

## IV. Gestalt und Künstliche Intelligenz (1943 bis zur Gegenwart)

Die Tradition des Gestaltbegriffs, von Goethe bis zur Gestaltpsychologie, lässt sich der breiteren philosophischen Grundausrichtung des Holismus zuordnen, dessen Widerpart ist der Atomismus. Diese Differenz wird im 20. und beginnenden 21. Jahrhundert dort problematisch, wo holistische Gegenstandsfelder in atomistisch operierende technische Systeme umgesetzt oder durch sie simuliert werden sollen. Nirgends zeigt sich der Zusammenstoß beider Paradigmen besser als in der fortdauernden Auseinandersetzung um die Begriffe und Vorgehensweisen zweier der jüngsten Disziplinen des letzten Jahrhunderts, der Informatik und der Kognitions- bzw. Neurowissenschaft. Dieses Spannungsverhältnis wird vor allem am Forschungsdiskurs um die Künstliche Intelligenz (KI) seit den 1940er Jahren deutlich, der hier nachgezeichnet werden soll.

### IV.1

Ist von ›Atomismus‹ die Rede, wird damit die Überzeugung bezeichnet, dass ein Phänomen oder ein Prozess, einschließlich seiner spezifischen Eigenschaften, durch die Analyse seiner konstitutiven Elemente erklärt werden könne und dass eine solche Erklärung abschließend sei; ›Holismus‹ meint die gegenteilige Auffassung. Auch der Begriff der ›Gestalt‹ ist in diesem Sinne holistisch und richtet sich explizit gegen zwei atomistische Grundthesen, die Max Wertheimer 1922 beschrieb: erstens gegen die »Mosaik- oder Bündelthese«, der zufolge Komplexe aus einfachen Elementen aufgebaut sind, und zweitens gegen die »Assoziationsthese«, für die Empfindungen lediglich miteinander korreliert werden, ohne auf Kausalitäten oder Gesetzmäßigkeiten zu achten.<sup>136</sup>

Ist der Holismus vor allem in den Geisteswissenschaften beheimatet – in der Philosophie, zumal der hermeneutischen Phänomenologie, aber auch der verstehenden Soziologie und der Geschichtswissenschaft –, kennen doch auch die Naturwissenschaften holistische Erklärungsmodelle: in der Biologie etwa als ›Organismus‹ und

134 Vgl. ebd., S. 406.

135 Vgl. Steve Heims: *Constructing a Social Science for Postwar America*, Cambridge, Mass. 1991, S. 201-223, 224-247.

136 Wertheimer: »Untersuchungen zur Lehre von der Gestalt. I« (Anm. 91), S. 47-58, hier S. 48 f.

›Ökologie‹, in der Physik als ›Feld‹, disziplinenübergreifend als ›Emergenz‹.<sup>137</sup> Die Informatik dagegen hat eine eindeutig atomistische Tendenz. Als automatisierte »Verarbeitung symbolisch repräsentierter Information«<sup>138</sup> sind ihre Grundelemente Algorithmen und Datenstrukturen.<sup>139</sup> Damit ist sie doppelt atomistisch, indem sie ihren *Gegenstandsbereich* in diskret zu verarbeitender, digitaler Information sieht und in ihrer *Methodik* auf explizite, in Algorithmen wohldefinierte Regelschritte abzielt. Dieser Atomismus hatte von den *computer sciences* ausgehend auch Einfluss auf die Neuro- und Kognitionswissenschaft. Denn indem die Informatik sich fast unmittelbar seit ihrer fachlichen Etablierung mit der Möglichkeit Künstlicher Intelligenz befasste, nahm sie immer auch an einem psychologischen Paralleldiskurs teil. Die disziplinäre Geschichte der KI war so durchweg von zwei spiegelbildlichen Erkenntnisinteressen durchdrungen: Einerseits ist KI *konstruktiv*, als »the art of creating machines that perform functions that require intelligence when performed by people«;<sup>140</sup> andererseits aber hat sie *heuristische* Funktion als »the study of mental faculties through the use of computational models«.<sup>141</sup>

Die Schnittmenge dieser beiden Ansätze ist die Idee, »that thinking and computing are radically the same«.<sup>142</sup> Weil KI und Kognitionswissenschaft dabei zunächst disziplinär nicht unterschieden waren, bestand von Anfang an ein Einfallstor für die Gestaltpsychologie. Sie nutzte es in der Nachkriegszeit zunächst, konnte sich aber nicht gegen die dominierende atomistische Tendenz durchsetzen; erst ab den 1980er Jahren wurden ihre Ideen wieder aufgegriffen, doch dann meist ohne expliziten Bezug auf diese Denkschule.

#### IV.2

Der Grundstein für die Kompatibilität von Informatik und Neurowissenschaft wurde von Warren McCullochs und Walter Pitts' Artikel *A Logical Calculus of the Ideas Immanent in Nervous Activity* von 1943 gelegt.<sup>143</sup> Nachdem Santiago Ramón y Cajal

137 Vgl. Georg Toepfer: »Ganzheit«, in: ders.: *Historisches Wörterbuch der Biologie*, Stuttgart 2011, S. 693-728; Michael Esfeld: »Holismus und Atomismus in den Geistes- und Naturwissenschaften. Eine Skizze«, in: Alexander Bergs/Soelve I. Curdts (Hg.): *Holismus und Individualismus in den Wissenschaften*, Frankfurt a.M. 2003, 7-21 sowie die Beiträge von Georg Toepfer und Alexandra Heimes in diesem Band, S. ###-### und ###-###.

138 Gesellschaft für Informatik (Hg.): *Was ist Informatik?*, Bonn 2015, S. 2.

139 Vgl. Nikolaus Wirth: *Algorithmen und Datenstrukturen*, Tübingen 1975.

140 Ray Kurzweil: *The Age of Intelligent Machines*, Cambridge, Mass. 1990, S. 14.

141 Eugene Charniak/Drew V. McDermott: *Introduction to Artificial Intelligence*, Reading, Mass. 1985, S. 6.

142 John Haugeland: *Artificial Intelligence. The Very Idea*, Cambridge, Mass. 1985, S. 2. Diese Idee geht zudem der Kybernetik als Disziplin digitaler Automaten voraus, wie etwa in Kenneth Craik: *The Nature of Explanation*, Cambridge 1943, das noch *analoge* Rechenmaschinen als Vorbild für Hirnprozesse nimmt.

143 Vgl. Warren S. McCulloch/Walter Pitts: »A Logical Calculus of the Ideas Immanent in Nervous Activity«, in: *Bulletin of Mathematical Biophysics* 5 (1943), S. 115-133; zur folgenden Darstellung vgl. Steve Joshua Heims: *The Cybernetics Group 1946-1953. Constructing a Social*

um die Jahrhundertwende die Gewebestruktur des Hirns als Verbund einzelner Zellen – über Synapsen verbundene und polarisierbare Neuronen – hatte nachweisen können, versuchten McCulloch und Pitts deren Funktionsweise als neuronale Implementierung von Aussagenlogik qua Boole'scher Algebra zu konzipieren. Unter der (heute widerlegten) Annahme, dass Neuronen nur binäre Zustände kennen, idealisierten sie diese als logische Gatter,<sup>144</sup> die im Verbund als »neuronales Netz« jede endliche logische Aussage umzusetzen in der Lage seien. Emergente Eigenschaften nahmen sie dabei gerade nicht an, sondern die atomistische und lineare Beschreibbarkeit von Elementarfunktionen.<sup>145</sup> Das legte nicht nur nahe, dass Hirnzustände logisch formalisiert werden können, sondern auch, dass neuronale Netze Turingvollständig sind, mit ihnen also jedes mögliche Programm ausgeführt werden kann: Das Hirn funktioniert wie ein digitaler Computer und Hirnfunktionen können in digitalen Computern nachgebaut werden – jedenfalls theoretisch, denn die praktische Bestätigung dieses Modells blieben McCulloch und Pitts schuldig.

Diese Interpretation war fruchtbar vor allem im Kontext der seit 1942 stattfindenden Macy-Konferenzen, die den Forschungsansatz der Kybernetik etablierten. Neben der Identifikation von Feedbackprozessen in Maschinen und Organismen gehörte eine funktionale (aber nicht histologische) Ähnlichkeit zwischen Hirn und Computer zu ihren Grundannahmen, wodurch sie mit dem seit den 1920er Jahren in den USA vorherrschenden Neobehaviorismus in Konflikt stand. Statt Geist lediglich als Blackboxartiges Aktions-Reaktions-System zu beschreiben, wollte sie auch den zugrunde liegenden Hirnmechanismen auf den Grund gehen, wofür McCulloch/Pitts ein erstes Modell boten.<sup>146</sup>

Auch die Vertreter der Gestaltpsychologie – bereits früh mit der Situation in den USA vertraut und dort rezipiert<sup>147</sup> – wandten sich gegen den Behaviorismus. Dennoch fanden, als die Protagonisten der Gestaltpsychologie nach der nationalsozialistischen Machtübernahme in die USA emigriert waren, die beiden Seiten nicht zueinander: Der »organic reductionism«<sup>148</sup> der Kybernetik mit seinen atomistischen Prämissen lag zu sehr auf der Linie jener von Wertheimer bemängelten Bündelthese. Stattdessen zog es vor allem Köhler vor, Hirnprozesse nicht als Effekt von diskreten Neuronenimpulsen, sondern als physikalische Feldprozesse zu beschreiben, in denen alle Elemente einander wechselseitig beeinflussen. Auf diese Weise seien Gestalteffekte der Wahrneh-

*Science for Postwar America*, Cambridge, Mass. 1991; David Bates: »Creating Insight. Gestalt Theory and the Early Computer«, in: Jessica Riskin (Hg.): *Genesis Redux. Essays in the History and Philosophy of Artificial Life*, Chicago 2007, S. 237-260.

144 Logische Gatter sind Schaltungen, die zwei oder mehr Eingangssignale nach einem logischen Operator (z. B. »und«, »oder«, »nicht«) in ein Ausgangssignal umwandeln; logische Aussagen sind endlich, wenn sie in einer endlichen Zahl von Schritten gelöst werden können.

145 Vgl. Bates: »Creating Insight« (Anm. 142), S. 241.

146 Vgl. Heims: *The Cybernetics Group* (Anm. 142), Kap. 9-10.

147 Vgl. Michael M. Sokal: »The Gestalt Psychologists in Behaviorist America«, in: *The American Historical Review* 89,5 (1984), S. 1240-1263. So erschienen zahlreiche Werke der Gestaltpsychologie bereits vor der Emigration ihrer Hauptvertreter auf Englisch.

148 Heims: *The Cybernetics Group* (Anm. 142), S. 42.

mung auch als Gestaltprozesse im Hirnsubstrat zu begründen. Eine experimentelle Bestätigung dieser These zu liefern, gelang ihm aber nicht.<sup>149</sup>

Diese Theorie war auch der Kybernetik-Gruppe bekannt. So traf das McCulloch-Pitts-Modell bereits beim ersten Macy-Treffen auf den Einwand einiger Mitglieder, dass gerade Gestaltwahrnehmung (als Form- und Mustererkennung) durch ein atomistisch-logisches Modell nicht erklärt werden könne.<sup>150</sup> McCulloch und Pitts reagierten auf diese Herausforderung 1947 mit dem ambitioniert betitelten Artikel *How We Know Universals*: Das Gehirn könne Akkorde unabhängig von der Tonhöhe und Formen unabhängig von ihrer Größe und Lage als Invarianten erkennen – dezidiert ist von »Gestalten« die Rede –, indem es die arithmetischen Mittel von als Mannigfaltigkeiten codierten Eindrücken berechne.<sup>151</sup> Holistische Eindrücke seien funktional über atomistische Grundlagen, als Listen von Eigenschaften zu erklären – nicht aber, wie bei Köhler, als Gestalteigenschaften des Hirnprozesses selbst, sondern über den Umweg von Neuronenaktivität, die Eindrücke binär codieren und sich informationstheoretisch beschreiben ließen.

Zur direkten Konfrontation des atomistischen und des holistischen Ansatzes kam es 1948 beim Hixon-Symposium über »Cerebral Processes in Behavior«, auf dem sich Köhler und McCulloch unversöhnlich gegenüberstanden: Waren für McCulloch die messbaren elektrischen Felder des Hirns bloße Schmiereffekte diskreter neuronaler Aktivität, hielt umgekehrt Köhler die Quanta der Neuronenimpulse nur kumulativ, eben als Felder für wirksam.<sup>152</sup> Dennoch vermochte Köhler, anders als McCulloch, keine stringent mathematisierbare und vor allem ingenieurstechnisch operationalisierbare Theorie zu formulieren: Maschinen konnte man mit seinem Modell gerade nicht bauen. Anders im Falle des McCulloch-Pitts-Modells: John von Neumann stellte, ebenfalls auf dem Hixon-Symposium, seine Automatentheorie vor, die explizit auf Neuronen als binäre Relais zurückgriff, mit denen sich alle formallogischen Sätze darstellen ließen. Er hatte keinen Zweifel daran, dass man *jedes* Verhalten »completely and unambiguously« in solchen Sätzen ausdrücken und damit auch in Computern nachbauen könne: »This description may be lengthy, but it is always possible. To deny it would amount to adhering to a form of logical mysticism which is surely far from most of us.«<sup>153</sup>

149 Köhler: *Die physischen Gestalten* (Anm. 100), insb. S. 70-79 zur Beschreibung der Feldtheorie und S. 180 f. zur Ablehnung atomistischer Hirntheorien.

150 Vgl. Heims: *The Cybernetics Group* (Anm. 142), S. 224.

151 Walter Pitts/Warren S. McCulloch: »How We Know Universals. The Perception of Auditory and Visual Forms«, in: *Bulletin of Mathematical Biophysics* 9 (1947), S. 127-147, hier S. 137.

152 Vgl. Warren S. McCulloch: »Why the Mind is in the Head«, in: Lloyd A. Jeffress (Hg.): *Cerebral Mechanisms in Behavior. The Hixon Symposium*, New York 1951, S. 42-82, hier S. 53 f.; Wolfgang Köhler: »Relational Determination in Perception«, in: ebd., S. 200-243, hier S. 208, 212, 227.

153 John von Neumann: »The General and Logical Theory of Automata«, in: Jeffress (Hg.): *Cerebral Mechanisms in Behavior* (Anm. 151), S. 1-41, hier S. 23.

Dass Gestaltprozesse atomistisch aufzulösen sind, war auch der Tenor von Norbert Wiensers enorm populärem Buch *Cybernetics* (1948).<sup>154</sup> Dem tat keinen Abbruch, dass sich im Laufe der Macy-Konferenzen herausstellte, dass die vorgeschlagene Lösung, Gestalten als Liste von Eigenschaften zu codieren, sich technisch nur unbefriedigend umsetzen ließ. Gegen Wiensers Version der Kybernetik erhob Köhler in einer Rezension Einspruch, wobei er, außer gegen die atomistische These, auch grundsätzlicher gegen die Idee schoss, Maschinen und Menschen seien auf dieselbe Weise zu beschreiben: »In the relation of human beings to the computing machines, thinking in the proper sense of the term appears to remain the task of the former.«<sup>155</sup> Köhler brachte nicht lediglich eine weltanschaulich motivierte Mechanismusskepsis zum Ausdruck,<sup>156</sup> sondern formulierte grundsätzlicher die Frage, ob Maschinen je menschenartig intelligent sein können. Dabei zog er den gestaltpsychologischen Begriff der »Einsicht« (*insight*) heran, den er bei seiner Forschung an Menschenaffen entwickelt hatte.<sup>157</sup> Gemeint war das plötzlich sich einstellende, gestalthafte Verstehen von Struktur- und Funktionszusammenhängen. Sein Schluss, der in der Folge für die Kritik an KI zur Trope wurde, lautete: »The machines do not know, because among their functions there is none that can be compared with *insight* into the meaning of a problem.«<sup>158</sup>

#### IV.3

Standen vor allem Probleme der *Wahrnehmung* am Beginn der KI-Forschung, verquickten sie sich bald mit der Frage nach der Simulierbarkeit von *Intelligenz*.<sup>159</sup> Als Disziplin mit eigener Identität begann die KI-Forschung 1956 auf dem »Dartmouth Summer Research Project on Artificial Intelligence«, das die Mathematiker John McCarthy und Marvin Minsky ausrichteten. Dort fiel zum ersten Mal der Begriff »Künstliche Intelligenz«, kamen seine Hauptakteure zusammen und wurden die Parameter bestimmt, unter denen das Forschungsfeld in der Folgezeit bestellt wurde.<sup>160</sup>

- 154 Vgl. Norbert Wiener: *Cybernetics, Or Control and Communication in the Animal and the Machine*, Cambridge, Mass. 1948.
- 155 Wolfgang Köhler: Rezension zu Norbert Wiener: *Cybernetics or Control and Communication in the Animal and the Machine*, in: *Social Research* 18.1 (1951), S. 125-130.
- 156 Das meint etwa Steve Joshua Heims, der Köhlers Holismus für das Resultat eines an Dilthey geschulten Antimechanismus hält; vgl. Heims: *The Cybernetics Group* (Anm. 142). Dass diese Identifikation von Gestalt und dem »Hunger nach Ganzheit« (Peter Gay) durchaus bestand, bestätigt Mitchell G. Ash in: *Gestalt Psychology* (Anm. 12), S. 297.
- 157 Vgl. Köhler: *Intelligenzprüfung an Menschenaffen* (Anm. 89), S. 137.
- 158 Köhler: Rezension zu Norbert Wiener (Anm. 154), S. 128, Hvh. H. B.
- 159 Taucht das Problem maschineller Denk- und Lernfähigkeit bereits auf bei Craik: *Nature of Explanation* (Anm. 141), S. 111, stellte Alan Turing 1950 explizit die Frage nach intelligenten Maschinen als lernenden; vgl. Alan M. Turing: »Computing Machinery and Intelligence«, in: *Mind* 59.236 (1950), S. 433-460.
- 160 Vgl. John McCarthy/Marvin L. Minsky/Nathaniel Rochester u. a.: »A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence«, in: *AI Magazine* 27.4 (2006), S. 12-14; zur Geschichte vgl. Nils J. Nilsson: *The Quest for Artificial Intelligence. A History of Ideas and Achievements*, Cambridge 2010, S. 52-56.

Das Atomismus-Holismus-Problem tauchte an dieser Stelle in neuer Form auf. War die Köhler'sche »Einsicht« eine wesentliche Komponente intelligenten Verhaltens bei Menschen, bestand die Herausforderung der KI-Forschung nun darin, diese Art von holistischer Intelligenz in Maschinen zu modellieren. Die Frage verschob sich von der Heuristik zur Konstruktion, bestand also nicht mehr nur darin, ob Hirn- als Feldprozesse oder als Feuern einzelner Neuronen zu artikulieren sind, sondern darin, wie Intelligenz zu operationalisieren ist – als explizite Regeln oder als implizites Wissen (in Gilbert Ryles einflussreicher Unterscheidung: als *knowing-that* oder als *knowing-how*).<sup>161</sup>

Die bis heute fortwirkende Unterscheidung zweier Typen von KI – »symbolisch« und »subsymbolisch« – hat ihren Ursprung in dieser Frage und wurde auf dem Workshop zum ersten Mal vorgestellt.<sup>162</sup> Symbolische Ansätze nähern sich künstlicher Intelligenz unter Annahme des atomistischen Paradigmas: Denken wird hier in erster Linie als die logische Manipulation von Symbolen verstanden, die atomistische Fakten repräsentieren. Die subsymbolische Familie setzt dagegen auf das Hirnmodell. Denken ist hier keine explizite, digitale, sequenzielle Symbolmanipulation (auch wenn deren Funktion post hoc aussagenlogisch ausgedrückt werden kann), sondern eine implizite, konnektionistisch organisierte und letztlich statistisch operierende Berechnung.

In Dartmouth wurden beide Ansätze diskutiert – Marvin Minsky, einer der Initiatoren, hatte etwa seine Dissertation über neuronale Netze geschrieben –, aber letztlich gab man symbolischen Modellen den Vorzug. Sie wurden oft in sogenannten »Expertensystemen« implementiert, unter denen der »General Problem Solver« von Allen Newell und Herbert Simon, beide Teilnehmer in Dartmouth, einer der frühesten und einflussreichsten war. Der GPS verschränkt eine Wissensdatenbank mit einer Schlussregeln enthaltenden Inferenzmaschine, so dass er aus der Kombination von Fakten Folgerungen ableiten konnte. Newell und Simon verbanden, ihrem Selbstverständnis nach, Einsichten aus Behaviorismus und Gestaltpsychologie gleichermaßen: So sollten subjektive Erfahrungsinhalte keine Rolle spielen, die von einer rein behavioristischen Perspektive ausgeschlossenen Funktionsbegriffe wie »Aufgabe« aber wieder in die Beschreibung menschlicher Intelligenz aufgenommen werden.<sup>163</sup> Menschliches Problemlösen betrachteten sie dabei, ganz auf der Linie von Neumanns, als Operationen, die sich aussagenlogisch ausdrücken lassen; wie Simon in einem Grundsatzartikel 1973 erläutert, sind Probleme, zu deren Lösung keine eindeutigen Verfahrensschritte angegeben werden können, in kleinere, wohldefinierte Teilprobleme zu

161 Gilbert Ryle: »Knowing How and Knowing That«, in: *Proceedings of the Aristotelian Society* 46.1 (1946), S. 1-16.

162 Für eine hilfreiche, allgemeinverständliche Einführung siehe Melanie Mitchell: *Artificial Intelligence. A Guide for Thinking Humans*, New York 2019, S. 17-34.

163 Vgl. Allen Newell/Herbert A. Simon: »GPS, a program that simulates human thought«, in: Heinz Billing (Hg.): *Lernende Automaten*, München 1961, S. 109-124. Auch hier findet sich wieder die Verschränkung von Informatik und Psychologie: »If the fit of such a program were close enough to the overt behavior of our human subject [...] then it would constitute a good theory of the subject's problem solving.« (Ebd., S. 114)

zerlegen, bis sie schließlich in logischer Form darstellbar und lösbar sind.<sup>164</sup> Im Gegensatz zum *Atomismus der Wahrnehmung* könnte man hier von einem *Atomismus der Rationalität* sprechen.

Im Vergleich zwischen dem Vorgehen eines Menschen und dem des GPS bei der Lösung logischer Probleme stellten Newell und Simon allerdings auch fest, dass die menschlichen Probanden nicht, wie im GPS, streng *sequenziell* voringen, sondern Denkschritte durchaus gleichzeitig verliefen.<sup>165</sup> Die *parallele* Verarbeitung ist aber die Stärke gerade des subsymbolischen Paradigmas. Als dessen einflussreiche technische Implementierung entwarf Frank Rosenblatt 1958 das »Perceptron«, das erste funktionale künstliche neuronale Netz.<sup>166</sup> Es ging über das McCulloch-Pitts-Modell hinaus, indem es nicht nur rechnen, sondern lernen konnte und emergente Qualitäten an den Tag legte. Dazu setzte Rosenblatt das 1949 von Donald Hebb entdeckte Gesetz um, nach dem Lernprozesse zwischen Neuronen als Verstärkung ihrer synaptischen Verbindungen ablaufen (»Hebb'sche Lernregel«).<sup>167</sup>

Das Perceptron bestand aus drei Schichten künstlicher neuronaler Zellen. Über eine Schicht von Fotozellen konnte es Muster aufnehmen und über eine Assoziationschicht lernen; genügend trainiert, zeigte eine Ausgabeschicht an, ob ein Muster erkannt worden war – unabhängig von Lage und Größe. Obwohl immer noch auf Wahrnehmung ausgerichtet, bewies Rosenblatts Erfindung, dass diese Operationen auch ohne explizite Regeln möglich waren. Wichtiger noch war das klarer integrierte Lernverfahren: Heißt »lernen« bei Expertensystemen die Erweiterung der Wissensdatenbank, ist es bei Perzeptronen ein Effekt des wiederholten Aufnehmens ähnlicher Beispiele, wobei sich die Verbindungen zwischen den einzelnen Schichten verstärken. Folgt das Expertensystem linearen Wenn-dann-Strukturen, hat die Schaltung des Perzeptrons einen parallelen Aufbau und kommt ohne die Trennung von Fakten und Regeln aus. Ihrer Struktur nach steht das Paradigma der symbolischen KI einer atomistischen, das der subsymbolischen KI einer holistischen, einer Gestaltlogik nahe.

Dennoch verloren die nichtsymbolischen Modelle, zumal Rosenblatts Perceptron, gegenüber dem symbolischen Ansatz bald an Akzeptanz. Marvin Minsky und Seymour Papert, zwei der wichtigsten Symbolisten, veröffentlichten 1969 eine Kritik des Perzeptrons, die es bis in die 1980er Jahre hinein ins Abseits drängte.<sup>168</sup> Obwohl Rosen-

164 »Given enough information about an individual, a program could be written that would describe the symbolic behavior of that individual. Each individual would be described by a different program, and those aspects of human problem solving that are not idiosyncratic would emerge as the common structure and content of the programs of many individuals.« (Herbert A. Simon: »The Structure of Ill Structured Problems«, in: *Artificial Intelligence* 4.3-4 (1973), S. 181-201, hier S. 190)

165 Vgl. Newell/Simon: »GPS« (Anm. 161), S. 122.

166 Frank Rosenblatt: »The Perceptron. A Probabilistic Model for Information Storage and Organization in the Brain«, in: *Psychological Review* 65.6 (1958), S. 386-408.

167 Vgl. Donald O. Hebb: *The Organization of Behavior. A Neuropsychological Theory*, New York 1949, Kap. 4.

168 Vgl. Marvin Minsky/Seymour Papert: *Perceptrons. An Introduction to Computational Geometry*, Cambridge, Mass. 1969.

blatt sich dezidiert gegen den Isomorphismus von Köhlers Feldtheorie wandte, formulierten Minsky und Papert ihre Einwände als Verteidigung einer atomistischen gegen eine ›unwissenschaftliche‹ holistische Erkenntnistheorie. Wie ein Echo auf von Neumanns Furcht vor einem logischen Mystizismus liest sich ihre erklärte Intention »to dispel what we feared to be the first shadows of a ›holistic‹ or ›Gestalt‹ misconception that would threaten to haunt the fields of engineering and artificial intelligence as it had earlier haunted biology and psychology.«<sup>169</sup> Mit äußerster Klarheit konnten sie nachweisen, dass Perzeptrone nicht alle logischen Operationen nachbilden konnte und es bei größeren Datenmengen schnell an seine Grenzen kam. Gemeinhin wird ihr Buch dafür verantwortlich gemacht, dass die Forschung an neuronalen Netzen in den 1970er Jahren zum Erliegen kam; mit Rosenblatts Tod 1971 war zudem der prominenteste Vertreter dieses Ansatzes verstummt.

#### IV.4

Doch auch die Symbolisten waren nicht ohne Gegner. Einer der ersten Philosophen, der eine holistisch fundierte Kritik des symbolistischen Ansatzes formulierte, war Hubert Dreyfus, der es für prinzipiell unmöglich ansah, auf diesem Weg menschenähnliche Intelligenz zu simulieren.<sup>170</sup> Der Symbolismus, so zeigte er, beruhte auf einer Reihe von Annahmen: der *biologischen* Identifikation des Gehirns mit einem digitalen Computer (wie bei Wiener), der *ontologischen* Annahme, dass die Welt aus isolierbaren Fakten bestehe (wie bei McCulloch/Pitts) und der *erkenntnistheoretischen* Annahme, dass der Verstand diese Fakten verarbeite (wie bei Newell/Simon).<sup>171</sup> Für einen solchen Atomismus kann Denken durch explizite Regeln formalisiert werden – *knowing-how* lässt sich als *knowing-that* ausdrücken.

Wie Köhler beharrte Dreyfus darauf, dass Menschen nicht nach dem simplistischen Maschinenmodell der Kybernetik zu erklären sind. Sie sind »embodied beings« und nicht bloß informationsverarbeitende Wesen; ein Gehirn ohne einen Körper ist zu keiner Intelligenz fähig.<sup>172</sup> Darüber hinaus ist Wissen, das im GPS in Form einer klar adressierbaren Datenbank abgelegt ist, oft gerade nicht explizit. Menschlicher intelligenter Umgang mit der Welt basiert auf implizitem Handlungs- und Hintergrundwissen – »Einsicht«. Um einem Computer Intelligenz zuzusprechen zu können, bräuchte dieser, so Dreyfus, dieses implizite Wissen, das nur durch die tatsächliche Begegnung mit der Welt in einem »being-in-a-situation« gewonnen werden könne.<sup>173</sup> Daher gelte: »being-in-a-situation turns out to be unprogrammable in principle using presently conceivable techniques.«<sup>174</sup>

169 Ebd., S. 19 f.

170 Vgl. Hubert L. Dreyfus: *What Computers Can't Do. A Critique of Artificial Reason*, New York 1972.

171 Vgl. ebd., S. 67-142.

172 Ebd., S. 149.

173 Ebd., S. 200.

174 Ebd., S. 215.

Dieser Einwand war überzeugend, solange die atomistische Annahme die Grundlage der »presently conceivable techniques« der KI-Forschung bildete. Rosenblatts Perzeptron, als Sackgasse gebrandmarkt, kam in Dreyfus' Buch nicht vor. Dennoch wurde die Arbeit am subsymbolischen System nicht völlig aufgegeben. Gerade in den Neurowissenschaften spielten mathematische Modellierungen von Hirnprozessen weiterhin eine große Rolle. Wegweisend waren John Hopfields 1982 publizierte Überlegungen zu den emergenten Eigenschaften neuronaler Netze,<sup>175</sup> vor allem aber die Veröffentlichung des zweibändigen *Parallel Distributed Processing* von 1986. Das PDP-Team um David Rumelhart und James McClelland konnte nachweisen, dass die Kritik von Minsky und Papert nicht für Modelle mit mehreren Schichten galt. Damit begann der bis heute anhaltende zweite Frühling des subsymbolischen Paradigmas.<sup>176</sup> Rumelhart und McClelland betonten explizit ihre Ablehnung atomistischer Annahmen und argumentierten für ein Verständnis von Kognition als *schwach emergentem*<sup>177</sup> Phänomen aus der Interaktion zwischen Elementen, in der Regelmäßigkeit kein Resultat von logisch formulierten Regeln sei.<sup>178</sup> In Anlehnung an das Motto der Gestalttheorie resümierten sie: »The whole is different than the *sum* of its parts.«<sup>179</sup>

Ist dies der späte Sieg der Gestaltpsychologie über die Kybernetik? Dreyfus wenigstens schwächte angesichts neuronaler Netze 1988 seine Kritik an der KI-Forschung ab und räumte ein, dass sie seiner ganzheitlichen Vorstellung von Wahrnehmung näherkämen. In *What Computers Can't Do* hatte er eine Version des Gestaltproblems formuliert, die er »perspicuous grouping« nannte: die Fähigkeit, Objektreihen anhand von zwischen ihnen herrschenden Ähnlichkeiten zu bilden, also ihre *kollektive Gestalt* zu erfassen.<sup>180</sup> 16 Jahre später gab er zu, dass »perspicuous grouping« im Wesentlichen gelöst sei. Doch trotz dieses Durchbruchs auf dem Feld der *Wahrnehmung* blieb er nach wie vor skeptisch, ob neuronale Netze tatsächlich als Baustein künstlicher maschineller *Intelligenz* taugen, und schloss: »Neural network modeling may simply be getting a deserved chance to fail, as did the symbolic approach.«<sup>181</sup> Über diesen Stand ist die Diskussion um KI bis heute im Wesentlichen nicht hinausgekommen.

175 Vgl. John Hopfield: *Neural Networks and Physical Systems with Emergent Collective Computational Abilities*, Proceedings of the National Academy of Sciences 79, April 1982, S. 2554-2558.

176 Vgl. David E. Rumelhart/James L. McClelland/PDP Research Group: *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, 2 Bde., Cambridge, Mass. 1986-1987.

177 *Schwache* Emergenz erklärt höherstufige Phänomene aus der komplexen Interaktion niedrigstufiger; *starke* Emergenz dagegen lehnt selbst diese Erklärung ab.

178 »[T]he apparent application of rules could readily emerge from interactions among simple processing units rather than from application of any higher level rules.« (David E. Rumelhart/James L. McClelland: »PDP Models and General Issues in Cognitive Science«, in: dies./PDP Research Group (Hg.): *Parallel Distributed Processing* (Anm. 174), Bd. 1: *Foundations*, S. 110-146, hier S. 120) »We are simply trying to understand the essence of cognition as a property emerging from the interactions of connected units in networks. [...] [W]e can't understand the behavior of networks of neurons from the study of isolated neurons.« (Ebd., S. 128)

179 Ebd., S. 128.

180 Dreyfus: *What Computers Can't Do* (Anm. 169), S. 32.

181 Hubert L. Dreyfus/Stuart Dreyfus: »Making a Mind versus Modeling the Brain. Artificial Intelligence Back at a Branchpoint«, in: *Daedalus* 117.1 (1988), S. 15-53, hier S. 37.

Und dennoch: Trotz Rumelharts und McClellands Anrufung der Gestalttheorie ist das Selbstverständnis der KI-Forschung auch im subsymbolischen Paradigma, das heute als *deep learning* firmiert, noch immer nicht tatsächlich holistisch: Phänomene schwacher Emergenz sind zwar nicht mehr explanatorisch reduktiv, aber genetisch ist weiterhin ein Bottom-up-Prozess am Werk, der auch ingenieurstechnisch umgesetzt wird. So operiert die erfolgreiche Architektur des *convolutional neural network* (CNN), das zum Erkennen und Generieren von Bildern eingesetzt wird, gewissermaßen gestaffelt emergent. Es baut auf der Entdeckung der Hirnforscher David Hubel und Torsten Wiesel auf, dass der visuelle Cortex Zellen besitzt, die nur auf Linien in bestimmten Winkeln reagieren, aber hintereinander angeordnet komplexe Muster codieren.<sup>182</sup> Ähnlich funktioniert auch das CNN: Erkennt die erste Schicht des Netzwerks Kanten, werden sie in der zweiten Schicht zu Ecken und Rundungen zusammengesetzt, in der dritten zu Objekten und so weiter.<sup>183</sup> Für Yann LeCun, KI-Forscher, der wesentlich an der Entwicklung neuronaler Netze in den 1980er Jahren beteiligt war, ist der Grund offensichtlich: »This works because the world is compositional.«<sup>184</sup> Die Struktur der Welt wird in der Struktur des Repräsentationsmodells der Welt wiederholt. So formuliert, wird aus einer technischen Lösungsanweisung eine ontologische Aussage, die ganz in Übereinstimmung mit klassischen atomistischen Welterklärungsmodellen steht, sei es das Carnaps oder das der Wittgenstein'schen Bildtheorie.<sup>185</sup>

Damit ergibt sich für die Gegenwart der paradoxe Befund, dass neuronale Netze unter atomistischen Annahmen holistisch operieren. Technisch könnte man sie, mit einem Begriff von John von Neumann, als »mixed systems« bezeichnen,<sup>186</sup> auf der

182 Vgl. David H. Hubel/Torsten N. Wiesel: »Receptive Fields of Single Neurons in the Cat's Striate Cortex«, in: *The Journal of Physiology* 148.3 (1959), S. 574-591. Auf dieser Forschung aufbauend konnte 1981 Geoffrey Hinton, Teil der PDP-Gruppe, nicht nur darauf aufmerksam machen, dass es möglich sein sollte, den visuellen Kortex mit neuronalen Netzen zu simulieren, sondern auch, dass hiermit Gestalteffekte erklärt wären; vgl. Geoffrey Hinton: »Shape Representation in Parallel Systems«, in: *Seventh International Joint Conference on Artificial Intelligence*, Vancouver, B. C. 1981, S. 1088-1096. Weiter ausgeführt hat er diesen Ansatz als Erklärung von Teil-Ganzes-Verhältnissen in Geoffrey E. Hinton: »Mapping Part-Whole Hierarchies into Connectionist Networks«, in: *Artificial Intelligence* 46.1-2 (1990), S. 47-75.

183 Vgl. Yann LeCun/Yoshua Bengio/Geoffrey Hinton: »Deep learning«, in: *Nature* 521 (Mai 2015), S. 436-444, hier S. 439.

184 Yann LeCun: »The Power and Limits of Deep Learning and AI«, zit. nach einer Vorlesungsmitschrift von Kiran Vodrahalli, 20.11.2017, [https://kiranvodrahalli.github.io/notes/live-Tex/yann\\_lecun\\_nov20.pdf](https://kiranvodrahalli.github.io/notes/live-Tex/yann_lecun_nov20.pdf) (aufgerufen am 15.02 2021).

185 Dass aus dieser ontologischen auch eine theologische Aussage folgen kann, kulminiert im apokryphen, aber oft dem Mathematiker Stuart Geman zugeschriebenen Aperçu »either the world is compositional or God exists« – entweder die Welt lässt sich atomistisch erklären oder es bedarf eines Gottes, um höhere Funktionen sicherzustellen; hier wird die Idee der Gestalt erneut der Mystik gezogen.

186 John von Neumann: *The Computer and the Brain*, New Haven 1958, S. 58-60.

konzeptionellen Ebene wären sie »quasi-holistisch«<sup>187</sup> zu nennen – weder völlig atomistisch, weil sie ohne explizite Regeln auskommen und emergent funktionieren, noch tatsächlich holistisch, weil sie immer noch bei der Verarbeitung kleinster Elemente ansetzen und natürlich weiterhin auf digitalen Systemen ausgeführt werden. Und insofern ›Gestalt‹ stets auch ein Verstehen und nicht nur ein Erkennen der Sache impliziert, sind sie weiterhin gestaltfremd. Denn von der Köhler’schen Einsicht sind auch die heutigen, beeindruckend leistungsfähigen KI-Modelle noch weit entfernt – nicht zuletzt, weil ihnen das *commonsense reasoning* abgeht und sie keine verkörperte Kognition besitzen, wie bereits Dreyfus feststellte. Zudem fehlt subsymbolischen Systemen gerade jene logische Inferenzleistung, die symbolische Systeme auszeichnet, so dass die Integration der Stärken beider Architekturen noch immer ein Desideratum darstellt.<sup>188</sup>

Die Problemfelder der Gestaltpsychologie beschäftigen die KI-Forschung also weiterhin. Während Wertheimers Kritik an der Mosaik- oder Bündelthese über emergente Interaktionen von Elementen durchaus lösbar ist, wie etwa im CNN, wiegt der Vorwurf der Assoziationsthese sehr viel schwerer: Noch immer sind KI-Systeme lediglich statistische Modelle, die ein Ereignis A mit einem Ereignis B korrelieren, aber nichts über deren kausale Verknüpfung sagen können. Für den Informatiker Judea Pearl ist daher klar, dass intelligente Systeme erst durch die Umstellung von Korrelation auf Kausalität möglich werden.<sup>189</sup>

Die Kognitions- und Neurowissenschaften rücken dagegen zusehends von ihrer Verschränkung mit der KI-Forschung ab. War die Identifikation des Gehirns als digital codierte Maschine bei Warren McCulloch noch eine ernstgemeinte Forschungsheuristik, wurde sie später zur hilfreichen, aber eben uneigentlichen Metapher abgewertet<sup>190</sup> und wird heute selbst als solche infrage gestellt.<sup>191</sup> Das Feld steht, so der Kognitionswissenschaftler Matthew Cobb, womöglich vor einem Metaphernwechsel, der das Informations-

187 Ich führe diesen Gedanken aus in Hannes Bajohr: »Die ›Gestalt‹ der KI. Jenseits von Holismus und Atomismus«, in: *Zeitschrift für Medienwissenschaft* 12.2 (2020), S. 168-181; ich wandle dafür den Begriff »quasi-analog« ab; vgl. Andreas Sudmann: »Szenarien des Postdigitalen. Deep Learning als MedienRevolution«, in: Christoph Engemann/Andreas Sudmann (Hg.): *Machine Learning – Medien, Infrastrukturen und Technologien der Künstlichen Intelligenz*, Bielefeld 2018, S. 55-73.

188 Dies wäre ein »master algorithm«, der die Vorteile aller Ansätze kombiniert; vgl. Pedro Domingos: *The Master Algorithm. How the Quest for the Ultimate Learning Machine Will Remake Our World*, New York 2015.

189 Vgl. Judea Pearl/Dana MacKenzie: *The Book of Why. The New Science of Cause and Effect*, New York 2018.

190 Vgl. S. Ryan Johnson: »The Brain’s Software. The Natural Languages and Poetic Information Processing«, in: Hermann Haken/Uno Svedin (Hg.): *The Machine as Metaphor and Tool*, Berlin 1993, S. 9-44.

191 Vgl. Romain Brette: »Is Coding a Relevant Metaphor for the Brain?« in: *Behavioral and Brain Sciences* 42.215 (2019), S. 1-58. Freilich gibt es, zumal in der Neurophilosophie, weiterhin eine einflussreiche Richtung, die diese Metapher aufrechterhält, vgl. Patricia Churchland/Terrence J. Sejnowski: *The Computational Brain. 25th Anniversary Edition*, Cambridge, Mass. 2017, wobei auch hier das Gestaltproblem wieder auftaucht, vgl. S. 83.

paradigma hinter sich lassen mag – wobei unklar ist, was folgen wird. Dagegen erscheint starke Emergenz, die Ganzheitsphänomene nicht atomistisch unter dem Rekurs auf ihre konstitutiven Elemente erklären will, heute wieder attraktiver.<sup>192</sup> Die Gestaltpsychologie ist hier freilich kein Bezugspunkt mehr, aber ihre Grundfragen fordern die Forschung weiterhin heraus.

*Hannes Bajohr*

192 Vgl. Matthew Cobb: *The Idea of the Brain. The Past and Future of Neuroscience*, New York 2020, S. 374f.