlich die „reale“ Außenwelt und die im Bewußtsein gespiegelte Welt. Er leugnete daher radikal, aber konsequent, daß die Außenwelt existiere!

Es ist eines der zahlreichen Kuriosa der Geschichte, daß Lenin in seinem Buch „Materialismus und Empiriokritizismus“ Berkeley heftig polemisierend angriff. Es ist schon sehr bemerkenswert, daß Berkeley durch die Auseinandersetzung mit der Natur, durch das Experiment, zu der Theorie des Immaterialismus geführt wurde, und daß Lenin, der niemals ein physikalischens Experiment ausgeführt hat, Berkeley im Namen der Naturwissenschaft und des Materialismus allein durch Bezug auf die „Klassiker“ angriff. Wenn es je ein schönes Beispiel für Dogmatismus gegeben hat: Hier ist es!

Das menschliche visuelle System wurde im Lotteriespiel der Evolution mit einer Anzahl von Defekten „begabt“. Das Auge selbst ist kein ideales optisches Instrument. Helmholtz soll einmal geäußert haben, das menschliche Auge sei, optisch gesehen, ein so unvollkommenes Gerät, daß er es einem Optiker zurückgeben würde. Berkeley hatte insofern recht, als das, was in diesem „unvollkommenen Gerät“ auf der Netzhaut dargestellt wird, keinesfalls das ist, was wir „sehen“. Vielmehr wird in den zahlreichen „Schaltstationen“ des Wahrnehmungssystems (siehe Abbildung 3, nach [10]) aus der optischen Information auf komplizierte und zu einem erheblichen Teil noch nicht verstandene Weise das erstellt, was wir als Bild von der Außenwelt empfinden.

Als einfaches Beispiel betrachten wir eine „subjektive Kontur“ nach G. Kanizsa. In Abbildung 4 sieht man in der Mitte des Bildes deutlich ein weißes Dreieck, welches im Original sicher nicht vorhanden ist. Es lohnt sich schon, ein wenig über diese Täuschung nachzudenken. Zum ersten handelt es sich hier um ein Beispiel dafür, daß man etwas deutlich wahrnimmt, was es nicht gibt. Weiterhin stellt sich die Frage, wie das Gehirn diese Illusion hervorbringt. Offenbar müssen die Informationen, die auf verschiedene, weit voneinander entfernte Teile der Netzhaut fallen, „irgendwie“ wieder zusammengebracht werden. Mehr noch: Wie kommt unser Wahrnehmungssystem dazu, gerade Linien so perfekt wahrzunehmen? Welches ist der evolutionäre Vorteil, gerade Linien als solche erkennen zu können? Sicher hatten unsere geschwänzten baumbewohnenden Vorfahren kaum Gelegenheit, irgendwo gerade Linien vorzufinden.

Eine weitere Eigenschaft des visuellen Systems ist noch viel bestürzender. Die Hauptschwierigkeit beim visuellen Wahrnehmen besteht darin, daß die Datenleitungen in unserem visuellen System sehr geringe Kapazität haben. Man kann recht gut angeben, was das visuelle System von der Außenwelt aufnimmt, etwa beim Fernsehen weiß man sogar ziemlich präzise, wieviel dem Auge angeboten wird. Beispeilweise beim HDTV sind dies einige 100 MBit/s. Im Auge hat man ca. 120 Millionen lichtempfindliche Zellen, aber nur etwa eine Million Sehnervenfasern, das heißt,

⁵Lloyd Kaufman and James H. Kaufman: Explaining the moon illusion, *Proceedings of the National Academy of Sciences*, 97(1):500–505, January 4, 2000.

⁶In seinem 1990 erschienenen Buch *Mentopolis* karikiert der bekannte Pionier der „künstlichen Intelligenz“, Marvin Minsky die Situation so:

Unsere Sinnesmaschinerie sendet Informationen an das Gehirn, wo sie auf eine Art innerer Filmleinwand projiziert werden. In diesem geisterhaften Kino betrachtet ein lauerndes Selbst die Szene; es handelt, indem es Signale zurück durch eine Serie von Fernsteuerungsmechanismen schickt. Aber dieses Konzept erklärt nichts, denn das Selbst bedürfte zur Erklärung seiner Wahrnehmungen einer weiteren Leinwand in seinem Inneren und so weiter.

(Zitiert nach Sebastian Stier, *Spektrum der Wissenschaft*, Mai 1992:138,140.)

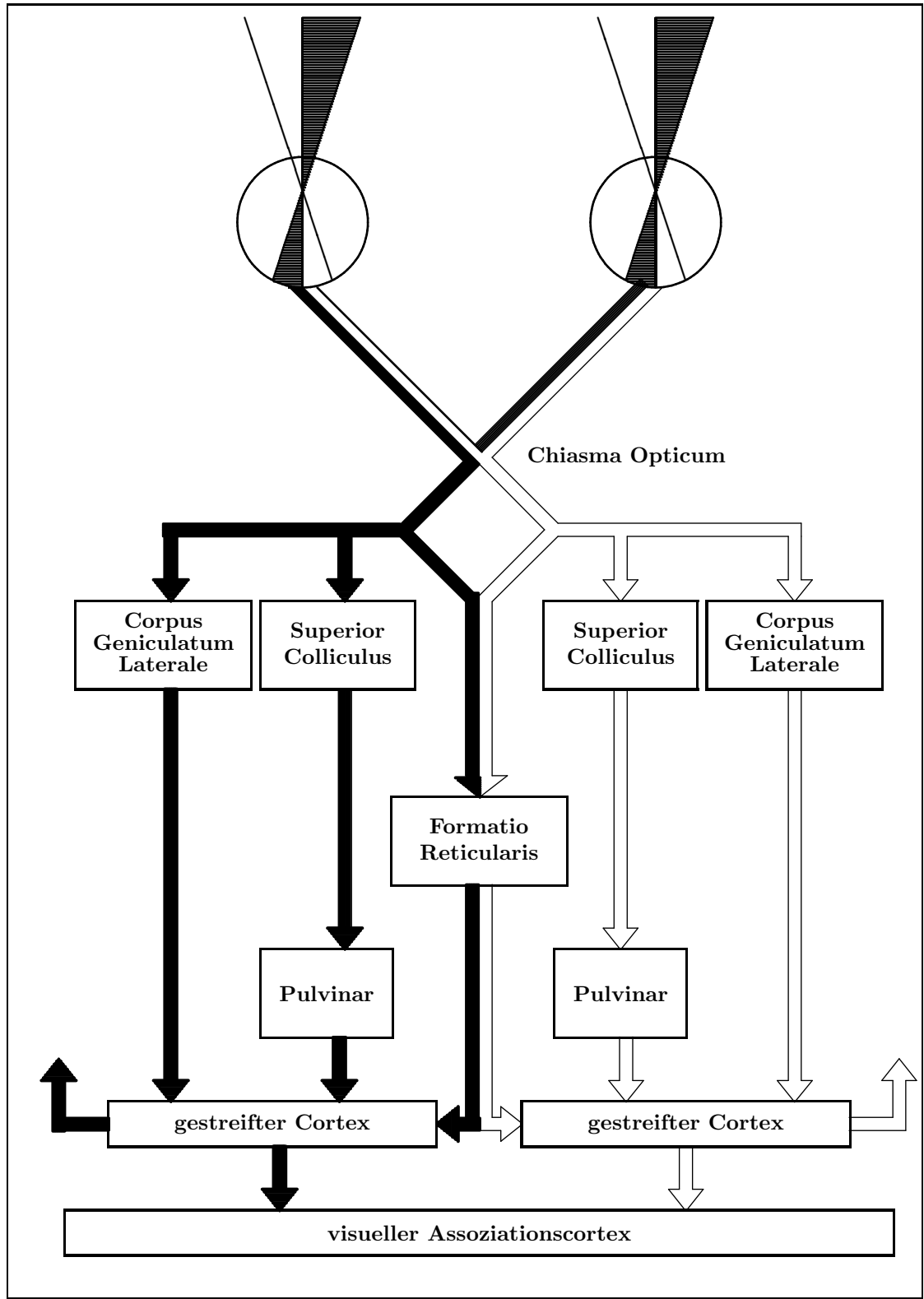


Abbildung 3: Schematische Darstellung des visuellen Systems

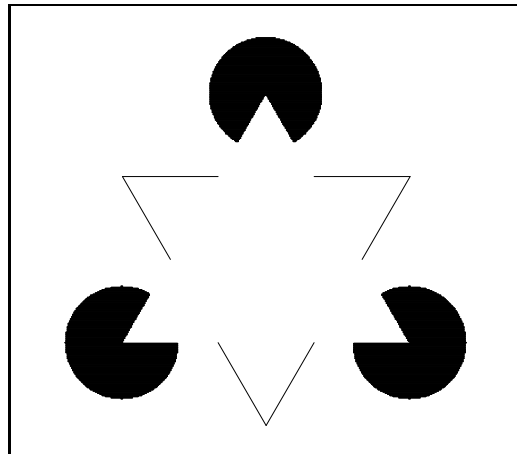


Abbildung 4: Das Kanizsa-Dreieck.

Das Kanizsa-Dreieck ist ein Beispiel einer „subjektiven Kontur“.
 Siehe: Gaetano Kanizsa: Subjective contours. *Scientific American*, 234(4):48–52, 1976.

es findet schon im Auge eine Datenreduktion von $120 : 1$ statt. Die durch das Auge wahrgenommene Information landet über verschiedene Zwischenstationen im primären visuellen Cortex („gestreifter“ Cortex, Feld 17). Dort wird die Information ausgebreitet und nach verschiedenen Gesichtspunkten sortiert. Die Information wird dann noch einmal komprimiert, gelangt schließlich ins Kurzzeitgedächtnis und nach einer weiteren Kompression ins Langzeitgedächtnis. Das Langzeitgedächtnis kann allenfalls wenige hundert Bit pro Sekunde aufnehmen (siehe Abbildung 5). Wenn man die in der Abbildung angegebenen Zahlen zugrundelegt (sie sind durchaus diskutierbar, dürften jedoch bis auf einige Größenordnungen korrekt sein), dann leistet das menschliche Wahrnehmungsvermögen auf dem Wege von den Rezeptoren zum Langzeitgedächtnis eine Reduktion von $1 : 10^8$. Das bedeutet, daß von einem Spielfilm von 90 Minuten Länge⁷ letztlich eine Datenmenge von einigen hundert Byte im Langzeitgedächtnis verbleibt. Mit anderen Worten: Die im Langzeitgedächtnis gespeicherte Information ließe sich in einem briefmarkengroßen Ausschnitt des allerersten Bildes der Sequenz unterbringen!

Unser visuelles System macht also aus der Not seiner Unvollkommenheit eine Tugend, indem es aus der immensen Flut von Informationen, die es von der Außenwelt erhält, einen winzigen Bruchteil eines Bruchteils extrahiert. Immerhin genügt dieser winzige Teil, um bei einer Seriensenkung im Fernsehen bei der nächsten Folge die Personen wiederzuerkennen und die Handlung (sofern eine solche vorhanden ist) zu rekonstruieren. Das visuelle System vollbringt diese unglaubliche Datenreduktion, indem es *semantische* Information zur Kompression nutzt.

Diese enorme Reduktion des Datenumfangs kann nur bedeuten, daß die Information im Langzeitgedächtnis in hochabstrakter Form gespeichert wird. Das bedeutet für uns Mathematiker, daß hier unsere Lieblingsbeschäftigung gefragt ist. Ein Ansatz zur Modellierung dieser abstrakten

⁷(PAL S-VHS Format: 720×576 Bildpunkte, 25 Bilder pro Sekunde) entspricht unkomprimiert eine Datenrate von etwa 166 Mbit/s. Für einen Spielfilm von 90 Minuten Länge kommt man auf einen Datenumfang von 112 Gbyte.

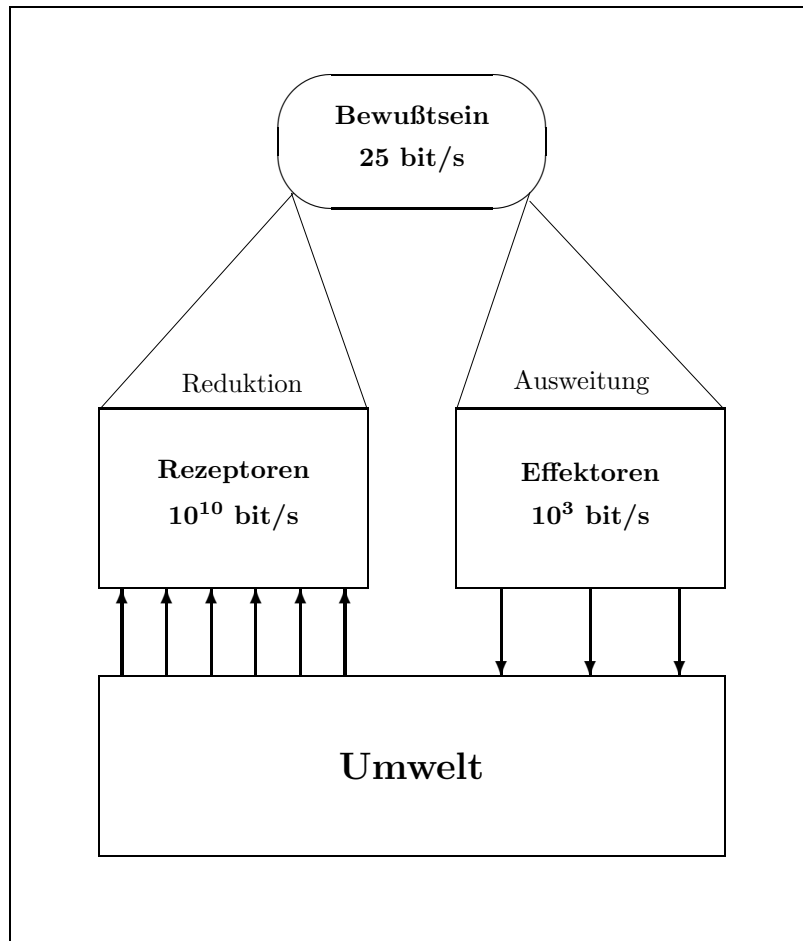


Abbildung 5: Schema zur Informationsverarbeitung im Menschen

Nach Georg Klaus und Heinz Liebscher, Wörterbuch der Kybernetik, 1979, Stichwort „Information“.

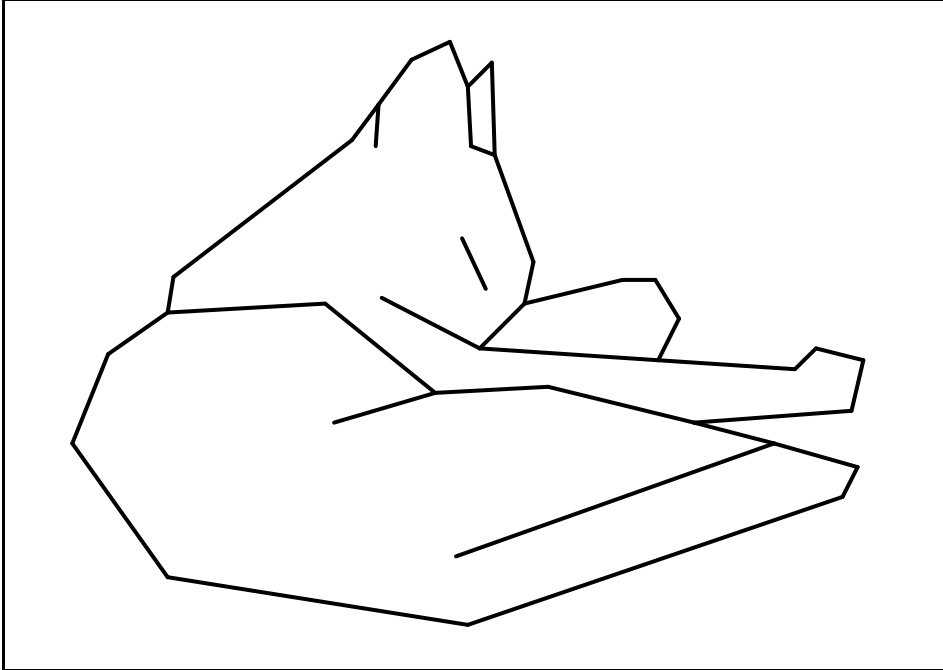


Abbildung 6: Reduktion eines Bildes auf „wesentliche“ Merkmale

Nach Fred Attneave: Informational aspects of visual perception. *Psychological Review*, 61:183–193, 1954. Im Bild wurden 38 markante Punkte ermittelt (Punkte maximaler Krümmung) und diese durch 40 Strecken miteinander verbunden.

Datenstrukturen im Langzeitgedächtnis wurde von dem Psychologen Fred Attneave vorgeschlagen⁸. Wie in Abbildung 6 dargestellt ist, genügt zur Erkennung der „schlafenden Katze“ eine Menge von 38 Punkten und 40 Relationen (Verbindungsstrecken). Diese Fähigkeit unseres visuellen Systems, aus wenigen Hinweisen ein Objekt in der Datenbank des Langzeitgedächtnisses zu assoziieren bildet die Existenzgrundlage zahlreicher Künstler, insbesondere der Karikaturisten!

Die Geschichte über die Problematik der menschlichen Wahrnehmung ist aber noch längst nicht zu Ende. Unser Langzeitgedächtnis ist kein Datenspeicher, es filtert die eingehende Information anhand von „Vorurteilen“. Wir erkennen die Welt nicht auf Grund eines Kantschen *a priori*, sondern durch den Zerrspiegel der dem Menschen durch die Evolution mitgegebenen Verhaltensnormen. Robert Havemann weist beispielsweise darauf hin, daß wir unsere Umwelt nicht metrisch wahrnehmen, sondern verzerrt. Ein Gegenstand, der sich in einer Entfernung von 10 m über uns befindet, erscheint „sehr hoch“, eine horizontale Entfernung von 10 m hingegen wird als „recht nah“ empfunden. Havemann spricht von der „Käseglockenstruktur“ des psychischen Raumes⁹ Nach Carnap¹⁰ nehmen wir unsere Umgebung überhaupt nicht „metrisch“, sondern

⁸Siehe auch Fred Attneave: Multistability in perception. *Scientific American*, 225 (6):62–71, December 1971.

⁹Robert Havemann: *Dialektik ohne Dogma? Naturwissenschaft und Weltanschauung; hat Philosophie den modernen Naturwissenschaften bei der Lösung ihrer Probleme geholfen? naturwissenschaftliche Aspekte, philosophische Probleme*, (Rororo; 683 : rororo aktuell), Rowohlt, Reinbek bei Hamburg, 1964, zweite Vorlesung.

¹⁰Rudolf Carnap: *Einführung in die Philosophie der Naturwissenschaften*, Ullstein Materialien, Ullstein Buch

„topologisch“ wahr:

Die »räumliche Anschauung«, wie sie genannt wird, ist nicht so sehr eine Anschauung einer metrischen Struktur als vielmehr eine Anschauung einer topologischen Struktur. Unsere Wahrnehmungen sagen uns, daß der Raum dreidimensional und kontinuierlich ist und daß jeder Punkt die gleichen topologischen Eigenschaften wie jeder andere Punkt hat. Aber in bezug auf die metrischen Eigenschaften des Raumes geben unsere Anschauungen und Intuitionen uns keine genauen und eindeutigen Antworten.

Verhaltenskonventionen, Aberglaube, Kultur, Vorurteile und das Erbe aus dem Holozän bestimmen unsere Sicht der Welt. Unsere Einschätzung der Welt ist keineswegs rational, sondern auf vielfältige Weise verzerrt und belastet. Man denke hier an das Verhalten von Menschen gegenüber Risiken am Beispiel von BSE oder Kernenergie im Vergleich zu etwa dem Straßenverkehr!

Damit noch nicht genug: Unser Langzeitgedächtnis ist kein passiver Speicher, der die in ihm eingeschriebene Information für alle Zeiten unverändert bewahrt, unser Gedächtnis ist aktiv. Ständig wird die in ihm enthaltene Information nach Art des Orwellschen „Wahrheitsministeriums“ umgeschrieben und redigiert. Dies wirkt sich eklatant aus beim Verhalten von Augenzeugen etwa bei Unfällen. Robert Buckhout¹¹ schreibt zu diesem Thema:

Human perception and memory function effectively by being selective and constructive. As Ulric Neisser of Cornell University has pointed out, “Neither perception nor memory is a copying process.” Perception and memory are decision-making processes affected by the totality of a person’s abilities, background, attitudes, motives and beliefs, by the environment and by the way his recollection is eventually tested. The observer is an active rather than a passive perceiver and recorder; he reaches conclusions on what he has seen by evaluating fragments of information and reconstructing them. He is motivated by a desire to be accurate as he imposes meaning on the overabundance of information that impinges on his senses, but also by a desire to live up to the expectations other people and to stay in their good graces. The eye, the ear and other sense organs are therefore social organs as well as physical ones.

Elizabeth F. Loftus¹² berichtet über Experimente, bei denen man bei Versuchspersonen „durch Suggestion und Einbildung . . . dem Gedächtnis Reminiszenzen von Ereignissen einpflanzen [lassen], die nicht so oder überhaupt nie stattgefunden haben.“

Wir können also als Resümé festhalten:

Wir können überhaupt nichts wahrnehmen!

Trotz dieser Negativaussage haben wir die Geheimnisse des Mikrokosmos und des Makrokosmos entschlüsseln können.

5 Modelle der Wahrnehmung

Es gibt eine große Anzahl von Versuchen, menschliche Wahrnehmungsprozesse auf Computern oder in Robotern nachzubilden. Diese Bestrebungen dienen praktischen Zielen, man möchte etwa

Nr. 35243, Verlag Ullstein GmbH, Frankfurt/Main, Berlin, 1986.

¹¹Robert Buckhout: Eyewitness testimony. *Scientific American*, 231(6):23–31, December 1974.

¹²Elizabeth F. Loftus: Falsche Erinnerungen. *Spektrum der Wissenschaft*, Januar 1998:63–67.

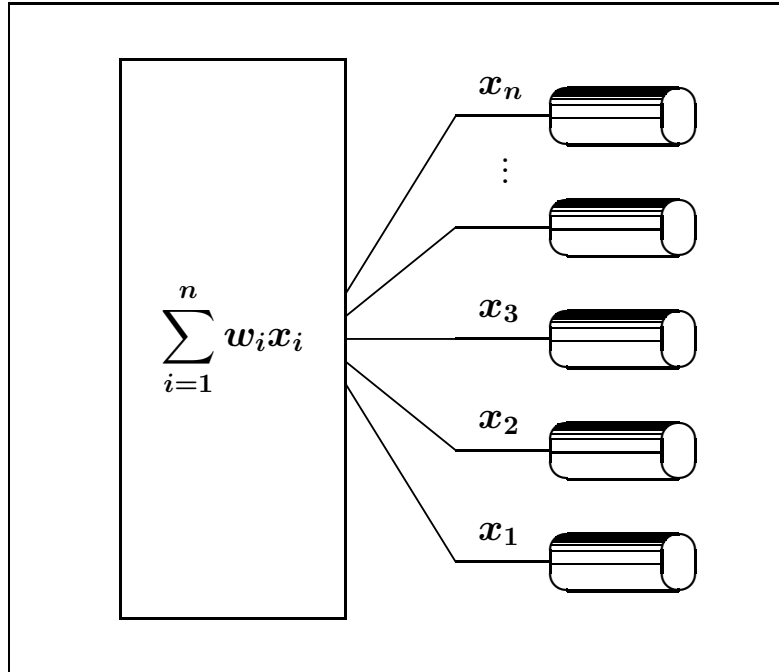


Abbildung 7: Perzeptron.

Die Sensoren nehmen jeweils einen Ausschnitt der Realität wahr, Die Leitungen haben begrenzte Kapazität. Aus den Signalen der Sensoren wird eine gewichtete Summe gebildet.

Systeme bauen, die Schriften lesen können, die medizinische Bilddaten interpretieren oder Roboter, die sich in unbekanntem Gelände selbständig zurechtfinden. Daneben hofft man, durch die Modellierung kognitiver Prozesse das Wesen der menschlichen Wahrnehmung zu verstehen.

5.1 Perzeptrons

Zu Beginn der sechziger Jahre wurden *Perzeptrons* genannte Geräte konstruiert und untersucht. Ein Perzeptron besteht aus einer Anzahl von Sensoren, die gewisse Ausschnitte der Umwelt wahrnehmen und die einen Hauptrechner über das Wahrgenommene informieren (Abbildung 7). Wir können uns die Funktionsweise einer solchen Maschine etwa wie folgt vorstellen: Jeder der Sensoren erhält zunächst eine Information über den zu erkennenden Gegenstand. Es wird etwa ein Programm heruntergeladen, welches die Sensoren befähigt, eine Aussage zu machen, ob in ihrem Gesichtsfeld Dinge zu sehen sind, die auf den gesuchten Gegenstand hinweisen.

Hierbei ist es sinnvoll, einige Annahmen über die Sensoren, die Datenleitungen und über den Hauptrechner zu machen, damit triviale Fälle ausgeschlossen werden. Wir wollen keine Annahmen über die Leistungsfähigkeit der beteiligten Komponenten machen, wir wollen uns nicht an die schwierige Frage der Charakterisierung von Rechenleistungen begeben. Jeder der Sensoren soll eine gewisse — nicht näher spezifizierte — Rechenleistung besitzen. Wir nehmen an, daß jeder

Sensor nur einen gewissen Ausschnitt der Umwelt wahrnehmen kann. Eine solche Annahme ist sinnvoll, würde nämlich einer der Sensoren die gesamte Umwelt wahrnehmen können, dann könnte er — hinreichende Rechenleistung vorausgesetzt — dem Hauptrechner einfach mitteilen, daß er den gesuchten Gegenstand gesehen hat. Der Hauptrechner, nachdem er festgestellt hat, daß der betreffende Sensor unfehlbar ist, einfach dessen Meinung übernehmen und die übrigen Sensoren ignorieren. Desgleichen dürfen die Datenleitungen nicht zu üppig ausgelegt werden. Dann nämlich könnte einfach jeder Sensor das von ihm gesehene Bild getreu übermitteln. Der Hauptrechner — ebenfalls hinreichende Rechenleistung vorausgesetzt — setzt dann die einzelnen Bilder zu einem Gesamtbild zusammen und analysiert dieses. Wir nehmen an, daß jeder Sensor etwa genau ein Bit oder auch eine reelle Zahl übertragen darf, wodurch er kundgibt, ob das von ihm wahrgenommene Teilbild einen Hinweis auf den gesuchten Gegenstand enthält oder nicht.

Der Hauptrechner hat dann die Aufgabe, aus den einlaufenden Signalen auf die Anwesenheit des gesuchten Gegenstandes zu schließen. Er könnte etwa eine Mehrheitsentscheidung fällen. Üblicherweise nimmt man an, daß der Hauptrechner die einlaufenden Signale mit Gewichtungsfaktoren multipliziert und die Produkte dann aufsummiert. Ist das Resultat größer als ein bestimmter Schwellwert, dann schließt der Hauptrechner darauf, daß der fragliche Gegenstand zu sehen ist, andernfalls nicht.

Eine besonders attraktive Eigenschaft des Perzeptrons ist, daß es *lernen* kann. Dazu nehmen wir an, es gebe eine große Anzahl von Testbildern, die man dem Perzeptron präsentieren kann, und für jedes dieser Testbilder sei bekannt, ob sich der gesuchte Gegenstand im Bilde befindet oder nicht. Man hat dann die mathematisch recht einfache Aufgabenstellung

Die Gewichte w_i sind so zu bestimmen, daß das Perzeptron alle Testbilder richtig klassifiziert.

Es wurde eine große Anzahl von Verfahren publiziert, mit deren Hilfe man diese Aufgabe praktisch lösen kann (oder aber feststellen, daß es keine Lösung gibt). Besonders beliebt waren inkrementelle Verfahren, bei denen dem Perzeptron die Testbilder einzeln und wiederholt präsentiert werden und dieses dann die Gewichte jeweils modifiziert, so daß die neu hinzugekommene Information richtig interpretiert wird. Man hat also eine Lernphase, in dem das Perzeptron seine Gewichte einstellt und eine Arbeitsphase, in dem es auf unbekannte Vorlagen angewandt wird. Stellt sich irgendwann einmal heraus, daß das Perzeptron ein Bild falsch klassifiziert, dann wird einfach „nachgelernt“.

Es ist nicht verwunderlich, daß in den frühen sechziger Jahren eine wahre Perzeptron-Euphorie ausbrach. Das Perzeptron hat eine entfernte Ähnlichkeit mit einem (winzigen) Kollektiv von Nervenzellen die über Axone und Dendriten miteinander kommunizieren. Auch bei den Nervenzellen spielen gewichtete Mittelwerte und Schwellwerte eine Rolle. Man hatte also den Eindruck, über das Studium von Perzeptrons Einsichten in das Wirken von kognitiven Vorgängen zu erhalten (daher auch der Name: lat. *perceptio* = Einsammeln, Erfassen, Begreifen, Erkenntnis). Attraktiv war auch das relativ einfache Lernmodell, das durch ein Perzeptron dargestellt wird. Zudem hatte man die Hoffnung, durch Perzeptrons komplexe und schwer durchschaubare Vorgänge modellieren zu können, ohne viel nachdenken zu müssen. Man hoffte — uneingestanden und auch nur partiell — vermittels von Perzeptrons auf eigene Lernvorgänge verzichten zu können und auch der Mühe enthoben zu sein, alles verstehen zu müssen!

Ein weiterer Punkt, der die Untersuchung von Perzeptrons beförderte, ist die Tatsache, daß die Perzeptronarchitektur eine einfache hierarchische Parallelrechnerarchitektur ist. Man konnte also hoffen, durch das Studium von Perzeptrons etwas über die Fähigkeiten von Parallelrechnern zu erfahren, und diese waren in den frühen sechziger Jahren noch nicht verfügbar.



Abbildung 8: Testbilder nach Minsky und Papert

Die Euphorie wurde nur beschränkt dadurch, daß in den sechziger Jahren die Rechner noch längst nicht so komfortabel waren wie heute, und daß die meisten der damals gebauten Perzeptrons Einzelanfertigungen waren, die in Labors „handgeschmiedet“ wurden. Hinzu kam ein ernüchterndes Forschungsergebnis, welches im Jahre 1969 publiziert wurde. Minsky und Papert [12] zeigten nämlich, daß Perzeptrons gewisse Prädikate nicht wahrnehmen können. Dies läßt sich im einfachsten Fall leicht einsehen (der allgemeine Fall, der in dem Buch von Minsky und Papert behandelt wird, bereitet dagegen sehr große Schwierigkeiten).

Als Testbilder wählten Minsky und Papert die vier in Abbildung 8 dargestellten Figuren. Wir sehen, daß zwei davon *zusammenhängend* sind, also „in einem Zuge“ nachzuzeichnen, und zwei der Figuren sind nicht zusammenhängend. Das Perzeptron soll herausfinden, welche der Figuren zusammenhängend sind und welche nicht. Wir nehmen an, das „Gesichtsfeld“ der einzelnen Sensoren sei derart, daß kein Sensor die gesamte Figur sehen kann, sondern nur höchstens ein Mittelstück — drei parallele Linien — oder ein rechtes oder linkes Endstück. Nur die letztgenannten Sensoren sind für die Entscheidung des Zusammenhanges relevant. Wir nehmen vereinfachend an, es gebe einen Sensor, der gerade das linke Ende der jeweiligen Figur sehen kann („linker Sensor“) und einen, der das rechte Ende sieht („rechter Sensor“). In Tabelle 1 sind die Gesichtsfelder



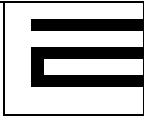
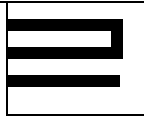




Testfigur	linker Sensor	rechter Sensor
S_1		
S_2		
S_3		
S_4		

Tabelle 1: Die vier Möglichkeiten für die relevanten Sensoren

dieser beiden Sensoren dargestellt.

Wir nehmen an, im Gesichtsfeld des linken Sensors befinde sich das folgende Bild:



In diesem Falle nimmt das Perzeptron entweder die nicht zusammenhängende Figur S_1 wahr oder die zusammenhängende Figur S_3 . Wie auch immer der linke Sensor auf Grund dieser Wahrnehmung votiert, er hat stets in einem Falle nicht recht. Analoges gilt für die anderen Sensoren. Man kann durch Untersuchung aller möglichen Fälle leicht zeigen, daß bei einfacher (also ungewichteter) Abstimmung durch das Perzeptron nicht entschieden werden kann, ob Zusammenhang vorliegt oder nicht.

Minsky und Papert zeigten noch viel mehr: Ein Perzeptron kann kein „topologisches“ Prädikat entscheiden (mit der Ausnahme „trivialer“ Prädikate). Diese Aussage ist wahrhaft niederschmetternd! Topologische Prädikate spielen beispielsweise bei der Dokumentanalyse eine sehr wichtige Rolle. Wenn wir eine wichtige Information auf einem Dokument zur Hervorhebung oder Abgrenzung in einen Kasten einschließen, dann wollen wir damit andeuten, daß das, was im Inneren des Kastens ist, einen besonderen semantischen Wert haben soll. Innen und Außen sind aber gerade topologische Begriffe. Wir werden uns mit diesem Dilemma noch beschäftigen.

An dieser Stelle höre ich bei ähnlichen Vorträgen häufig einen Einwand, typischerweise von einem Informatiker vorgetragen: Schließlich seien Minsky und Papert in diesem Punkte längst wider-

legt, man müsse nur eine weitere Schicht von Rechnern einführen, dann gäbe es diese Art von Problemen nicht mehr. Im Dezember 1984 hatte sich A. K. Dewdney in der Rubrik „Computer–Kurzweil“ im *Spektrum der Wissenschaft* mit Perzeptrons beschäftigt¹³ Im Februar 1985 fügte er in einem Nachtrag zu den Perzeptrons hinzu¹⁴, daß John M. Evans aus Hartford in Connecticut gerade eben dies bewiesen habe. Ich möchte diese Geschichte zur „Perceptron–Folklore“ zählen. Die Gründe für meine Skepsis sind:

- Bislang konnte mir niemand Details nennen, Recherchen schlugen fehl, anscheinend hat noch niemand eine Arbeit von John M. Evans zu diesem Thema gesehen.
- Im Jahre 1988 schrieben Minsky und Papert eine „Expanded Edition“ ihres Werkes. Wäre schon im Jahre 1985 bekannt gewesen, daß der „Hauptsatz der Theorie der Perzeptrons“ nur für einschichtige Perzeptrons gilt, dann hätten sie dies gewiß erwähnt.
- Sobald die Information die Sensoren der ersten Schicht verlassen hat, liegt nur ein Bitmuster von Stimmen für oder gegen Zusammenhang vor. In unserem Perzeptron–Modell kann der Hauptrechner — ganz unabhängig von seiner Leistungsfähigkeit — nicht mehr entscheiden, ob Zusammenhang vorliegt oder nicht. Gäbe es eine Architektur, die dies leisten kann, dann brauchte der Hauptrechner lediglich diese Architektur zu simulieren. Man muß also annehmen, daß nicht nur Voten von der Sensorschicht weitergegeben werden. Damit wären aber die eingangs genannten „Spielregeln“ verletzt.
- Die leistungsfähigste Rechnerarchitektur, die wir kennen, unser Gehirn, welches hochparallel arbeitet, hat Schwierigkeiten, den Weg durch einen gezeichneten Irrgarten zu finden, es muß gezwungen werden, sequentiell zu verfahren, etwa indem man die Wege des Irrgartens mit einem Bleistift verfolgt (siehe Abbildung 9).

Unabhängig davon gab es in den frühen sechziger Jahren eine Anzahl von seriösen Anwendungen von Perzeptrons. Das bekannteste Beispiel eines Perzeptrons war der OCR–Leser (OCR = **O**ptical **C**haracter **R**ecognition) der American Bankers Association¹⁵, der von der Firma IBM gebaut wurde. Bei diesem System werden die Ziffern auf Schecks normiert dargestellt wie etwa in Abbildung 10. Jede Ziffer wird in 7×10 Bildpunkte aufgelöst (beziehungsweise in 7×8 wesentliche Bildpunkte) und jeder der Sensoren des Perzeptrons ist einem Bildpunkt zugeordnet.

In Abbildung 11 sind die Bildpunktmuster für die Ziffern 1, 3 und 8 dargestellt sowie eine Gewichtsmatrix, die die Ziffer 1 von allen anderen unterscheiden soll. Man addiert einfach für jeden Bildpunkt die entsprechende Zahl in der Maske und erhält dann für die Ziffer 1 einen hohen positiven Wert, für die anderen Ziffern jedoch hohe negative Werte. Wenn man einmal annimmt, daß ein oder mehrere der Bildpunkte falsch besetzt sind, dann wird das Perzeptron trotzdem nicht notwendig falsch klassifizieren, es reagiert also „robust“ auf Störungen der Eingabewerte.

5.2 Neuronale Netzwerke

In den letzten Jahrzehnten nahm die Leistungsfähigkeit der Rechner rasant zu, und es war möglich, Perzeptrons und noch kompliziertere Architekturen auf einem gewöhnlichen PC zu

¹³A. K. Dewdney: Computer–Kurzweil: Die Schwächen des elektronischen Auges oder: Warum Computer sehr wohl sehen, in den meisten Fällen aber nichts erkennen können. *Spektrum der Wissenschaft*, Dezember 1984:4–10.

¹⁴A. K. Dewdney: Computer–Kurzweil: Die kybernetische Saga vom Planeten Wa–tor oder: Wie man die Populationsdynamik eines Ökosystems auf dem Computer simuliert. *Spektrum der Wissenschaft*, Februar 1985:6–13.

¹⁵J. S. Griffin, Jr., J. H. King, Jr. and C. J. Tunis: A pattern identification system using linear decision functions. *IBM Systems J.*, 2:248–267, 1963.

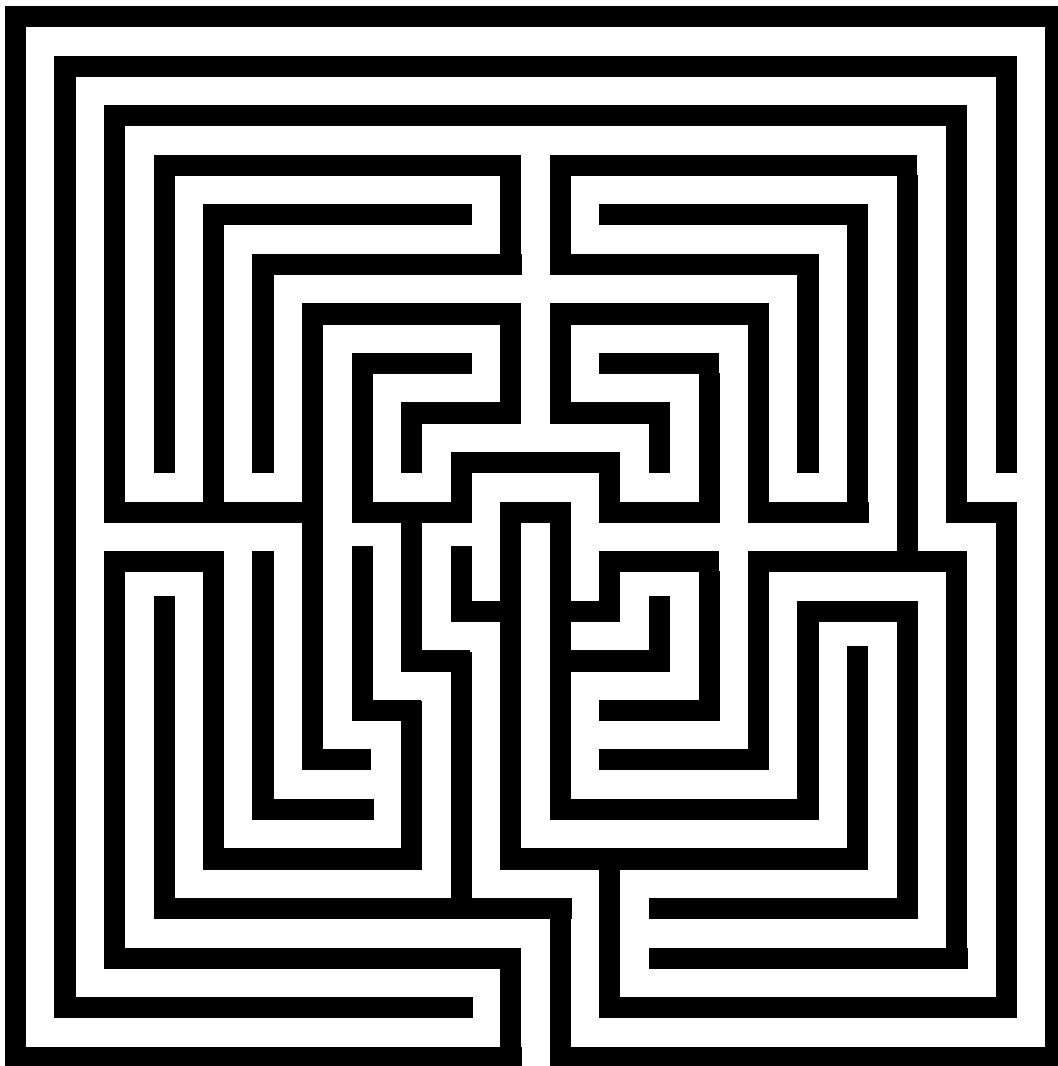


Abbildung 9: Das Heckenlabyrinth von Chevening

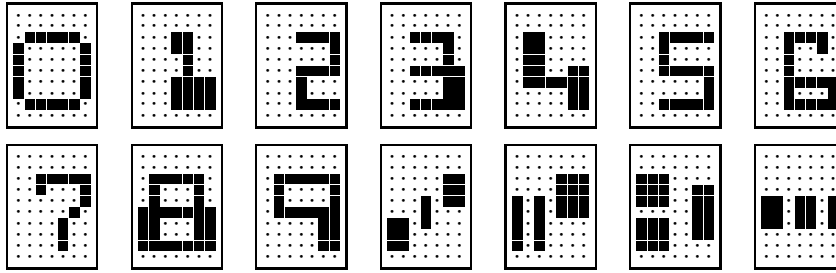


Abbildung 10: Die „American Bankers Association E-13 B Fonts“.

simulieren. Dies führte zu einer Wiederbelebung der Perzeptronidee — allerdings mit weitaus nüchteren Erwartungen als in den sechziger Jahren. Besondere Bedeutung erlangten dann die *neuronalen Netzwerke*, die eine Weiterentwicklung der Perzeptrons darstellen.

Ein neuronales Netzwerk ist eine hierarchische Anordnung von miteinander vernetzten Rechnern. Aufgabe dieser Anordnung ist es, eine komplizierte Funktion, die man nur aus Beobachtungen kennt, durch Einstellung der inneren Parameter der Rechner (der sogenannten *Gewichte*) darzustellen.

Zur Bestimmung der Gewichte kann man ähnliche Lernalgorithmen verwenden, wie bei den Perzeptrons. Für geeignete Netzstrukturen kann man die Gewichte recht gut durch den sogenannten „Backpropagation-Algorithmus“ ermitteln. Dann stellt das Netz dann einen Rechner dar, der nicht programmiert werden muß, der vielmehr in der Trainingsphase „lernt“. Durch Hinzunahme zusätzlicher Daten kann man leicht die Approximation verbessern. Eine Vergrößerung des Netzes durch Hinzunahme weiterer Knoten und Kanten ist leicht möglich¹⁶. Es ist kaum möglich, einen Überblick über die Anwendungsmöglichkeiten neuronaler Netzwerke zu geben. Die Literatur ist sehr umfangreich, und die Hersteller liefern phantastische „Verkaufsaussagen“. Vor einigen Jahren brachte Siemens einen Netzwerkchip auf den Markt. In der Werbung wurde als Anwendung die Voraussage von Börsenkursen genannt. Ich bezweifle, daß sehr viele Käufer mit diesem Netzwerkchip reich geworden sind.

Ein neuronales Netz dient dazu, Funktionen zu approximieren. Dazu sei $F : \mathbb{R}^d \rightarrow \mathbb{R}$ eine unbekannte Funktion, beispielsweise der morgige Börsenkurs einer bestimmten Aktie als Funktion etwa der Börsenkurse ausgewählter Aktien in den vergangenen Tagen. Das Netz soll durch Einstellen seines „internen Zustandes“, das heißt, der Gewichte, diese Funktion approximieren. In die Eingangsknoten werden die Werte der Komponenten des Argumentvektors $x \in \mathbb{R}^d$ eingegeben. Der Ausgangsknoten soll dann den Funktionswert $F(x)$ ausgeben.

Es kann für ein neuronales Netzwerk grundsätzlich nicht möglich sein, eine vorgelegte Funktion zu approximieren.

¹⁶Es gibt eine große Anzahl von Publikationen über dieses Gebiet. Hier seien nur einige „populärwissenschaftliche“ Titel genannt: Geoffrey E. Hinton: Wie neuronale Netze aus Erfahrung lernen. *Spektrum der Wissenschaft*, November 1992:134–143. Richard P. Lippmann: An introduction to computing with neural nets. *IEEE ASSP Magazine*, April 1987:4–22. David E. Rumelhart, Bernard Widrow and Michael A. Lehr: The basic ideas in neural networks. *Communications of the ACM*, 37:87–92, 1994. David Tank und John J. Hopfield: Kollektives Rechnen mit Neuronenähnlichen Schaltkreisen. *Spektrum der Wissenschaft*, Februar 1988:46–54.

Für gehobeneren mathematischen Ansprüchen sei auf den Übersichtsartikel von Ferber verwiesen (Reginald Ferber: Neuronale Netze. In S. D. Chatterji, Benno Fuchssteiner, Ulrich Kulisch, Roman Liedl und Walter Purkert, Hrsg.: *Jahrbuch Überblicke Mathematik 1992*, Seiten 137–157, Friedr. Vieweg & Sohn Verlagsgesellschaft mbH, Braunschweig/Wiesbaden, 1992.)

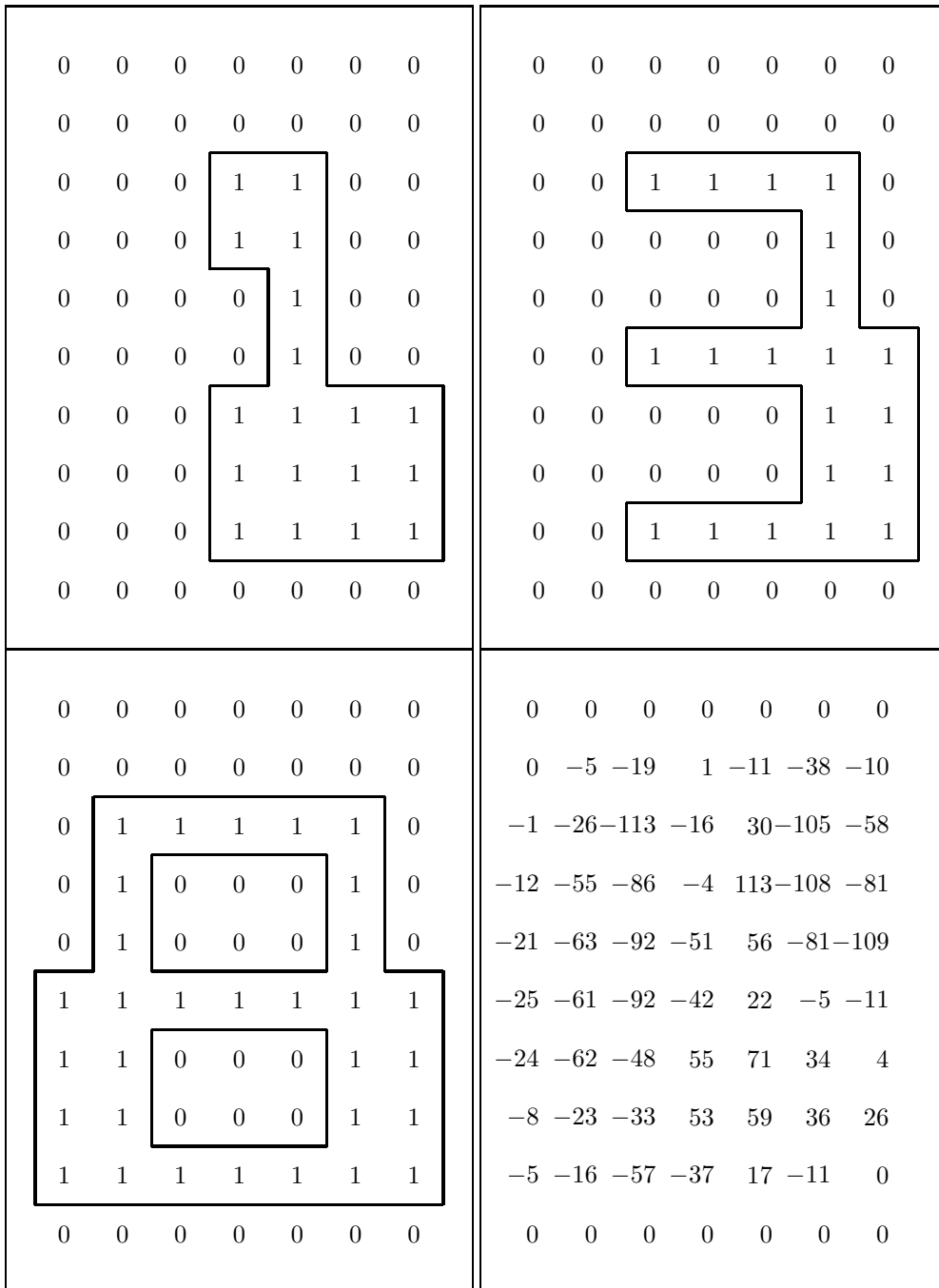


Abbildung 11: Die Ziffern 1, 3 und 8 der American Bankers Association E-13 B Fonts. in der rechten unteren Abbildung sind die Gewichtungsfaktoren für die Erkennung der Ziffer 1 angegeben. Skalare Multiplikation mit den dargestellten Codierungen der Ziffern ergeben für die Ziffer 1 den Wert +508, für die Ziffer 3 den Wert -509 und für die Ziffer 8 den Wert -877.

Modellierungsfehler Es kann sein, daß die darzustellende Funktion überhaupt nicht von einigen der Eingabevariablen abhängt. Das wäre nicht so schlimm, man hat dann einfach „redundante“ Variable. Schwierig wird es, wenn es Variable gibt, die nicht durch Eingabeknoten dargestellt werden, die jedoch Einfluß auf die darzustellende Funktion haben. Ein solcher Fehler kann nur durch eine Analyse des untersuchten „realen“ Systems entdeckt und möglicherweise behoben) werden. Dies ist nicht unmittelbar Gegenstand mathematischer Betätigung.

Approximationsfehler Es kann sein, daß bei ungeschickter Wahl der Knotenfunktionen (und der Netzwerk-„Topologie“) gewisse Funktionen nicht dargestellt werden können (sogenannte „Nemesis-Funktionen“). Dies ist eine durchaus mathematische Fragestellung, die man so formulieren kann: Gegeben sei eine Klasse von Funktionen. Liegen die durch das Netzwerk repräsentierbaren Funktionen in dieser vorgegebenen Klasse dicht? Das heißt: Welche Funktionen sind approximierbar, wenn man voraussetzt, daß man die Anzahl der Knoten des Netzes beliebig vergrößern kann (wobei man sich auf eine gewisse Graphenstruktur gestlegt).

Fehlerordnungen Eine Frage, auf die wir hier nicht eingehen können, ist die nach der Approximationsgüte: Wieviele Knoten benötigt man, um eine Funktion mit gewissen Eigenschaften mit einer gewissen Genauigkeit darzustellen.

Bevor wir auf diese Fragen eingehen, konkretisieren wir die Aufgabenstellung. Zunächst einmal sei $F : \mathbb{R}^d \rightarrow \mathbb{R}$ die durch das Netz darzustellende Funktion. Die durch das Netz darstellbaren Funktionen sollen die Form $\Phi : \mathbb{R}^d \times \mathbb{R}^W \rightarrow \mathbb{R}$ haben, wobei W die Anzahl der Gewichte sein soll. Hat man nun Beobachtungen $x^{(1)}, x^{(2)}, \dots, x^{(N)} \in \mathbb{R}^d$, dann sind die Gewichte so zu wählen, daß (etwa) gilt

$$\sum_{i=1}^N \left[F(x^{(i)}) - \Phi(x^{(i)}, w) \right]^2 \rightarrow \text{Minimum.}$$

Dies ist eine Minimierungsaufgabe zur Bestimmung von $w \in \mathbb{R}^W$.

In Abbildung 12 ist ein solches Netz schematisch dargestellt. Die erste Schicht links ist die Eingabeschicht. In der nächsten Schicht werden die Eingabewerte mit Gewichten w_{ikn} aufsummiert und dann auf das Resultat die Funktion e^{-z} angewandt. In der nächsten Schicht wird die einfache Summe (mit Gewichten 1) gebildet und dann die Funktion $1/(1+t)$ angewandt. Die resultierende Funktion $1/(1+e^{-z})$ ist eine sogenannte *Sigmoidfunktion*. Für $z \rightarrow -\infty$ konvergiert sie gegen 0 und für $z \rightarrow \infty$ gegen 1. Sigmoidfunktionen kann man als geglättete Varianten der beim Perzeptron auftretenden Signumfunktion ansehen. Von dieser Eigenschaft haben sie ihren Namen erhalten.

In der vierten Schicht wird die mit w_i gewichtete Summe der Ergebnisse der dritten Schicht gebildet. Insgesamt werden durch das Netz also alle Funktionen der folgenden Bauart dargestellt.

$$\Phi(x, w) = \sum_{i=1}^{\ell} w_i \left[1 + \sum_{k=1}^K w_{ik} \exp \left(- \sum_{n=1}^d w_{ikn} x_n \right) \right]^{-1}. \quad (1)$$

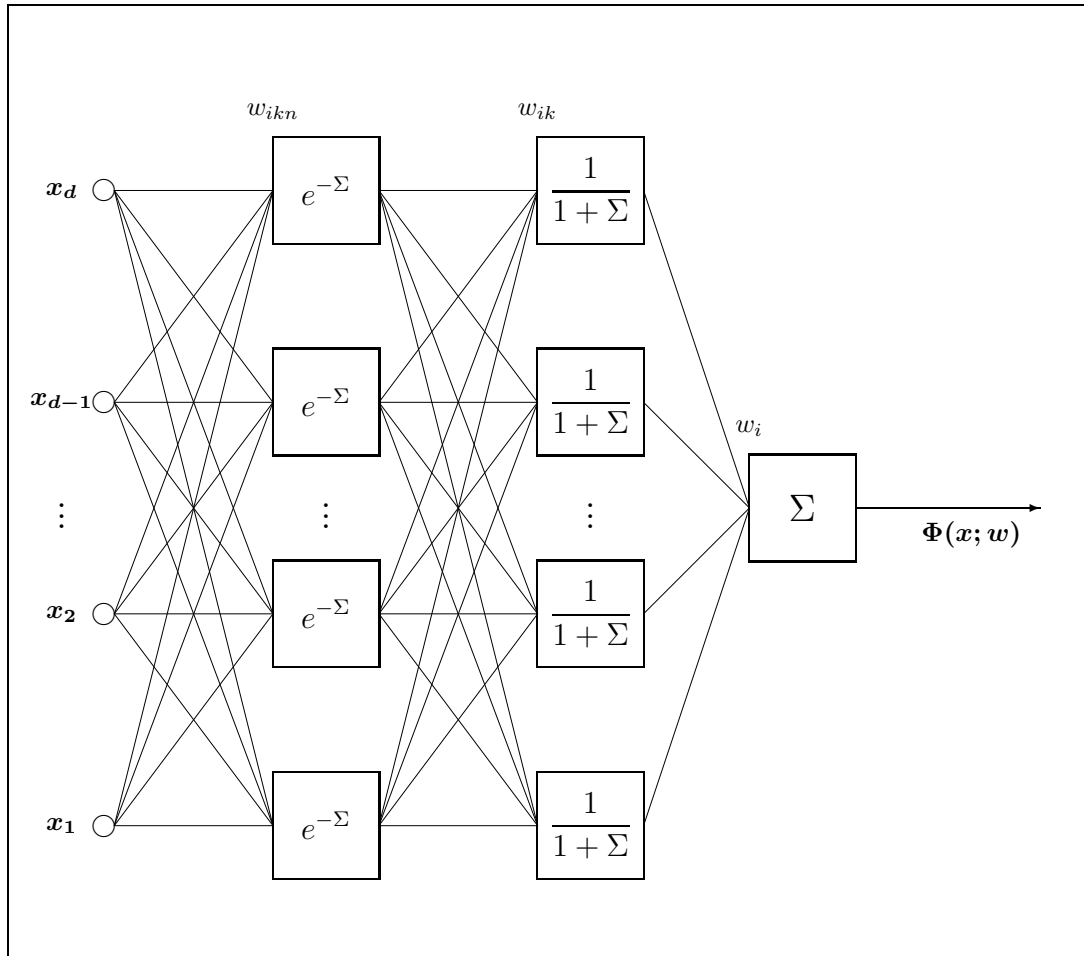


Abbildung 12: Beispiel eines neuronalen Netzes mit zwei „verborgenen“ Schichten.

5.2.1 Der Satz von Kolmogorow

Im Jahre 1900 hatte David Hilbert auf dem zweiten Internationalen Mathematikerkongreß in Paris seine berühmten „Hilbertschen Probleme“ vorgestellt¹⁷. Das dreizehnte dieser Probleme behandelte die Frage, ob sich eine Funktion mehrerer Veränderlicher als Funktion einer Veränderlichen darstellen lasse. Ganz speziell untersuchte Hilbert die Gleichung $x^7 + ax^3 + bx^2 + cx + 1 = 0$. Die Nullstellen dieser Gleichung sind Funktionen der Koeffizienten a, b, c . Er stellte die Frage, ob sich die Nullstellen wenigstens als als endliche Summen von Funktionen zweier Variabler darstellen lassen. Der Hintergrund dazu war, wie Hilbert angibt, daß sich Funktionen von zwei Variablen gut graphisch darstellen lassen (etwa durch Höhenlinienbilder oder Nomogramme). Hilbert hatte diese Frage verneint.

Für die Gründe, aus denen Hilbert ausgerechnet diese spezielle Polynomgleichung wählte, sei auf den Beitrag von A. G. Vituškin in dem erwähnten Buch verwiesen. Man findet dort gleichfalls eine Darstellung der Gedankengänge, die Hilbert zu seiner negativen Aussage bewogen hatten.

Jedoch fanden im Jahre 1957 A. N. Kolmogorov¹⁸ und V. I. Arnol'd¹⁹ eine überraschend einfache bejahende Antwort auf die Fragestellung von Hilbert (die Antwort ist einfach zu formulieren aber keineswegs einfach zu beweisen!): Jede stetige Funktion mehrerer Veränderlicher besitzt nämlich eine Darstellung als Superposition von stetigen Funktionen einer Veränderlichen. Das heißt, jedes $f : [0, 1]^d \rightarrow \mathbb{R}$ besitzt eine Darstellung der Form

$$f(\xi_1, \xi_2, \dots, \xi_d) = \sum_{i=1}^{2d+1} g_i \left(\sum_{j=1}^d h_{ij}(\xi_j) \right),$$

wobei die g_i und die h_{ij} stetige Funktionen einer Variablen sind. Die h_{ij} können dabei unabhängig von der Funktion f gewählt werden, die Funktionen g_i hängen dagegen von der Funktion f ab. Das Problem hierbei ist, daß die zur Darstellung benutzten Funktionen zwar stetig, aber doch höchst irregulär sind, wie David A. Sprecher im Jahre 1965 zeigen konnte²⁰.

Diese Darstellung als Superposition läßt natürlich an neuronale Netzwerke denken. Tatsächlich hat Robert Hecht-Nielsen im Jahre 1987 auf diesen Zusammenhang aufmerksam gemacht²¹. An die Arbeit von Hecht-Nielsen schloß sich eine heftige Debatte in der Fachliteratur über neuronale Netze an, die Frage betreffend, welchen Rang diese Beobachtung für die Praxis habe. Der „Altmeister“ der Wahrnehmungstheorie, Tomaso Poggio, nannte den Satz von Kolmogorow schlicht „irrelevant“²².

¹⁷Es gibt eine ausgezeichnete kommentierte Darstellung der Hilbertschen Probleme: *Die Hilbertschen Probleme. Vortrag „Mathematische Probleme“ von D. Hilbert, gehalten auf dem 2. Internationalen Mathematikerkongreß, Paris 1900, erläutert von einem Autorenkollektiv unter der Redaktion von P. S. Alexandrov, (Ostwalds Klassiker der exakten Wissenschaften, Band 252)*, Akademische Verlagsgesellschaft Geest & Portig K.-G., Leipzig, 1971.

¹⁸A. N. Kolmogorow: Über die Darstellung stetiger Funktionen mehrerer Veränderlicher als Superposition stetiger Funktionen einer Veränderlichen und deren Addition. (Russisch). *Doklady Akademii Nauk SSSR*, 114(5):953–956, 1957.

¹⁹V. I. Arnol'd: Über Funktionen dreier Veränderlicher. (Russisch). *Doklady Akademii Nauk SSSR*, 114(4):679–681, 1957.

²⁰David A. Sprecher: On the structure of continuous functions of several variables. *Trans. Amer. Math. Soc.*, 115:340–355, 1965.

²¹Robert Hecht-Nielsen: Kolmogorov’s mapping neural network existence theorem. In *Proceedings of the International Conference on Neural Networks, (III, 11–14)*, IEEE Press, New York, 1987. Siehe auch Věra Kůrková: 13th Hilbert’s problem and neural networks. In *Theoretical Aspects of Neurocomputing*, pages 213–216, World Scientific, Singapore, 1991.

²²Federico Girosi and Tomaso Poggio: Representation properties of networks: Kolmogorov’s theorem is irrelevant. *Neural Computation*, 1:465–469, 1989, siehe jedoch Věra Kůrková: Kolmogorov’s theorem is relevant. *Neural Computation*, 3:617–622, 1991.

Literatur

- [1] A. G. Arkadjew und E. M. Brawerman: *Zeichenerkennung und maschinelles Lernen, (Beiheft Elektronische Rechenanlagen, Bd. 13)* R. Oldenbourg Verlag, München und Wien, 1966.
- [2] George Berkeley: *Essay Towards a New Theory of Vision*, Printed by Aaron Rhames, at the Back of Dick's Coffee House, for Jeremy Pepyat, Bookseller in Skinner-Row, MDCCIX.
The Theory of Vision, or Visual Language Shewing The Immediate Presence and Providence of a Deity, Vindicated and Explained by the Author of Alciphron, or, The Minute Philosopher, Printed for J. Tonson in the Strand, London, MDCCXXXIII.
- [3] Karol Borsuk: *Theory of Shape, (Monografie Matematyczne, Tom 59)*, PWN Polish Scientific Publishers, Warszawa 1975.
- [4] Richard G. Casey and George Nagy: Advances in pattern recognition. *Scientific American*, 224(4):56–71, 1971.
- [5] Ian L. Dryden and Kanti V. Mardia: *Statistical Shape Analysis, (Wiley Series in Probability and Statistics)*, John Wiley & Sons, Chichester · New York · Weinheim · Brisbane · Singapore · Toronto, 1998.
- [6] Ulrich Eckhardt and Gerd Maderlechner: Invariant thinning. *International Journal of Pattern Recognition and Artificial Intelligence, (Special Issue on Techniques for Thinning Digitized Patterns)*, 7:1115–1144, 1993.
- [7] Josef Hofer-Alfeis and Gerd Maderlechner: Automated conversion of mechanical engineering drawings to CAD models: Too many problems? *IAPR Workshop on Computer Vision — Special Hardware and Applications —, October 12–14, 1988, Tokyo, Proceedings*, pages 206–209.
- [8] Axel Korn: *Bildverarbeitung durch das visuelle System, (Fachberichte Messen — Steuern — Regeln, Band 8)*, Springer-Verlag, Berlin, Heidelberg, New York, 1982.
- [9] Longin Jan Latecki: *Discrete Representation of Spatial Objects in Computer Vision, (Computational Imaging and Vision)*, Kluwer Academic Publishers, Dordrecht / Boston / London, 1998.
- [10] Martin D. Levine: *Vision in Man and Machine*, McGraw-Hill Book Company, New York, St. Louis, San Francisco, Auckland, Bogota, Hamburg, Johannesburg, London, Madrid, Mexico, Montreal, New Delhi, Panama, Paris, Sao Paulo, Singapore, Sydney, Tokyo, Toronto, 1985.
- [11] E. A. Lord and C. B. Wilson: *The Mathematical Description of Shape and Form*, Ellis Horwood Limited, Chichester; John Wiley & Sons, New York, Chichester, Brisbane, Toronto, 1984.
- [12] Marvin Minsky and Seymour Papert: *Perceptrons. An Introduction to Computational Geometry*, The MIT Press, Cambridge, Massachusetts, London, England, 1969.
Expanded Edition, The MIT Press, Cambridge, Massachusetts, London, England, 1988.
- [13] N. J. Nilsson: *Learning Machines*, McGraw-Hill Book Company, New York, St. Louis, San Francisco, Toronto, London, Sydney, 1965.
- [14] Karl R. Popper und John C. Eccles: *Das Ich und sein Gehirn*, 2. Auflage, R. Piper & Co. Verlag, München, Zürich, 1982.

- [15] Frank Rosenblatt: *Principles of Neurodynamics*, Spartan Books, Washington, D. C., 1962.
- [16] Jean Serra: *Image Analysis and Mathematical Morphology*, English version rev. by Noel Cressie. Academic Press, Inc. (Harcourt Brace Jovanovich, Publishers), London, Orlando, San Diego, New York, Austin, Boston, Sydney, Tokyo, Toronto, 1982.
- [17] D'Arcy Wentworth Thompson: *Über Wachstum und Form*, In gekürzter Fassung neu herausgegeben von John Tyler Bonner. Ins Deutsche übersetzt von Ella M. Fountain und Magdalena Neff. Mit einem Geleitwort von Adolf Portmann. (*suhrkamp taschenbuch wissenschaft 410*), Suhrkamp Taschenbuch Verlag, Frankfurt a. M., 1983.
- [18] Alan M. Turing: Computing machinery and intelligence. *MIND*, LIX:433–460, 1950.
In D. C. Ince, editor: *Collected Works of A. M. Turing. Volume 1: Mechanical Intelligence*, pages 133–460, North-Holland, Amsterdam, London, New York, Tokyo, 1992.
- [19] P. T. Saunders, editor: *Collected Works of A. M. Turing. Volume 3: Morphogenesis*, North-Holland, Amsterdam, London, New York, Tokyo, 1992.