



universität  
wien

# DIPLOMARBEIT / DIPLOMA THESIS

Titel der Diplomarbeit / Title of the Diploma Thesis

Theorie der erweiterten Kognition: Eine kritische Analyse

verfasst von / submitted by

Sebastian English

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of  
Magister der Philosophie (Mag. phil.)

Wien, 2020 / Vienna, 2020

Studienkennzahl lt. Studienblatt /  
degree programme code as it appears on  
the student record sheet:

UA 190 406 299

Studienrichtung lt. Studienblatt /  
degree programme as it appears on  
the student record sheet:

Lehramtsstudium UF Mathematik UF Psychologie  
und Philosophie

Betreut von / Supervisor:

ao. Univ.-Prof. Mag. DDr. Andreas Hergovich Bakk.

# Theorie der erweiterten Kognition: Eine kritische Analyse

Sebastian Englich

31. Januar 2020

# Inhaltsangabe

|          |   |           |
|----------|---|-----------|
| <b>1</b> | <b>Einleitung</b>   | <b>4</b>  |
| <b>2</b> | <b>Kontextualisierung I: Das Computerbild von Kognition</b> | <b>7</b>  |
| 2.1      | Der (lange) Weg zur Kognitionswissenschaft . . . . .        | 7         |
| 2.1.1    | Logik . . . . .   | 7         |
| 2.1.2    | Psychologie und Linguistik . . . . .                        | 16        |
| 2.1.3    | Künstliche Intelligenz . . . . .                            | 18        |
| 2.1.4    | Die drei Ebenen von Marr . . . . .                          | 21        |
| 2.2      | Die Repräsentationale Theorie des Geistes . . . . .         | 23        |
| 2.3      | Anwendungen . . . . .                                       | 30        |
| 2.3.1    | SHRDLU . . . . .  | 30        |
| 2.3.2    | SHAKEY . . . . .  | 32        |
| 2.4      | Zusammenfassung . . . . .                                   | 34        |
| <b>3</b> | <b>Kontextualisierung II: Konnektionismus</b>               | <b>36</b> |
| 3.1      | Grundaufbau . . . . .                                       | 37        |
| 3.2      | Lernen . . . . .  | 39        |
| 3.3      | Charakteristika . . . . .                                   | 41        |
| 3.4      | Kognitivismus vs. Konnektionismus . . . . .                 | 42        |
| 3.5      | Kritik . . . . .  | 44        |
| 3.6      | Zusammenfassung . . . . .                                   | 49        |
| <b>4</b> | <b>Kontextualisierung III: Situierete Kognition</b>         | <b>51</b> |

|          |  |           |
|----------|--|-----------|
| 4.1      | Zwei Merkmale situierter Kognition . . . . .               | 54        |
| 4.2      | Methode . . . . .  | 55        |
| 4.3      | Künstliches Leben . . . . .                                | 57        |
| 4.4      | Theorie Dynamischer Systeme . . . . .                      | 61        |
| 4.5      | Zusammenfassung . . . . .                                  | 65        |
| <b>5</b> | <b>Der dynamische Komputationalismus von Andy Clark</b>    | <b>67</b> |
| 5.1      | Die evolutionsbiologisch-ökologische Perspektive . . . . . | 68        |
| 5.1.1    | Mechanismen intelligenten Verhaltens . . . . .             | 70        |
| 5.1.2    | Die Rolle des Gehirns . . . . .                            | 75        |
| 5.1.3    | Intelligentes Verhalten als emergentes Phänomen . . . . .  | 77        |
| 5.1.4    | Methodischer Pluralismus . . . . .                         | 78        |
| 5.1.5    | Die Bedeutung von Repräsentationen . . . . .               | 82        |
| 5.1.6    | Das konnektionistische Gehirn . . . . .                    | 84        |
| 5.1.7    | Fazit . . . . .  | 87        |
| 5.2      | Die kulturelle Perspektive . . . . .                       | 87        |
| 5.2.1    | Kognitive Nischenbildung . . . . .                         | 90        |
| 5.2.2    | Die Mangrovenbaum-Metapher . . . . .                       | 91        |
| 5.2.3    | Turingmaschine reloaded . . . . .                          | 92        |
| 5.3      | Zusammenfassung . . . . .                                  | 94        |
| <b>6</b> | <b>These der erweiterten Kognition</b>                     | <b>96</b> |
| 6.1      | These der erweiterten Kognition . . . . .                  | 97        |
| 6.1.1    | Ausgangspunkt . . . . .                                    | 98        |
| 6.1.2    | Schlussfolgerung . . . . .                                 | 99        |
| 6.1.3    | Einordnung . . . . .                                       | 103       |
| 6.1.4    | Gehirn und Körper . . . . .                                | 104       |
| 6.2      | Kritik . . . . .   | 106       |
| 6.2.1    | Der Kopplungs-Konstitutions-Fehlschluss . . . . .          | 107       |
| 6.2.2    | Ein Kennzeichen für Kognition . . . . .                    | 108       |
| 6.2.3    | Intrinsische Inhalte . . . . .                             | 110       |

|          |   |            |
|----------|---|------------|
| 6.2.4    | Natural Kinds . . . . .                         | 111        |
| 6.2.5    | Die These der eingebetteten Kognition . . . . . | 114        |
| 6.3      | Zusammenfassung . . . . .                       | 116        |
| <b>7</b> | <b>Fazit</b>                                    | <b>118</b> |
|          | <b>Literaturverzeichnis</b>                     | <b>124</b> |

# Kapitel 1

## Einleitung

Mit der *Theorie der erweiterten Kognition* hat der Philosoph Andy Clark die These aufgestellt, dass kognitive Prozesse (manchmal) auch *nicht-neuronale* Ressourcen umfassen können. In diesen Fällen dehnt sich die kognitive Aktivität des Menschen über die eigenen biologischen Grenzen hinweg auf Teile der Umwelt aus und konstituiert dadurch, so Clark, einen *erweiterten* kognitiven Prozess.

Ursprünglich wurde diese Erweiterungsthese von Andy Clark in Zusammenarbeit mit David J. Chalmers in dem Artikel *The Extended Mind* im Jahre 1998 zum Ausdruck gebracht und führte wenig später bereits, zu intensiven Debatten innerhalb der Kognitionswissenschaft und der Philosophie des Geistes.

Mit der hier vorliegenden Arbeit soll nun der Versuch unternommen werden, möglichst genau darzustellen, was genau unter der These der erweiterten Kognition (TEK) zu verstehen ist und von welchen Vorstellungen sie sich dabei anleiten lässt. Es wird sich dabei zeigen, dass ihre Grundlagen tief im Gesamtwerk von Andy Clark verwurzelt sind, weshalb ein wesentlicher Teil der Arbeit auch darin bestehen wird, die der Erweiterungsthese zugrundeliegenden theoretischen Anschauungen von Clark klar zu umreißen.

Um dies gewährleisten zu können, ist es notwendig, auf die bedeutsamsten theoretischen Strömungen innerhalb der Kognitionswissenschaft und deren Überzeugungen einzugehen. Andy Clark vereint in seiner Theorie sowohl Vorstellungen, die sich im Bereich der traditionellen Kognitionswissenschaft — auch *Kognitivismus* genannt — verorten lassen, als auch Ideen aus dem *Konnektionismus* und der *Theorie situierter Kognition*.

Der originäre Beitrag von Clark besteht darin, dass er zu all diesen Ansätzen kritisch Stellung bezieht, mit eigenen Gedanken anreichert und zu einer ganz eigenständigen kognitionswissenschaftlichen Theorie ausarbeitet. Dieser *dynamische Komputationalismus* stellt gewissermaßen eine Synthese aus Kognitivismus, Konnektionismus und situierter Kognition dar, die eine ganz neuartige Sicht auf kognitive Phänomene eröffnet.

Erst vor diesem Hintergrund wird es möglich, die These der erweiterten Kognition vollumfänglich nachzuvollziehen und einer kritischen Untersuchung unterziehen zu können. Dabei wird auch herausgearbeitet, dass Clark mit seiner Erweiterungsthese nicht *ausschließlich* eine kognitionswissenschaftliche Hypothese im engeren Sinne postuliert, sondern mit ihr auf eine tiefliegende metaphysische Grundüberzeugung verweist, die ihrer Stoßrichtung nach eine philosophisch-anthropologische Bestimmung des Wesens des Menschen darstellt.

Insgesamt ist die Arbeit wie folgt strukturiert: In Kapitel 2 wird ein kurzer Abriss über jene wichtigen wissenschaftlichen Disziplinen gegeben, die maßgeblich an der Hervorbringung der Kognitionswissenschaft als eigenen Forschungszweig beigetragen haben. Aus diesen kristallisierte sich ein erstes zentrales Leitmotiv heraus, nämlich der Vorstellung von (menschlicher) Intelligenz, als ein durch informationsverarbeitende Prozesse hervorgebrachtes Phänomen. In enger Analogie zu digitalen Computern etablierte sich dort bald das sogenannte *Computerbild des Geistes*. Ebenfalls wird hier die *Repräsentationale Theorie des Geistes* vorgestellt, die wohl einflussreichste Theorie des Geistes innerhalb der Kognitionswissenschaft. Schließlich werden auch zwei Arbeiten aus der klassischen KI-Forschung präsentiert werden.

In Kapitel 3 wird die Theorie des *Konnektionismus* vorgestellt, die als großer Gegenspieler der traditionellen Sichtweise von Kognition angesehen werden kann. Die künstliche Modellierung neuronaler Netzwerke hat zu ganz neuartigen Vorstellungen über das Wesen des Kognitiven geführt und soll deshalb ebenfalls in angemessenem Umfang in der Arbeit abgebildet werden.

Die *Theorie situierter Kognition* begründet wiederum eine alternative Sichtweise auf das Wesen intelligenten Verhaltens. Anhand der *Künstliche-Leben-Forschung* und der *Theorie Dynamischer Systeme* wird aufgezeigt, dass das Nachdenken über Intelligenz auch jenseits informationsverarbeitender Prozesse vorgestellt werden kann. Dies sind die In-

halte von Kapitel 4.

Kapitel 5 schließlich widmet sich eingehend den Gedanken von Andy Clark und versucht schrittweise, dessen *dynamischen Komputationalismus* aus einer Diskussion verschiedenster Aspekte des Kognitivismus, Konnektionismus und situierter Kognition heraus zu entwickeln.

Im vorletzten Kapitel schließlich wird die These der erweiterten Kognition vorgestellt und einer kritischen Untersuchung unterzogen.

Im abschließenden Kapitel wird, ausgehend von der TEK, versucht zu zeigen, dass in ihr auch eine tiefliegende metaphysische Überzeugung von Clark liegt: Menschen gelten Clark ihrer Natur nach als *natural-born cyborgs*.

# Kapitel 2

## Kontextualisierung I: Das Computerbild von Kognition

[...] cognition is computation, computation operates over symbols, symbols begin with inputs to the brain and end with outputs from the brain, and so it is in the brain alone that cognition takes place and it is with the brain alone that cognitive science need concern itself.

(Shapiro, 2011, 27f.)

### 2.1 Der (lange) Weg zur Kognitionswissenschaft

#### 2.1.1 Logik

Schon seit jeher interessierte sich die Menschheit für das Wesen logisch vernünftigen Denkens. Mit der im 4. Jahrhundert v. u. Z. von Aristoteles entwickelten, klassischen Logik (auch *Syllogistik* genannt), wurde erstmals ein umfassender Katalog logisch korrekter Schlussfolgerungen (das sind: Syllogismen) erstellt. Wodurch, so kann in Betrachtung solcher Syllogismen gefragt werden, zeichnet sich die *Gültigkeit* einer logischen Schlussfolgerung aus? Auf Basis des wohl bekanntesten Syllogismus, soll dies anschaulich gemacht werden.

|                  |                              |
|------------------|------------------------------|
| Prämisse 1       | Alle Menschen sind sterblich |
| Prämisse 2       | Aristoteles ist ein Mensch   |
| Schlussfolgerung | Aristoteles ist sterblich    |

Die Vorgehensweise innerhalb der aristotelischen Logik bleibt dabei immer dieselbe: Zunächst werden zwei Annahmen (bzw. Prämissen) aufgestellt aus denen in weiterer Folge der Versuch unternommen wird, einen logisch korrekten Schluss zu ziehen. Offensichtlich ist die im oben angeführten Beispiel abgeleitete Schlussfolgerung gültig.

*Gültigkeit* bedeutet in diesem Zusammenhang, dass aus der Wahrheit der Prämissen sich unmittelbar auch die Wahrheit des Schlusses ergibt. Wahrheit und Gültigkeit sind dabei zwei streng voneinander zu unterscheidende Begriffe.

*Wahrheit* bezieht sich immer nur auf den *Inhalt* einer Aussage. Manche Aussagen drücken, um mit Wittgenstein zu sprechen, einen wahren Sachverhalt über die Welt aus, und genau solche Aussagen sind dann auch wahr.

Gültigkeit hingegen hat im eigentlichen Sinne nichts mit der Wahrheit von Aussagen zu tun. Sie ist kein Merkmal von Aussagen, sondern ein Kennzeichen logischer Schlussfolgerungen, d.h. selbst aus falschen Aussagen können logisch korrekte Schlussfolgerungen abgeleitet werden. Dies ist deshalb möglich, weil logische Schlüsse immer dann schon gültig sind, wenn in ihnen etwas Wahres ausgesagt wird, vorausgesetzt dass auch die Prämissen wahr sind. Die logische Gültigkeit achtet also nicht auf den Wahrheitsgehalt der Prämissen; es geht ihr einzig darum sicherzustellen, dass im Falle des Wahr-Seins von Prämissen auch die daraus gezogene Schlussfolgerung wahr ist. Ihr Fokus liegt damit auf der *Form* eines gezogenen Schlusses unter Ausblendung der inhaltlichen Ebene.

Dies kann auch an obigem Beispiel deutlich gemacht werden. Wird dort nämlich für *Menschen* die Variable  $X$ , für *sterblich* ein  $y$ , sowie  $Z$  für *Aristoteles* eingesetzt, werden die formalen Eigenschaften deutlich sichtbar:

|                  |                   |
|------------------|-------------------|
| Prämisse 1       | Alle $X$ sind $y$ |
| Prämisse 2       | $Z$ ist ein $X$   |
| Schlussfolgerung | $Z$ ist $y$       |

Setzt man nun für  $X$  und  $Z$  ein beliebiges Subjekt oder Objekt und für  $y$  ein beliebiges Prädikat ein, so kann festgestellt werden, dass die jeweilige Schlussfolgerung stets etwas

Korrektes aus den Prämissen folgern wird.

Freilich ergibt sich aus inhaltlich falschen Prämissen (in der Regel) auch ein unwahrer Schluss; umgekehrt allerdings ist auch klar, dass aus wahren Prämissen niemals eine falsche Schlussfolgerung gezogen werden kann. Ist also sichergestellt, dass die Prämissen wahr sind, so stellen die formalen Eigenschaften einer logisch gültigen Argumentationsfigur sicher, dass daraus stets auch eine wahre Aussage abgeleitet wird.

Die formale Logik unterscheidet deshalb zwischen einer *semantischen* und *syntaktischen* Ebene logischer Systeme. Die semantische Ebene betrachtet dabei die inhaltliche Seite von Aussagen, das heißt, es geht hier um die *Bedeutung* logischer Ausdrücke. Demgegenüber beschäftigt sich die syntaktische Ebene ausschließlich mit den formalen Eigenschaften von Schlussfolgerungen.

Zwischen diesen beiden Ebenen besteht allerdings ein wichtiger Zusammenhang, der bereits aus dem Begriff der logischen *Gültigkeit* hervorgeht: Aus inhaltlich wahren Prämissen kann (für den Fall, dass syntaktisch korrekt abgeleitet wurde) nie eine falsche Aussage abgeleitet werden: „Just follow the rules and truth will be preserved“ (Clark, 2014, S. 9).

In der Mitte des 19. Jahrhunderts nahm die formale Logik, insbesondere durch die Arbeiten des deutschen Mathematikers und Logikers Gottlob Frege, einen entscheidenden Entwicklungsschritt. Frege setzte sich zur Aufgabe, die Logik zum unumstößlichen Fundament der Mathematik zu machen und sah sich deshalb der Herausforderung gestellt, diese erheblich ausbauen zu müssen. Die Logik müsste damit nämlich in der Lage sein, sämtliche mathematischen Erkenntnisse in eine Sprache der Logik zu übersetzen (Davis, 2011, S. 41), und dafür reichte die aristotelische Logik bei weitem nicht aus. Frege entwickelte deshalb in seinem Werk *Begriffsschrift. Eine der arithmetischen nachgebildete Formelsprache des reinen Denkens*, eine auf der Logik basierende, künstliche Sprache, die diesem Erfordernis entsprechen sollte.

Eine dieser Ansprüche, die diese neue logische Sprache einbringen musste, war eine Abbildung der Tiefenstruktur *individueller* Propositionen. Die aristotelische Logik war dafür ungeeignet, weil diese bloß Beziehungen *zwischen* Propositionen, also zwischen Prämissen und einer Schlussfolgerung, abbilden, nicht aber die logische Struktur einer einzelnen Proposition selbst sichtbar machen konnte. Frege musste dementsprechend ein

logisches System mit einer höheren Auflösungsrate erarbeiten und erkannte, dass „the same relations that connect propositions can also be used to analyze the structure of individual propositions, and he made these relations the basis of his logic“ (Davis, 2011, S. 41).

Dies kann anhand der beiden Aussagen *Alle lieben jemanden* und *Jemand liebt alle* deutlich gemacht werden. Sieht man sich diese beiden Sätze mit Hilfe der aristotelischen Logik an, so kann über diese beiden lediglich gefolgert werden, dass sie sich voneinander unterscheiden. Die elementaren Basiseinheiten der Syllogistik sind Propositionen und auf ihnen kann nur erkannt werden, ob zwei Aussagen dieselben sind oder nicht. Ganz anders stellt sich dies nun für das von Frege entwickelte logische System dar. Mit Hilfe eines als Allquantor bezeichneten Symbols  $\forall$ , das *für alle* bedeutet und einem durch das Symbol  $\exists$  gekennzeichneten Existenzquantors, welcher *es existiert* meint, können die obigen beiden Aussagen folgendermaßen dargestellt werden:

$$\begin{array}{l|l} \text{Alle lieben jemanden:} & (\forall x)(\exists y) : x \text{ liebt } y. \\ \text{Jemand liebt alle:} & (\exists x)(\forall y) : x \text{ liebt } y. \end{array}$$

Die Einführung des All- und Existenzquantors führt im Ergebnis dazu, dass hier nicht nur die Unterschiedlichkeit der beiden Aussagen zu erkennen ist, sondern auch, *worin genau* sie sich voneinander unterscheiden (Davis, 2011, S. 44). Freges Sprache ist also in der Lage, tiefer in die Struktur von Propositionen einzudringen und besitzt insofern auch eine höhere Auflösungsrate als jene der Syllogistik.

Des Weiteren verlangte die Abbildung mathematischer Aussagen einer ihr gerecht werdenden logischen Sprache und selbstverständlich auch eine sehr strenge „grammatikalische“ Struktur. Die Mathematik zeichnet sich allem voran durch ihre Genauigkeit aus. Diese Eigenschaft muss also auch jedes logische System berücksichtigen, das eine Basis für die Mathematik sein möchte.

Die grammatikalische Struktur einer logischen Sprache, auch als *Syntax* bezeichnet, sucht, wie zuvor bereits bei der aristotelischen Logik gezeigt wurde, nach jenen regelbasierten Beziehungen, die als Schlussfolgerung im Lichte ihrer Prämissen als logisch korrekt anzusehen sind.

Frege setzte sich hierbei zum Ziel, diese regelbasierten Beziehungen so exakt zu for-

mulieren, dass diese im Prinzip auch rein *mechanisch* ausgeführt werden könnten. Mit anderen Worten sollte eine Person unter Befolgung dieser Regeln selbst dann richtige Schlussfolgerungen in Freges System ableiten können, wenn diese (abgesehen von diesen Regeln) keinerlei Verständnis dafür besitzt, was sie hier gerade macht. Um dies verständlich zu machen, sei im Folgenden kurz die wichtigste dieser Regeln vorgestellt:

Gilt für zwei beliebige Sätze  $A$  und  $B$ , dass

- (i)  $A$  gilt, und
- (ii)  $A \supset B$  gilt, dann folgt
- (iii) daraus, dass auch  $B$  gilt.<sup>1</sup>

Die Regel macht deutlich, dass das System Freges ausschließlich regelbasiert, d.h. syntaktisch arbeitet und deshalb auch keinerlei inhaltliche Interpretation benötigt. Freilich wurde die Sprache dahingehend entwickelt, dass sich mit ihr sinnvolle Aussagen der Mathematik abbilden ließen; allerdings kann allein auf Basis der Regeln mit der Sprache gearbeitet werden, d.h. ohne zu wissen, was die verwendeten Symbole eigentlich bedeuten. Gilt sowohl  $A$ , als auch  $A \supset B$ , dann sagt oben angeführte Regel aus, dass daraus die Gültigkeit von  $B$  schlussgefolgert werden kann. Dafür ist es eben nicht notwendig zu wissen, wofür etwa das Symbol  $\supset$  steht; auch ist dabei unerheblich zu wissen, welche inhaltliche Bedeutung die Sätze  $A$  und  $B$  besitzen (Davis, 2011, S. 45). Formale Logiken sind (auf der syntaktischen Ebene) also nichts weiter, als „systems comprising sets of symbols, ways of joining the symbols so as to express complex propositions, and rules specifying how to legally derive new symbol complexes from old ones“ (Clark, 2014, S. 9). Frege konnte dabei zeigen, dass eine solche Menge von Symbolen und syntaktischen Regeln dazu in der Lage ist, sämtliche Aussagen der Mathematik nicht nur darzustellen, sondern ebenfalls deren Gültigkeit nachweisen.

---

<sup>1</sup>Kurz: ( $A$  und  $A \supset B$ ) impliziert  $B$ , bzw.  $(A \ \& \ (A \supset B)) \rightarrow B$ .

Sieht man, so wie viele das tun, die Mathematik als Inbegriff logischen Denkens an, dann erscheint es nachvollziehbar, dass mit der von Frege entwickelten Logik sogar erstmalig ein System entwickelt wurde „that could plausibly account for all of human deductive reasoning“ (Davis, 2011, S. 129).

Was allerdings fehlte, war die Existenz einer Methode zur Bestimmung der Gültigkeit einer spezifischen, in der logischen Sprache Freges ausgedrückten, mathematischen Formel (Davis, 2011, S. 49). Sämtliche mathematisch bereits bewiesenen Formeln ließen sich ohne weiteres auch in Freges logischem System als korrekt ausweisen; ob allerdings innerhalb dieses Systems auch etwas über die Korrektheit mathematisch *unbewiesener* Formeln ausgesagt werden kann, blieb demgegenüber offen.

Das vom deutschen Mathematiker David Hilbert formulierte *Entscheidungsproblem* wies genau auf diese Lücke hin. Das Entscheidungsproblem stellt damit also die Frage, ob eine unbewiesene, in eine logische Zeichenkette übersetzte, mathematische Aussage, innerhalb des logischen Systems von Frege, auf ihre Gültigkeit hin überprüft werden kann. Eine solche Überprüfung müsste im Kern aus einem Verfahren bestehen, das sich aus einer endlichen Anzahl wohldefinierter Schritte bzw. Operationen zusammensetzt (Davis, 2011, S. 87).

Solche Verfahren werden auch als *Algorithmus* bezeichnet. *Algorithmen* sind ihrem Wesen nach Verfahren, die, ausgehend von einem Startwert, in einer endlichen Anzahl an Umtransformationen dieses Startwerts, ein Ergebnis liefern (Crane, 2003, S. 88). Das Verfahren bei der schriftlichen Multiplikation von Zahlen etwa stellt ein Beispiel eines mathematischen Algorithmus dar.

Auf das Entscheidungsproblem wurde auch der britische Mathematiker und Logiker Alan Turing aufmerksam, der sich infolgedessen eingehend um die Lösung dieses Problems bemühte. Den Ausgangspunkt seiner Überlegungen stellte dabei die Frage nach dem Wesen von *Berechnungen* dar (Barker-Plummer, 2016), denn wenn das von Hilbert eingeforderte Verfahren im Kern einen Algorithmus bezeichnet und Algorithmen, mathematisch gesprochen, als eine Methode zur regelgeleiteten Berechnung von Funktionen angesehen werden können, dann muss in diesem Zusammenhang insbesondere ein Verständnis des Begriffes der *Berechnung* erarbeitet werden. Die Identifikation eines Algorithmus mit der Berechnung von Funktionswerten ergibt sich daraus, dass jedes

regelgeleitete mathematische Verfahren auf die Ausführung einer Funktion rückgeführt werden kann. Beispielsweise ist auch die Multiplikation zweier Zahlen nichts weiter als die Abbildung zweier zuvor multiplizierter Zahlen auf das Multiplikationsergebnis.

Funktionen können in einer (noch allgemeineren) Betrachtungsweise wiederum als (mathematische) Objekte verstanden werden, die einen gegebenen Input auf regelbasierter Weise in einen Output umwandeln. Als berechenbar gilt eine Funktion dann, wenn ein Algorithmus zur Berechnung ihrer Werte existiert (Crane, 2003, S. 89). Das Entscheidungsproblem war demzufolge also aufs Engste mit dem Begriff des *Algorithmus* verbunden und dieser wiederum verwies mittels des Funktionsbegriffes auf regelgeleitete Umtransformationen (von Input in Output).

Turing fokussierte sich nun, im Hinblick auf Algorithmen, weniger auf deren Regelbasiertheit, sondern richtete stattdessen seine Aufmerksamkeit „to what [a] person actually did when carrying them out“ (Davis, 2011, S. 131). Er stellte dabei fest, dass sich die Ausführung eines Algorithmus unter Angabe einiger weniger sequentieller Schritte charakterisieren ließ (Davis, 2011, S. 134):

1. Der gesamte Berechnungsprozess ergibt sich aus einem sequentiellen Ablauf von Schritten, die das Eintragen bzw. Löschen von Symbolen in die (teilweise mit Symbolen beschriebenen) Kästchen eines linierten Papierbandes zur Folge haben.
2. Bei jedem einzelnen Schritt des Berechnungsprozesses richtet die Person ihre Aufmerksamkeit auf exakt ein Kästchen auf dem Papierband.
3. Dort wird die Handlung der Person allein durch den Eintrag in das Kästchen und durch jenen Zustand festgelegt, der sich aus dem Verlauf vorangegangener Schritte ergibt.
4. Die Handlung selbst besteht schließlich in der Eintragung bzw. Löschung eines Symbols in diesem Kästchen und einer anschließenden Verlagerung der Aufmerksamkeit auf das unmittelbar daneben liegende Kästchen.

Jeder nur denkbare Algorithmus kann demnach also prinzipiell auf Basis einer endlichen Abfolge dieser vier Schritte von einer Person ausgeführt werden. „Turing’s next

step was to see that the person could be replaced by a machine capable of performing these basic actions“ (Davis, 2011, S. 131). Damit allerdings ist nicht gemeint, dass Turing in weiterer Folge dann an der Realisierung eines materiellen Gerätes arbeitete; vielmehr übersetzte er diese informelle Beschreibung zur Ausführung eines beliebigen Algorithmus nun in ein rigoroses mathematisches Modell (Rescorla, 2017), bei dem es sich im Kern um eine „[...] abstract, theoretical specification of a possible machine“ (Crane, 2003, S. 92) handelte, einem „[...] abstract model of an idealized computing device [...]“ (Rescorla, 2017), der Symbole in genau derselben Art und Weise manipuliert, wie Personen dies in seiner Analyse zur Ausführung einer Berechnung getan haben (Rescorla, 2017).

Diese abstrakte Spezifikation eines idealisierten Berechnungssystems wird seinem Erfinder zu Ehren als *Turingmaschine* bezeichnet. Turingmaschinen geben dem Begriff des Algorithmus damit eine ungemein präzise, mathematische Bedeutung und legen darüber hinaus auch das Wesen von Berechenbarkeit offen (Crane, 2003, S. 92). Jede algorithmische Berechnung kann diesem Verständnis nach in so einfache Subaufgaben zerlegt werden, dass diese von einer Turingmaschine ausgeführt werden können (Crane, 2003, S. 108).

What is important is that on the basis of Turing’s analysis of the notion of computation, it is possible to conclude that anything computable by any algorithmic process can be computed by a Turing machine. So if we can prove that some particular task can not be accomplished by a Turing machine, we can conclude that no algorithmic process can accomplish that task.

(Davis, 2011, 134f.)

In Bezug auf das oben erwähnte *Entscheidungsproblem* konnte Turing zeigen, dass das dort eingeforderte Verfahren unmöglich von einer Turingmaschine ausgeführt werden kann. Nachdem Turing allerdings zeigen konnte, dass eine Turingmaschine de facto jedes algorithmische Verfahren durchführen kann, hatte dies zur Konsequenz, dass eine Methode zur Lösung des Entscheidungsproblems nicht existiert.

Die Turingmaschine ist ein theoretisches, rein mathematisches Modell einer Maschine, doch bald schon wurde erkannt, dass sie auch das Potential besitzt, tatsächlich verwirklicht zu werden. Eine solch materiell realisierte Turingmaschine wäre dazu in der Lage, in automatisierter Form algorithmische Berechnungen durchzuführen. Gesteuert würde diese Maschine durch einen Maschinencode, und dieser Code wäre im Kern nichts anderes als eine formallogische Sprache im Sinne Freges. Turing legte damit also den Grundstein für die Computerwissenschaften und schaffte mit dem Modell der Turingmaschine eine Blaupause zur Schaffung von Computern (Rescorla, 2017).

Wenn (i) logisch vernünftiges Denken im Kern mit mathematischem Nachdenken gleichgesetzt werden kann, (ii) mathematische Aussagen in der logischen Sprache Freges ausgedrückt werden können, (iii) eine Turingmaschine sämtliche Algorithmen auf mechanische Weise zum Ausführen imstande ist und (iv) schließlich ein Computer die physikalische Realisierung einer Turingmaschine ist, die Berechnungen auf Basis einer logischen Symbolsprache durchführt, so scheint all dies folgenden Schluss nahezulegen:

The physical world suddenly included, clear, nonevolved, nonorganic examples of what Daniel Dennett would later dub ‚syntactic engines‘— quasi-autonomous systems whose sheer physical makeup ensured (under interpretation) some kind of ongoing reason-respecting behavior.

(Clark, 2014, S. 13)

Es kann dementsprechend auch nicht allzu viel überraschen, dass bald schon das menschliche Gehirn in Analogie zu Computern gesetzt wurde. Könnte es sein, dass menschliche Intelligenz ihrem Wesen nach eine Form algorithmischer Symbolverarbeitung ist? Damit eine solche Analogie an Plausibilität gewinnen kann, ist es notwendig, „[...] to investigate in detail what thinking and other mental processes are. So this approach will need a psychological theory behind it: for it will need to figure out what the processes are before finding out what sort of computational mechanisms carry out these processes [...]“ (Crane, 2003, 115f.).

### 2.1.2 Psychologie und Linguistik

Innerhalb der Psychologie formierte sich etwa zur selben Zeit eine allmähliche Kritik an (besonders in den USA dominierenden) behavioristischen Theorien. Der Behaviorismus besitzt eine große (methodische) Skepsis gegenüber der Annahme mentaler Zustände beim Menschen und will demgegenüber psychologische Untersuchungen allein auf von außen beobachtbaren Verhaltens einschränken: „According to behaviorists, psychology is really the science of behavior“ (Bermúdez, 2014, S. 7), wobei mentale Eigenschaften hier dann als bloße Dispositionen für ein bestimmtes Verhalten interpretiert werden (Clark, 2014, S. 253). In Bezug auf lerntheoretische Fragestellungen behauptet der Behaviorismus, dass (i) alles Lernen das Ergebnis von Konditionierung ist und (ii) Konditionierungen sich aus assoziativen und verstärkenden Prozessen ergeben.

Allerdings konnte ein im Jahr 1930 von Tolman und Honzik durchgeführtes Experiment an Ratten zeigen, dass diese sich auch ohne jegliche Art von Konditionierung eine räumliche Vorstellung von einem Labyrinth bildeten, in das sie gesetzt wurden. Ein Folgeuntersuchung im Jahre 1946 legte überdies hinaus den Schluss nahe, dass Ratten dabei eine Art *kognitive* Landkarte des Labyrinths erstellten (Bermúdez, 2014, 7ff.). Wahrnehmungsinhalte wurden also zur systematischen Erstellung einer räumlichen Abbildung des Labyrinths verwendet, welches ihnen dann als Orientierungshilfe zur Verfügung stand. Intelligentes Verhalten kann demnach nicht ausschließlich über die Außenperspektive eines Lebewesens verstanden werden. Offensichtlich geschehen im Gehirn zielgerichtete Verarbeitungen von Umweltreizen und dienen als Grundlage für Handlungen.

Diese und ähnliche Experimente führten dazu, dass der Behaviorismus allmählich an Bedeutung verlor, und sich an seiner statt *informationsverarbeitende* Modelle zur Erklärung von Verhalten durchsetzten.

Nicht nur in der Psychologie, sondern auch seitens der Linguistik erfuhr der Behaviorismus heftige Kritik. Insbesondere Noam Chomsky und die von ihm 1957 entworfene *Theorie der Generativen Grammatik* setzte behavioristische Sprachtheorien unter Druck. Chomsky argumentierte, „[...] that language (specifically grammar) could not be learned by the principles of Skinnerian learning theory, and that much of what we know about our

languages must be innate“ (Garnham, 2009, S. 103). Er bezweifelte dabei, dass Kinder, allein auf Basis gehörter Sprache ihres Umfelds dazu in der Lage sind, ein ausreichendes Verständnis der ihr zugrunde liegenden Grammatik entwickeln zu können. Zweifellos ist es für das Erlernen von Sprache notwendig, in ein sprachliches Umfeld hineingeboren zu werden; allerdings variiert ein dort vernommener sprachlicher Input so stark, „[t]hat there is next to no prospect of generalizations framed over the states of mature linguistic competence and input kinds“ (Collins, 2011, S. 168). Das Erlernen der Prinzipien und Regeln einer Sprache kann Chomsky zufolge also nicht allein durch externe Sprachinputs erfolgen; vielmehr muss zusätzlich auch von der Existenz *angeborener innerer* Mechanismen im Gehirn ausgegangen werden (Figdor, 2019, S. 86). Chomsky vertritt damit eine *internalistische Auffassung von Sprache*: Sprache und Spracherwerb können nur durch eine genaue Untersuchung der Anwender von Sprache verständlich gemacht werden. Sprachliche Fähigkeiten sind demnach auf kognitive Fähigkeiten zurückzuführen (Collins, 2011, S. 167).

Dementsprechend geht Chomsky in seiner Theorie der Generativen Grammatik dann auch davon aus, dass Menschen über ein bereits angeborenes, internes Wissenssystem von Sprache besitzen, welches im Wesentlichen eine *universelle Grammatik* umfasst. Eine solche Grammatik setzt sich dabei aus sogenannten *Phrasenstruktur-* und *Transformationsregeln* zusammen. Während *Phrasenstrukturregeln* es ermöglichen, grammatikalisch korrekte Sätze zu bilden, tragen *Transformationsregeln* dafür Sorge, neuartige Satzkonstruktionen aus gegebenen Satzkonstruktionen auf korrekte Art und Weise zu generieren. *Universell* ist dieses grammatikalische Wissenssystem deshalb, weil Chomsky davon ausgeht, „that the underlying principles of grammar are the same in all languages“ und ein solches sprachliches Wissenssystem dazu befähigt „to learn any human language“ (Garnham, 2009, S. 103).

Damit lieferten die Entwicklungen in Psychologie und Linguistik mit ihrer Hinwendung auf angeborene, informationsverarbeitende Prozesse des Gehirns einen fruchtbaren Nährboden für die aus den Computerwissenschaften hervorgehende Analogie zwischen Computer und Gehirn, die wiederum den zentralen Rahmengedanken der neu aufstrebenden Kognitionswissenschaft darstellte. Diese, als *Computerbild des Geistes* bezeichnete, Leitidee betrachtete die geistigen Fähigkeiten des biologischen Gehirns analog zu den

Berechnungsprozessen digitaler Computer und schien den Schlüssel zum Verständnis menschlicher Intelligenz in sich zu tragen:

The attractions of such a view can hardly be overstated. It makes the mental special without making it ghostly. It makes the mental depend on the physical, but in a rather complex and [...] liberating way. And it provides a ready-made answer to a profound puzzle: how to get sensible, reason-respecting behavior out of a hunk of physical matter.

(Clark, 2014, 8f.)

Man schien damit also nun auch für den Geist die Möglichkeit einer rein materialistischen Deutung gefunden zu haben: Gedanken und Denkprozesse werden vom Gehirn realisiert. Dieses Gehirn ist, seinem Organisationsprinzip zufolge, ein (biologischer) Computer, dem Wahrnehmungsreize als Input zugeführt werden, und diesen Input anschließend regelgeleitet in Erkenntnisse und Handlungsanweisungen übersetzt. Was die Untersuchungen formallogischer Systeme und Computer zutage gefördert hat, ist, dass Intelligenz durch (automatisierte) physikalische Symbolsysteme realisiert werden kann (Clark, 2014, S. 30) und also nahelegt, dass „[...] mental activity [...] be modelled as the processing of information using an internal symbol system, with discrete abstract symbols“ (Garnham, 2009, S. 99). Vernunftangeleitetes Handeln „[...] it seems, makes good scientific sense if we imagine a neural economy organized as a syntax-driven engine that tracks the shape of semantic space“ (Clark, 2014, S. 16).

### 2.1.3 Künstliche Intelligenz

In diesem Zusammenhang ist auch der wesentliche Beitrag der *Künstliche-Intelligenz-Forschung* (KI-Forschung) hervorzuheben, der für die Entwicklung des Computerbildes des Geistes eine wichtige Rolle spielte. Gegenstandsbereich der KI-Forschung ist die Nachbildung intelligenter Prozesse. „[I]t aims to construct computing machines that execute core mental tasks such as reasoning, decision-making, problem solving, and so on“ (Rescorla, 2017). Setzte sie sich in ihrer Anfangszeit allein mit der Simulation bzw. Imitation von Intelligenz auseinander, so erhob sie allmählich immer mehr auch den Anspruch, Intelligenz in Maschinen tatsächlich auch realisieren zu können. Die KI-Forschung

war damit bald auch schon am Projekt zur Erklärung menschlicher Intelligenz intensiv beteiligt:

The philosophically interesting idea behind AI is the idea of building a thinking computer [...]. Obviously, this is an interesting question in itself; but, [...] the project of building a thinking computer should help us understand what intelligence (or thought) is in general. That is, by building a thinking computer, we can learn about thought

(Crane, 2003, S. 115)

Ihren markantesten Ausdruck findet das Computerbild des Geistes in der sogenannten *Hypothese über physikalische Symbolsysteme*, kurz HPSS. Bevor diese hier vorgestellt wird, ist es allerdings entscheidend, sich die grundlegende Funktionsweise eines Computers zu vergegenwärtigen. Ein Computer ist im Kern eine physikalische Realisierung eines automatisch arbeitenden, formallogischen Systems. Physische Eingaben (oder auch interne Computersignale) werden im Computer auf Basis einer symbolischen Computersprache regelbasiert verarbeitet und in Ausgaben überführt. Diese Computersprache ist ihrem Wesen nach eine logische Sprache, ganz analog zu jener die Gottlob Frege entwickelt hat. „From this point of view, the Begriffsschrift was the ancestor of all computer programming languages common use today“ (Davis, 2011, S. 45). Die in dieser Programmiersprache verwendeten Symbole und Ausdrücke repräsentieren dabei den Computerinput, und Computerberechnungen verarbeiten diese Repräsentationen anschließend in systematischer Weise in Ausgabesignale (Crane, 2003, S. 131). Ein Computer ist damit ein Informationsverarbeiter: Er manipuliert die (als Repräsentationen für den Input fungierenden) Symbole einer Programmiersprache auf algorithmische Weise, und ist damit ein kausaler Mechanismus über Repräsentationen (Crane, 2003, S. 83).

Die HPSS nun postuliert, dass sämtliches intelligentes Verhalten auf regelgeleitete Transformationen physikalischer Symbole rückgeführt werden kann (Bermúdez, 2014, S. 141). Sie wurde von den beiden Informatikern Herbert Simon und Allen Newell erstmalig formuliert und geht in Bezug auf *menschliche* Intelligenz davon aus, dass „cognitive information processing has to be understood in terms of the rule-governed transformation of

physical symbols. This way of thinking about information processing is inspired by the metaphor of the mind as computer“ (Bermúdez, 2014, S. 139).

Die HPSS ist insofern eine recht starke These über Intelligenz, als sie behauptet, dass ein physikalisches Symbolsystem, also ein Computer, sowohl die *notwendigen* als auch *hinreichenden* Eigenschaften zur Hervorbringung von Intelligenz besitzt (Bermúdez, 2014, S. 142). Die HPSS

[...] does not simply hold that the mind is like a computing system. [It] holds that the mind literally is a computing system. [It] holds that a suitable abstract computational model offers a literally true description of core mental processes.

(Rescorla, 2017)

Menschliches Denken wird damit mit den Berechnungen einer (physikalisch realisierten) Turingmaschine gleichgesetzt (Rescorla, 2017), zeichnet intelligentes Verhalten durch ihnen zugrunde liegende, informationsverarbeitende Prozesse aus und sieht Informationsverarbeitung selbst wiederum als durch regelgebundene Manipulationen von Symbolen bestimmt.

Zwar weist die HPSS damit schon eine recht eindeutige Richtung in Bezug auf die Frage, auf, auf welche Art Intelligenz in der Welt in Erscheinung tritt; gleichzeitig bleibt sie aber bei Fragen der Implementierung stumm. Wie exakt Intelligenz physikalisch in Maschinen oder Lebewesen realisiert wird bzw. realisiert werden kann, findet in der HPSS keine Berücksichtigung. In Bezug auf menschliche Intelligenz stellt die HPSS damit einen eher *regulativen* Denkraum über die qualitative Struktur menschlichen Denkens dar (Bermúdez, 2014, S. 142) und enthält sich auf Ebene der Implementierung ihrer Stimme.

Methodisch betrachtet ist dies allerdings durchaus so gewollt. Die HPSS, und damit das Computerbild des Geistes insgesamt, will ja eben auch deutlich machen, dass die konkrete materielle Realisierung eines formallogischen Symbolsystems eigentlich nebensächlicher Natur ist. Die Möglichkeit, Computer aus unterschiedlichsten Materialien zusammensetzen zu können, ändert nichts an deren Wesen als Computer. Computer zeichnen sich auf einer abstrakten, nicht-materiellen Ebene dadurch aus, dass sie automatisierte formallogische Rechensysteme sind. Wird der Geist mit einem informationsverarbeitenden

und regelgeleiteten System identifiziert, dann lassen sich die essentiellen Fragestellungen in Bezug auf Denkprozesse, gerade auch unter Auslassung biologischer Eigenschaften des Gehirns, verständlich machen. So wie es innerhalb der Informatik möglich ist, eine Trennung zwischen Hardware- und Softwareebene durchzuführen, so ist auch bei der menschlichen Intelligenz eine Trennung zwischen Neurowissenschaften einerseits und der Kognitions- bzw. Intelligenzforschung andererseits methodisch sinnvoll.

Zweifellos besteht dabei zwischen den neuronalen Vorgängen im Gehirn und mentalen Prozessen ein enger Zusammenhang; dies impliziert allerdings nicht, dass mentale Prozesse nicht auch auf einer autonomen Ebene beschrieben werden können, d.h. ohne Berücksichtigung neuronaler Eigenschaften des Gehirns.

Aus diesem Verständnis heraus ergab sich dann auch der Tätigkeitsbereich der Kognitionspsychologie, die sich fortan auf die Ausarbeitung informationsverarbeitender Modelle hin ausrichtete (Bermúdez, 2014, S. 29).

#### **2.1.4 Die drei Ebenen von Marr**

Das Aufgabengebiet der Psychologie kann auch an den vom Psychologen David Marr eingeführten drei Ebenen zur Beschreibung eines kognitiven Prozesses verständlich gemacht werden. Mithilfe dieser drei Ebenen wollte Marr zeigen, welches Verhältnis zwischen einer (i) psychologischen, (ii) algorithmischen und (iii) neurologischen Betrachtungsweise menschlicher Intelligenz besteht (Garnham, 2009, S. 107). Kognitive Leistungen des Menschen können demzufolge immer auf dreifache Weise beschrieben werden.

Auf der sogenannten *Rechenebene* (computational level) soll ein kognitiver Tätigkeitsbereich identifiziert werden, und es soll erklärt werden, welchen Zweck diese kognitive Leistung für das System erfüllt. Auf der daran anschließenden *algorithmischen Ebene* (algorithmic level) muss verständlich gemacht werden, woraus genau sich diese kognitive Leistung funktional ergibt. Es muss hier also geklärt werden, aus welchen Komponenten sich eine gegebene kognitive Leistung zusammensetzt und wie diese miteinander interagieren. Es geht hier also um eine detaillierte Beschreibung des Bauplans jenes Algorithmus, der gefragte kognitive Leistung hervorzubringen vermag. Auf der dritten und letzten Ebene schließlich, die Marr als *Ebene der Implementation* (implementational le-

vel) bezeichnet, ist zu untersuchen, wie der auf der zweiten Ebene ermittelte Algorithmus neuronal im Gehirn realisiert worden ist (Bermúdez, 2005, 18f.).

Die (kognitive) Psychologie ist, diesem Schema zufolge, also sowohl auf der Rechen-ebene, als auch auf der algorithmischen Ebene tätig. Aus Sicht der HPSS jedoch hat die algorithmische Ebene die größte Bedeutung, weil hier die informationsverarbeitenden Prozesse ermittelt werden. Zur Modellierung kognitiver Prozesse auf algorithmischer Ebene greift die Kognitionspsychologie auf eine *Komponentenanalyse* zurück, die im wesentlichen eine *Zweck-Mittel-Analyse* ist und manchmal auch als *Boxologie* bezeichnet wird. Ziel ist es, mit ihrer Hilfe eine funktionale Beschreibung eines kognitiven Prozesses zu liefern, d.h. aufzuzeigen, aus welchen kognitiven *Subeinheiten* sich dieser zusammensetzt. „Processes (such as sentence comprehension) and functions (such as memory) were divided and subdivided and represented in models containing interconnected sets of boxes representing subprocesses or subfunctions“ (Garnham, 2009, S. 105). Sie ist damit eine „[...] analysis of the working of the machine into the functions of its component parts“ (Crane, 2003, S. 105); wie ein Uhrmacher durch das Zerlegen einer Uhr in ihre einzelnen Bestandteile sich deren Funktionsweise verständlich machen kann, so will hier auch die Psychologie kognitive Prozesse in ihre funktionalen Einheiten aufspalten, um dadurch ein genaueres Bild von ihnen zu bekommen.

In den Anfangsjahren wurde also den neuronalen Prozessen biologischer Gehirne wenig Beachtung geschenkt. Dies hing unter anderem auch damit zusammen, dass Fragen der Implementation als weniger zentral angesehen wurden: Während die algorithmische Ebene das eigentliche Wesen von Kognition herauszuarbeiten hatte, mussten auf Ebene der Implementation trivialere Fragen technischer Natur einer Antwort zugeführt werden.

Damit nun ist also auch das theoretische Selbstverständnis der Kognitionswissenschaft offengelegt: „Cognitive Science is the interdisciplinary study of mind and intelligence, embracing philosophy, psychology, artificial intelligence, neuroscience, linguistics, and anthropology“ (Thagard, 2019). Sie ist interessiert am Wesen intelligenter Prozesse, geht traditionellerweise davon aus, dass diese durch regelhafte Berechnungen charakterisiert sind, und ihr Ziel ist demzufolge eine möglichst exakte Beschreibung jener informationsverarbeitenden Prozesse, die kognitiven Leistungen zugrunde liegen (Shapiro, 2011,

S. 14).

Die Kognitionswissenschaft, die sich seit ca. 1970 fest innerhalb der anderen Wissenschaften etablierte (Bermúdez, 2014, S. 5), machte sich damit auf den Weg, das Wesen des Geistes zu entmystifizieren:

Mind was not ghostly stuff, but the operation of a formal computational system implemented in the meatware of the brain [...] Mindware, it was claimed, is to the neural meat machine as software is to the computer.

(Clark, 2014, S. 13)

## 2.2 Die Repräsentationale Theorie des Geistes

Innerhalb der Philosophie des Geistes, der Kognitionswissenschaften und auch der Psychologie ist die führende philosophische Grundvorstellung immer eine an der Theorie des *Funktionalismus* ausgerichtete Sichtweise (Heil, 2013, S. 87). Der Funktionalismus kann direkt auf die Entwicklung der ersten Computer in den Jahren zwischen 1950-1960 zurückgeführt werden (Heil, 2013, S. 88) und vertritt die Ansicht, dass Systeme genau dann einen Geist besitzen, wenn sie über eine bestimmte *funktionale* Organisationsweise verfügen.

Mental states are states that play appropriate roles in the system's functional organization. Each mental state is individuated by its interactions with sensory input, motor output, and other mental states.

(Rescorla, 2017)

Wird über die funktionale Organisation eines Systems gesprochen, so bezeichnet dies eine Beschreibungsebene, bei der der materiellen Zusammensetzung des Systems keinerlei Beachtung geschenkt wird. Wird in funktionalistischer Hinsicht über das Mentale und mentale Prozesse gesprochen, so werden ebene Aspekte, die zu deren Realisierung beitragen, außen vorgelassen und der Fokus auf eine abstraktere Ebene gerichtet.

Der Funktionalismus als *metaphysische* bzw. *ontologische* Position geht davon aus, dass mentale Eigenschaften von Systemen auf einer höheren Ebene liegen, als deren materielle Eigenschaften. In Bezug auf das menschliche Gehirn würde dies also bedeuten, dass mentale Eigenschaften keine neuronalen Eigenschaften sind, sondern auf einer höheren Stufe liegen (Heil, 2013, S. 93). „Minds are not identifiable with brains [...] Talk of minds is at bottom talk of material systems at a ‚higher level‘, a level that abstracts from the ‚hardware‘ [...]“ (Heil, 2013, S. 90). Dabei will eine funktionalistische Theorie allerdings nicht zum Ausdruck bringen, dass das Mentale völlig unabhängig von der materiellen Ebene ist. Geistige Zustände und Eigenschaften werden zweifellos durch ihr zugrundeliegende materielle Zustände und Eigenschaften eines Systems realisiert bzw. instantiiert; nichtsdestotrotz geht der Funktionalismus davon aus, dass das Mentale nicht mit etwas Materiellem identifiziert werden kann. „Pains, for instance, are, according to functionalists, realized in the nervous system. But the property of being in pain is not a material property“ (Heil, 2013, S. 92). So wie etwa im Hinblick auf Computer zwischen einer Hardware- und Softwareebene unterschieden werden kann, so gilt dies auch in Bezug auf das Gehirn und das Geistige.

Funktionale Eigenschaften werden dabei anhand *kausaler Rollen* bestimmt (Heil, 2013, S. 96), d.h. ein System besitzt genau dann eine bestimmte funktionale Eigenschaft, wenn etwas innerhalb dieses Systems eine bestimmte (kausale) Rolle einnimmt. Verdeutlicht kann dies anhand eines Unternehmens werden, dass im Folgenden nun als System aufgefasst werden soll. Vizepräsident dieses Unternehmens zu sein kann bspw. als eine funktionale Rolle aufgefasst werden; in die Funktion des Vizepräsidenten tritt eine Person nicht kraft irgendeiner materiellen Eigenschaft, sondern weil sie innerhalb des Unternehmens eine (abstrakte) Position einnimmt, die in ganz bestimmten Beziehungen zu den übrigen Berufspositionen und Abteilungen dieses Unternehmens steht. Diese Beziehungen können durch ein Organigramm dargestellt werden, und dieses Beziehungsmuster ist es dann auch, welches die Position des Vizepräsidenten eindeutig allen anderen Positionen gegenüber auszeichnet. Prinzipiell könnte jede Person diese Rolle im Unternehmen einnehmen; es kommt wie gesagt nicht auf irgendwelche physischen Eigenschaften dieser Person an, um Vizepräsident werden zu können; was eine Person als Vizepräsidenten auszeichnet, ist, dass sie ebenjenen Platz in der funktionalen Organisation des Betriebes

ausfüllt, die sie als Vizepräsident eindeutig bestimmen lässt.

Allgemeiner gesprochen ist eine kausale Rolle im Sinne des Funktionalismus dadurch gekennzeichnet, dass sie einen bestimmten Input in einen adäquaten Output transformiert (Rescorla, 2017). In Bezug auf einen physischen Schmerzzustand würde dies bedeuten, dass das „[w]hat makes the state a realization of pain, [...] is not its material makeup, but its playing a particular kind of causal role within [the] nervous system“ (Heil, 2013, S. 98).

To be in such and such a mental state is simply to be a physical device, of whatever composition, that satisfies a specific formal description. Mindware, in humans, happens to run on a meat machine. But the very same mindware [...] might run in some silicon device or in the alien organic matter of a Martian.

(Clark, 2014, S. 14)

Der Funktionalismus zielt damit auf eine *strukturelle* Betrachtungsweise mentaler Entitäten ab: Auf funktionaler Ebene werden Zustände und Eigenschaften nicht durch Charakteristika identifiziert, die diesen selbst eigen sind, sondern durch jene spezifischen Beziehungen bzw. Relationen, die sie anderen Entitäten gegenüber aufweisen.

Für die Kognitionswissenschaften stellt der sogenannte *Maschinenfunktionalismus*, einer speziellen Strömung innerhalb des Funktionalismus, den zentralen Bezugspunkt dar. Dieser gilt als „[...] leading philosophical offspring of the developments in artificial intelligence [...]“ (Clark, 2014, S. 14) und wird auch als *klassische komputationale Theorie des Geistes* bezeichnet. Er wurde in der Mitte der 1960er Jahre von Hilary Putnam entwickelt und vertritt, der HPSS sehr ähnlich, die These, dass mentale Aktivitäten durch Berechnungen einer Turingmaschine realisiert werden (Rescorla, 2017). Ein mentaler Zustand wird auch hier über das Ausfüllen einer bestimmten (kausalen) Rolle charakterisiert; hier allerdings wird diese kausale Rolle schon etwas genauer spezifiziert, namentlich als eine bestimmte informationelle Berechnung einer Turingmaschine.

Schließlich entwickelte sich daraus in den 1970er Jahren die von Jerry Fodor begründete *Repräsentationale Theorie des Geistes* (Heil, 2013, S. 110), einer speziellen Variante

des Maschinenfunktionalismus (Heil, 2013, S. 120). Fodor betrachtete mentale Aktivitäten ebenfalls als informationsverarbeitende Prozesse einer Turingmaschine, gab darüber hinaus ein genaueres Bild über das Wesen dieser Prozesse. (Rescorla, 2017).

Um die von Fodor erstellte Theorie bestmöglich nachvollziehbar zu machen, soll hier über den Weg der Psychologie eine Hinführung zu dessen Gedanken erfolgen und zwar anhand folgenden Problems:

[...] we are, by and large, successful in explaining and predicting other people's behavior in terms of what they believe about the world and what they want to achieve. This success is something that itself needs explanation. Why is it that our vocabulary of beliefs and desires (our belief-desire psychology or propositional attitude psychology) is so deeply ingrained and indispensable in our social interactions and social coordination.

(Bermúdez, 2014, S. 152)

Auf personaler Ebene besitzen Menschen ein intuitives Verständnis jener Motive, die den Handlungen ihrer Mitmenschen zugrunde liegen. Dieses *alltagspsychologische* Wissen des Menschen lässt dabei das Handeln des eigenen Umfeldes nicht nur verständlich machen, sondern besitzt darüber hinaus auch das Potential, zukünftige Handlungen anderer Personen prognostizieren zu können. Aus Sicht der wissenschaftlichen Psychologie ist es dabei von Interesse, erklären zu können, wodurch sich dieses alltagspsychologische Wissen des Menschen ergibt. Diese Problemstellung hat Bermúdez (2005, S. 35) als *Schnittstellenproblem* (interface problem) bezeichnet und wie folgt umrissen:

„How does commonsense psychological explanation interface with the explanations of cognition and mental operations given by scientific psychology, cognitive science, cognitive neuroscience and the other levels in the explanatory hierarchy?“

In welchem Verhältnis steht die (auf personaler Ebene in Erscheinung tretende) Alltagspsychologie des Menschen zu subpersonalen Ebenen bzw. — um mit David Marr zu sprechen — welche Mechanismen der algorithmischen Ebene bzw. der Ebene der Implementierung bringen dieses Wissen ursächlich hervor? Alltagspsychologische Erklärungen treten auf personaler Ebene in Form von propositionalen Einstellungen in Erscheinung. Das beobachtete Verhalten der Mitmenschen kann dahingehend verständlich

gemacht werden, als diesen Menschen gewisse, auf Sachverhalte bezogene, Überzeugungen, Hoffnungen, Ängste o.ä. unterstellt werden, die für ihr tatsächliches Verhalten kausal wirksam sind.

Für Jerry Fodor sind alltagspsychologische Erklärungen und Prognosen deshalb so erfolgreich, weil die Annahme, dass menschliches Handeln durch *intentionale innere Zustände* verursacht wird, auch tatsächlich wahr ist. Er vertritt damit einen *Realismus in Bezug auf propositionale Einstellungen*, womit gemeint ist, dass die (auf personaler Ebene wirkenden) propositionellen Einstellungen auf subpersonaler Ebene durch physikalisch real existierende (neuronal) Prozesse erzeugt werden. Die Struktur geistiger Zustände wird durch eine ihr zugrunde liegende neuronale Struktur realisiert (Crane, 2003, S. 137), wobei die beiden Strukturen zueinander *ähnlich* sind. Aus dieser *strukturellen Ähnlichkeit* zwischen propositionellen Einstellungen einerseits und den sie realisierenden neuronalen Strukturen andererseits folgt für Fodor, dass das menschliche Gehirn über eine Art *Sprache des Geistes* verfügt: Alltagspsychologische Erklärungen fungieren über propositionelle Einstellungen und diese wiederum sind als sprachliche Einheiten anzusehen. Denken Menschen über das Verhalten ihrer Mitmenschen nach, so geschieht dies auf sprachliche Art und Weise. Fodor zieht aus dieser Erkenntnis und der Annahme struktureller Gleichheit zwischen Intentionen und ihren neuronalen Realisierern den Schluss, dass solche neuronalen Strukturen im Kern *sprachliche Symbole* repräsentieren und bestimmte Aktivitäten des Gehirns diese sprachlichen Symbole regelgeleitet manipulieren.

Diese Auffassung lässt Fodor schließlich seine bekannte *Hypothese über eine Sprache des Geistes* (HSG) formulieren. Sie kann als spezifische Variante der HPSS angesehen werden und behauptet, dass neuronale Strukturen ihrem Wesen nach Symbole sind und diese Symbole das Alphabet einer Sprache des Geistes darstellen.

According to Fodor's language of thought hypothesis, the basic symbol structures in the mind that carry information are sentences in an internal language of thought (sometimes called Mentalese). Information processing works by transforming those sentences in the language of thought.

(Bermúdez, 2014, S. 151)

Als Grundelement dieser Sprache dienen Sätze, wobei diese Sätze ihre inhaltliche Bedeutung aus kleineren semantischen Einheiten, ähnlich den Wörtern der Sprachen von Menschen, erhalten. Diese symbolischen „Atome“ besitzen Inhalte, weil sie für bestimmte Sachverhalte der Welt eintreten, d.h. sie sind Repräsentationen von Aspekten der Welt (Crane, 2003, S. 140). Wie bei anderen Sprachen muss bei der Sprache des Geistes ebenfalls zwischen *Satzinhalt* und *Satzausdruck* unterschieden werden. Bei gesprochener Sprache wird der Satz durch das Gesprochene ausgedrückt, d.h. als Träger bzw. Vehikel des Satzinhaltens fungieren hier Schallwellen. Diese Unterscheidung zwischen den *Trägern* von Inhalten und den Inhalten *selbst* kann analog zur semantischen und syntaktischen Ebene logischer Sprachen verstanden werden. Satzausdrücke beschreiben die Art und Weise, wie Inhalte zum Ausdruck gebracht werden, sind gewissermaßen eine formale Beschreibung des in die Welt Kommens von Inhalten. Diese Unterscheidung wird in weiterer Folge wichtig sein, wobei die Frage, wodurch Sätze ihre inhaltliche Bedeutung erlangen, so wichtig sie ist, hier nicht näher betrachtet wird.

Die Repräsentationale Theorie des Geistes „provides a way of understanding how minds, higher-level entities, could systematically affect and be affected by bodily goings-on“ (Heil, 2013, S. 110) und postuliert die Existenz einer Sprache des Geistes, einem „[...] system of symbols that function as ‚mental representations‘. These symbols make up [...] a biologically fixed ‚code‘ analogous to the ‚machine code‘ hard-wired into an ordinary computing machine“ (Heil, 2013, S. 110).

Hier also ist jener Punkt erreicht, an dem sich das von der Psychologie ins Treffen geführte Schnittstellenproblem mit dem Computerbild des Geistes vollends verbindet. Das menschliche Gehirn wird als biologischer Computer betrachtet, und kognitive Aktivitäten des Gehirns als über eine (biologische) Programmiersprache vollzogene Informationsverarbeitung vorgestellt. Ihrem Wesen nach wird diese Sprache eine große Ähnlichkeit zu formallogischen Sprachen wie der von Frege aufweisen.

Menschen sind dieser Sichtweise zufolge also deshalb in der Lage, die Welt um sie herum sprachlich zu deuten, weil ihr biologisches Gehirn eine Art *semantische Maschine* ist. Zwar operiert auch das Gehirn ganz wie ein Computer über uninterpretierte Symbole, d.h. es macht sich keinen Begriff über die inhaltliche Seite der Symbole, allerdings werden die, auf syntaktischen Regeln basierenden, symbolischen Berechnungen so durchgeführt,

dass sie auch zu keinen verzerrten inhaltlichen Ergebnissen führen (Heil, 2013, S. 112). „Symbols in the Language of Thought [...] are presumed to have meanings. The idea is just that these meanings play no role in the operations of devices that process them“ (Heil, 2013, S. 118).

[A]t the subpersonal level, functional states are realized by physical structures [...] and the causal relations into which those physical structures enter are not a function of their semantic properties. The semantic properties that are so important at the personal level no longer have a role to play at the subpersonal level.

(Bermúdez, 2005, S. 74)

Auf der personalen Ebene stellt sich das alltagspsychologische Wissen des Menschen als inhaltlich bedeutungsvoll dar; realisiert werden diese Inhalte aber durch subpersonale Prozesse, die selbst wiederum nur formale Eigenschaften der Symbole der Sprache des Geistes berücksichtigen.

Damit also scheint eine gute Antwort auf das Schnittstellenproblem gefunden worden zu sein: Die Alltagspsychologie gibt den Menschen (zuverlässig) Auskunft über das beobachtete Verhalten der Mitmenschen, indem sie dieses Verhalten auf die Existenz propositioneller Einstellungen zurückführt. Diese propositionellen Einstellungen existieren gemäß der Repräsentationalen Theorie des Geistes tatsächlich und werden durch Prozesse im Gehirn realisiert. Die Gehirnprozesse sind genau solche Berechnungsprozesse, wie sie von modernen Computern durchgeführt werden. Berechnet wird über symbolische Repräsentationen, die inhaltliche Aspekte der Welt abbilden, durch elektrochemische Signale im Gehirn in Erscheinung treten und als Alphabet einer Sprache des Geistes fungieren. Diese Sprache des Geistes ist analog zu einer Programmiersprache, stellt einen biologischen Code zur Einschreibung von Informationen für das Gehirn zur Verfügung und ist, seiner Funktionsweise nach, analog zu formalen Sprachen der Logik anzusehen.

Damit also liefert die Repräsentationale Theorie des Geistes eine Vorstellung über die kognitive Architektur des menschlichen Gehirns.

## 2.3 Anwendungen

Mit der Formulierung der Hypothese über physikalische Symbolsysteme (HPSS) und der Repräsentationalen Theorie des Geistes (RTG) als einem an der HPSS ausgerichteten Modell einer kognitiven Architektur ist die traditionelle Sichtweise der Kognitionswissenschaft als einem Computerbild des Geistes fest umrissen. Dieses klassische Verständnis von Kognition wird in weiterer Folge auch unter den Begriff des *Kognitivismus* gefasst werden.

In diesem Abschnitt werden nun zwei Arbeiten aus der KI-Forschung vorgestellt, die sich zur Modellierung intelligenten Verhaltens stark von der traditionellen Sichtweise der Kognitionswissenschaft haben anleiten lassen.

### 2.3.1 SHRDLU

Der Spracherwerb gilt als eine der herausragendsten kognitiven Fähigkeiten des Menschen. Zugleich gilt dieser als geradezu paradigmatischer Fall eines *regelbasierten* Lernprozesses. Einerseits muss auf syntaktischer Ebene eine Menge grammatikalischer Regeln zur korrekten Bildung von Sätzen erlernt werden; andererseits erfordert der semantisch korrekte Umgang mit Sprache die richtige Zuordnung zwischen Wörtern und die durch sie bezeichneten Objekten und Sachverhalten. Auch dies kann regelbasiert vorgestellt werden.

Der Kognitivismus schien dahingehend über den richtigen theoretischen Rahmen zur Modellierung sprachlicher Fähigkeiten zur Verfügung zu stellen, weshalb bereits sehr früh künstliche Sprachprogramme entwickelt wurden, von denen sich erhofft wurde, wichtige Einblicke in die kognitiven Mechanismen von Sprachverständnis zu erhalten.

Das Sprachprogramm SHRDLU etwa wurde in den 1970er Jahren von Terry Winograd am MIT im Zuge seiner Dissertation erstellt.

SHRDLU setzt sich dabei aus mehreren Teilprogrammen zusammen. Eines dieser Teilprogramme bettet die sprachliche Kernkomponente in eine virtuelle Mikrowelt ein, die sich aus farbigen geometrischen Körpern und einer Box zusammensetzt. Die sprachliche Kernkomponente, auf die sich in weiterer Folge der Begriff *SHRDLU* beziehen soll, ist dabei auf doppelte Weise mit dieser Welt verbunden: Einerseits besitzt sie ein sprachlich

formuliertes, inneres Modell dieser Mikrowelt, das sich aus einer Liste logischer Sätze zusammensetzt und im Zuge von Änderungen der Mikrowelt aktualisiert wird; andererseits kann sie auf einen virtuellen Roboterarm zugreifen, mittels dessen sie mit der Mikrowelt interagieren kann.

Damit das innere Weltmodell ausgelesen werden kann, verfügt SHRDLU über einen einprogrammierten Wortschatz samt einfacher grammatikalischer Regeln. Mittels gewisser Eingabebefehle ist es möglich, schriftliche Aufträge an SHRDLU zu übermitteln, die dieser dann mittels seines Roboterarms in der Mikrowelt ausführen soll.

Der eigentliche Zweck von SHRDLU besteht also darin, die von außen an ihn herangetragenen Anweisungen sprachlich wahrzunehmen, zu interpretieren und umzusetzen. Wird etwa der Eigenname eines geometrischen Objekts mit dem Befehl *CLEARTOP* kombiniert, so ist SHRDLU angewiesen, ebenjenes Objekt von allen anderen Objekten, die sich auf diesem befinden, freizuräumen.

Dafür stehen SHRDLU mehrere Subsysteme zur Verfügung. Ein syntaktisches System wird zur grammatikalischen Analyse des schriftlichen Auftrages eingesetzt. Es klärt dabei, ob es sich bei diesen um eine syntaktisch korrekt formulierte Zeicheneingabe handelt. Sofern keine Fehler entdeckt wurden, übernimmt das semantische System und gleicht die Inhalte des Auftrages mit dem eigenen Wortschatz und den Inhalten des Weltmodells ab. So SHRDLU den Auftrag auch semantisch interpretieren kann, muss nun das kognitive System einen Handlungsplan ausarbeiten. Im Kern verwendet es dafür den Auftrag als gewünschten Zielzustand, gleicht diesen mit dem aktuellen Zustand der Mikrowelt ab und versucht nun über eine Reihe logisch-sequentieller Schritte, vom derzeitigen Zustand hin zum Zielzustand zu gelangen. Im Ergebnis führt dies zur Erstellung einer Liste von Handlungsanweisungen, die an den virtuellen Roboterarm gesendet werden.

Obwohl die von SHRDLU bewohnte Mikrowelt nur sehr einfach ausgestattet ist und auch seine kommunikativen Fähigkeiten recht rudimentär sind, steht das Programm doch für eine erste Vorstellung davon, über welche internen Mechanismen sprachliche Fertigkeiten realisiert werden können. Dabei ist die grundlegende Architektur von SHRDLU eng an die Repräsentationale Theorie des Geistes angelehnt und steht deshalb auch paradigmatisch für einen *kognitivistischen* Ansatz zur Modellierung intelligenten Verhaltens.

Im Kern ist SHRDLU ein symbolverarbeitendes Programm. Die grundlegenden Symbole sind hier Wörter, z.B. „CLEARTOP“, die kombiniert komplexere symbolische Strukturen ergeben, nämlich Sätze. Wenn SHRDLU einen (syntaktisch und semantisch korrekten) Auftrag erhält, so dient ihm diese Symbolstruktur als Zielvorgabe. Als Startwert liegt dem Programm der, ebenfalls als Symbolstruktur gegebene, Istzustand der Mikrowelt zur Verfügung. Auf algorithmische Art und Weise manipuliert SHRDLU nun den Startwert schrittweise so um, dass zum Schluss eine Symbolstruktur vorliegt, die mit der Zielvorgabe übereinstimmt. Wenn diese gelingt, liegt ein Handlungsplan vor, den er umsetzen kann.

Sprachverständnis wird hier somit als durch ein auf Regeln basierendes, physikalisches Symbolsystem realisiert angesehen. Auch der modulare Aufbau von SHRDLU verweist auf den charakteristischen *Top-Down-Ansatz* des Kognitivismus hin: Intelligentes Verhalten wird hier als durch das Zusammenwirken *funktionaler Subeinheiten* verstanden, wobei jede dieser Einheiten ganz spezifische *informationsverarbeitende* Prozesse erledigt:

SHRDLU illustrates a view of linguistic understanding as resulting from the interaction of many, independently specifiable cognitive processes. Each cognitive process does a particular job – the job of identifying noun phrases, for example. We make sense of the complex process of understanding a sentence by seeing how it is performed by the interaction of many simpler processes (or procedures). (Bermúdez, 2014, S. 39)

### 2.3.2 SHAKEY

Der Roboter SHAKEY wurde entwickelt, um in einer festgelegten, echten Umgebung, die sich aus miteinander verbundenen Räumen zusammensetzt, Kisten zu transportieren. Damit SHAKEY dies erfüllen kann, wurde er mit einem inneren Modell seiner Umwelt ausgestattet, das ihm auch seinen eigenen Zustand repräsentiert und bei Bedarf aktualisiert werden kann. Das Modell besteht dabei aus einer Menge von Aussagen einer einstufigen Prädikatenlogik. In ihr enthalten sind Ausdrücke für Objektamen sowie Ausdrücke für SHAKEY selbst und dessen Zustände.

Des Weiteren besitzt SHAKEY ein Programm zur Planung und Ausführung von Handlungen, das ebenfalls in der Prädikatenlogik formuliert ist und über ein Vokabular zur Bezeichnung von Handlungen verfügt.

Insgesamt sind die von SHAKEY beherrschten Verhaltensweisen hierarchisch organisiert. An unterster Stufe stehen sogenannte *elementare Verhaltensweisen* und aus ihnen sind alle weiteren, komplexeren Handlungen kompositional zusammengesetzt. Zur Initiierung einer elementaren Handlung greift das Programm auf die Hardwareebene des Roboters zu und aktiviert damit sensorische oder motorische Einheiten. Dadurch kann SHAKEY sich in Bewegung setzen oder seine Sensoren neu ausrichten. Komplexere Handlungen setzen sich aus einer Sequenz elementarer Handlungen zusammen und können als *Handlungsroutinen* vorgestellt werden. Dabei können auch Handlungsroutinen selbst Teil einer Handlungsroutine sein.

Eine Stufe über den Handlungsroutinen befindet sich jene Ebene, die Zielanweisungen in eine *Sequenz von Handlungsroutinen* zu übersetzen hat. Eine solche Zielanweisung wäre etwa der Auftrag, eine bestimmte Kiste in jenen Raum zu transportieren, in dem SHAKEY sich gerade befindet. Dafür muss der Auftrag mit der aktuellen Position von Roboters und Box abgeglichen werden, wofür das Umweltmodell herangezogen wird. Der Istzustand von SHAKEY und seiner Umwelt, dargestellt als formallogische Aussage, dient hier dann zur Erstellung einer *Sequenz logisch gültiger Schlüsse*, die als letzten logischen Schluss ein Umweltmodell generieren soll, die der Zielvorgabe entspricht. Im Kern sucht SHAKEY dabei einen *Entscheidungsbaum möglicher Handlungsroutinen* ab, bei der jeder Zweig einen möglichen Kurs beschreibt.

Darüber hinaus existiert auf der höchsten Ebene des Programmes ein *Monitoringsystem*, das die Ausführung des Handlungsplans überwacht. Wenn sich beispielsweise die Box nicht dort befindet, wo SHAKEY sie erwartet hätte (weil bspw. das Umweltmodell nicht oder falsch aktualisiert wurde), so weist dieses System den Roboter an, einen neuen Handlungsplan zu erstellen.

Im Ergebnis liefert SHAKEY damit ein verkörpertes Modell eines kognitiven Systems. Die Verwendung der Prädikatenlogik, verschiedener funktionaler Subsysteme und algorithmische Prozesse zur Ausführung von Handlungen verweist auch hier klar auf die Leitideen des Kognitivismus.

## 2.4 Zusammenfassung

Das Nachdenken über die menschliche Fähigkeit zu logisch vernünftigen Schlussfolgerungen nahm ihren Ausgangspunkt in den Arbeiten der klassischen Logik, wurde durch Frege erheblich erweitert und fand in den Arbeiten von Alan Turing einen bahnbrechenden Höhepunkt. Das mathematische Modell einer Turingmaschine zeigte die Möglichkeit einer vollständigen Mechanisierung logisch-korrektur Denkprozesse auf und fand in der Entwicklung der ersten Computersysteme ihre materielle Realisierung.

Zu etwa derselben Zeit führten Entwicklungen innerhalb der Psychologie und Linguistik zu einem stärkeren Fokus auf die Rolle innerer und angeborener Mechanismen des Gehirns zur Hervorbringung intelligenten Verhaltens.

Aus der KI-Forschung heraus schließlich wurde die Hypothese über physikalische Symbolsysteme formuliert, einer These, die Intelligenz im Kern als durch informationsverarbeitende Prozesse hervorgebracht ansieht.

Dieses informationsverarbeitende Paradigma verfestigte sich zum regulativen Leitgedanken der traditionellen Kognitionswissenschaft. Mit dem Funktionalismus, insbesondere der Repräsentationalen Theorie des Geistes wurde eine zugrundeliegende philosophische Theorie formuliert, die die Entwicklungen der Kognitionswissenschaft auch in das Gebiet der Philosophie des Geistes hineintrug.

Die Repräsentationale Theorie des Geistes liefert dabei auch eine erste Vorstellung über die allgemeine Struktur kognitiver Systeme. Eng mit der HPPS verwandt, wird hier das Gehirn als eine Art universale Turingmaschine vorgestellt, die auf Basis einer Sprache des Mentalen algorithmische Berechnungen über symbolische Strukturen durchführt und dabei intelligentes Verhalten erzeugt.

Methodisch betrachtet wurde damit ein Top-Down-Ansatz innerhalb des Kognitivismus etabliert: Ausgehend von den 3 Ebenen von Marr war es Aufgabe der Psychologie zunächst auf personaler Ebene kognitive Phänomene zu identifizieren und anschließend auf der algorithmischen Ebene in ihre funktionalen, informationsverarbeitenden Subeinheiten zu zerlegen. Die dort angewendete Komponentenanalyse soll dabei das Wesen von Intelligenz freilegen, wodurch der algorithmischen Ebene aus Sicht des Kognitivismus der eigentlich zentrale Status zukommt.

Demgegenüber wird der Ebene der Implementation eine bloß nebensächliche Rolle zugewiesen. Die genaue materielle Umsetzung informationsverarbeitender Prozesse wurde eher als technische Herausforderung, nicht aber als Intelligenzforschung im eigentlichen Sinne verstanden.

Mit der in Kapitel 2 vorgestellten Theorie des Konnektionismus etablierte sich Mitte der 1980er Jahre eine neue, konkurrierende Sichtweise im Hinblick auf Kognition. Inspiriert von der netzwerkartigen Struktur biologischer Gehirne, versuchen konnektionistische Theorien alternative Modelle kognitiver Prozesse aufzustellen. Ein zentraler Gedanke ist dabei, dass insbesondere die Ebene der Implementation auch Auswirkung auf die genaue Ausgestaltung informationsverarbeitender Prozesse auf der algorithmischen Ebene besitzt.

# Kapitel 3

## Kontextualisierung II: Konnektionismus

Tatsächlich spielten in den formierenden Jahren der Kognitionswissenschaft die Neurowissenschaften eine verhältnismäßig nebensächliche Rolle (Bermúdez, 2014, S. 59). Allen voran hatte dies damit zu tun, dass der Kognitivismus intelligentes Verhalten hauptsächlich auf die algorithmische Ebene rückzuführen versuchte und dementsprechend Fragen der Implementierung in den Hintergrund rückte:

Almost all cognitive scientists are convinced that in some fundamental sense the mind just is the brain, so that everything that happens in the mind is happening in the brain [...] But for a long time in the history of cognitive science it was widely held that we are better off studying the mind by abstracting away from the details of what is going on in the brain.

(Bermúdez, 2014, S. 6)

Dies allerdings änderte sich schlagartig mit dem Aufkommen neuer bildgebender Verfahren zur Visualisierung neuronaler Aktivitäten des Gehirns.

Es mag aus heutiger Sicht verwunderlich wirken, dass der konkreten neuronalen Organisation des Gehirns nicht von Beginn an eine wichtige Bedeutung zur Klärung menschlicher Intelligenz beigemessen wurde: Es scheint offensichtlich zu sein, dass die Ausarbeitung eines konkreten Algorithmus nicht bloß die auf personaler Ebene identifizierte

kognitive Leistung adäquat darzustellen hat, sondern auch die Spezifika neuronaler Prozesse berücksichtigen muss. Neuronale Signalübertragungen bspw. sind um ein Vielfaches langsamer als in elektronischen Transistoren (Clark, 2014, S. 16). Einerseits ging vom Computerbild des Geistes eine starke suggestive Kraft aus; andererseits gab es in der Anfangszeit der Kognitionswissenschaften schlicht und einfach keine technischen Möglichkeiten zur Untersuchung im Gehirn ablaufender Vorgänge.

Die durch die neu entwickelten, bildgebenden Verfahren zur Verfügung gestellten Daten über die Funktionsweise des Gehirns rückten dann aber schon bald Fragen der konkreten Implementation in den Fokus (Garnham, 2009, S. 107). Sowohl auf Ebene einzelner Neuronen, als auch auf Ebene größerer neuronaler Systeme konnte das Wissen um das Gehirn dank dieser Techniken erheblich gesteigert werden (Bermúdez, 2014, S. 93).

Allerdings kam die Neurowissenschaft dann auch zu der Erkenntnis, dass der Schlüssel zum Verständnis neuronaler Prozesse auf der Ebene von *Populationen* von Neuronen liegt, und genau für diese Ebene existierten keinerlei Verfahren. Populationen von Neuronen lagen gewissermaßen genau zwischen der Ebene einzelner Neuronen und der Ebene größerer neuronaler Systeme, und genau hier eben fehlten technische Methoden zu deren Messung (Bermúdez, 2014, S. 93).

Aus dieser Tatsache heraus gründete sich der Konnektionismus, der Populationen von Neuronen in Form künstlicher neuronaler Netzwerke mathematisch nachzubilden versuchte, um zumindest auf diesem Wege Forschung auf dieser Ebene betreiben zu können.

### 3.1 Grundaufbau

Ihrem Grundaufbau nach sind alle künstlichen neuronalen Netzwerke gleich. Ihr *schichtweiser Aufbau* setzt sich aus einer Eingangsschicht, einer oder mehreren verborgenen Schichten und einer Ausgangsschicht zusammen. Jede dieser Schichten besteht ihrerseits aus *Knotenpunkten*, die die Grundelemente eines jeden Netzwerks bilden und mit den Knotenpunkten davor- und nachliegender Schichten verbunden sind. Innerhalb einer Schicht sind die Knotenpunkte nicht miteinander verbunden.

Die Eingangsschicht fungiert dabei als Tor zur Außenwelt, empfängt dementsprechend also ein externes Eingangssignal. Dieses wird von den Knoten der Eingangsschicht emp-

fangen, verarbeitet und an die Knotenpunkte der ihr nachfolgenden (verborgenen) Schicht weitergeleitet. Dort verarbeiten die Knoten das Signal weiter und senden es anschließend wiederum an die nächste Schicht. Dieser Prozess wiederholt sich so lange, bis das Signal schließlich in der Ausgangsschicht ankommt, dort ein letztes Mal verarbeitet wird und dann als Ausgangssignal das Netzwerk verlässt.

Ein konnektionistisches Netzwerk kann damit als eine, aus einer Menge von Knotenpunkten bestehende, Struktur aus Verbindungen begriffen werden (Buckner & Garson, 2019). Es ist seiner Natur nach ein mathematisch-algorithmisches Computermodell mit einer *netzwerkartigen Struktur* (Besold & Kühnberger, 2013, S. 164). Diese Netzwerkstruktur konnektionistischer Modelle ist dabei bewusst der neuronalen Architektur des Gehirns nachempfunden.

Die in den Knotenpunkten stattfindende Signalverarbeitung ist dementsprechend an der Funktionsweise eines Neurons angelehnt. Neben dem Zellkern besitzt ein Neuron *Dendriten*, die Signale anderer Neuronen empfangen können und ein *Axon*, das Signale an andere Neuronen weiterleitet (Besold & Kühnberger, 2013, S. 164). Die *Verbindungen* die in konnektionistischen Netzwerken zwischen den Knotenpunkten der unterschiedlichen Schichten existieren, übernehmen die Rolle der Dendriten und des Axons. Im Gehirn bestimmt die Stärke des von einem Neuron ausgesendeten Signals über die Stärke der von ihm empfangenen Signale, was als *Aktivitätsniveau* eines Neurons bezeichnet wird. Die Verbindungen, die zwischen den Knotenpunkten im künstlichen Netzwerk existieren, besitzen *numerische Werte*, die angeben, wie stark das Signal von einem Knotenpunkt zum nächsten gesendet wird. Während positive numerische Werte signalverstärkend wirken, sorgen negative Werte für eine Signalhemmung (Buckner & Garson, 2019). Die Gesamtheit der von einem Knotenpunkt empfangenen Signale wird im Knotenpunkt anhand einer (mathematischen) Aktivitätsfunktion in jenes Signal umtransformiert, das der Knotenpunkt an die mit ihm verbundenen Knotenpunkte der nächsten Schicht weiterleitet.

Obwohl damit, also im Zuge der Modellierung künstlicher neuronaler Netzwerke, durchaus auf Eigenschaften echter Gehirne Rücksicht genommen wurde, muss nichtsdestotrotz hervorgehoben werden, dass künstliche neuronale Netzwerke die Funktionsweise echter Gehirne bloß recht grob approximieren:

[m]ost neural network research abstracts away from many interesting and possibly important features of the brain. For example, connectionists usually do not attempt to explicitly model the variety of different kinds of brain neurons, nor the effects of neurotransmitters and hormones.

(Buckner & Garson, 2019)

Einerseits hat dies damit zu tun, dass Computermodelle immer auch den berechnungstechnischen Grenzen der sie ausführenden Computer unterworfen sind; andererseits sind viele Details über die Funktionsweise des Gehirns schlichtweg noch nicht hinreichend verstanden und können deshalb auch nicht modelliert werden (Sun, 2014, S. 109).

Innerhalb des konnektionistischen Ansatzes gibt es eine große Bandbreite an verschiedenen *Netzwerkarchitekturen*. Die vorangegangenen Beschreibungen charakterisierten sogenannte *Feedforwardnetzwerke*. Diese propagieren das Eingangssignal linear von einer Schicht zur nächsten, bis schließlich die Ausgangsschicht erreicht ist, und ein Ausgabesignal kreiert wurde. Demgegenüber verfügen *rekurrente* neuronale Netzwerke über *Rückkoppelungsstrukturen*; sie senden Teilsignale zurück an vorhergehende Schichten und ermöglichen damit komplexere Signalverarbeitungen (Besold & Kühnberger, 2013, S. 166). Rekurrente Netzwerke sind Turing-Äquivalent und können damit sämtliche, auf Turingmaschinen implementierbare, Algorithmen ausführen (Sun, 2014, S. 111).

In einem konnektionistischen Netzwerk kommt den bereits erwähnten *Verbindungsgewichten* zwischen den Knotenpunkten eine besondere Rolle zu: Sie legen fest, wie ein ins Netzwerk eingespeistes Signal umtransformiert und ausgespeist wird. Neben der verwendeten Netzwerkarchitektur legen damit die numerischen Gewichtungen der Verbindungen die Identität eines Netzwerks fest (Buckner & Garson, 2019).

## 3.2 Lernen

Es ist des Weiteren auch möglich, Verbindungsgewichte graduell so abzuändern, dass sie in Bezug auf eine vorgegebene Aufgabenstellung stetig bessere Ergebnisse liefern. Dies geschieht mithilfe von *Lernalgorithmen*, wobei hier grob zwischen *überwachten* und

*nicht-überwachten* Lernalgorithmen unterschieden werden kann. Bei überwachten Lernalgorithmen wird das Ausgabesignal eines Netzwerks mit einem festgelegten Zielsignal abgeglichen. Das als Maßstab dienende Zielsignal repräsentiert dabei eine optimale Lösung für eine Berechnung. Je kleiner die Differenz zwischen Ausgabe- und Zielsignal ist, desto besser arbeitet das Netzwerk. Als Vertreter eines überwachten Lernalgorithmus spielt der *Backpropagationsalgorithmus* die ermittelte Differenz als Feedbacksignal in das Netzwerk zurück und versucht dabei die Verbindungsgewichte zwischen den Knotenpunkten so abzuändern, dass das Netzwerk bei zukünftigen Berechnungen näher an das optimale Ergebnis herankommt (Sun, 2014, S. 111). Künstliche neuronale Netzwerke sind dabei also in der Lage, aus vorangegangenen „Erfahrungen“ zu lernen, und das Ergebnis dieses Lernprozesses manifestiert sich in den abgeänderten Gewichtungen der Verbindungen: „Connection weights implicitly contain the record of past activation and so collectively embody what the network has learned from experience“ (Figdor, 2019, S. 287). Unbeaufsichtigte Lernalgorithmen statten demgegenüber das Netzwerk mit sich selbstorganisierenden Eigenschaften aus.

Verdeutlicht kann dies anhand des am *Prinzip des Hebb'schen Lernens* ausgerichteten Lernalgorithmus werden:

Wiederholt gemeinsam aktive Zellen oder Zellsysteme weisen eine Tendenz auf, sich enger miteinander zu verknüpfen, was nachfolgend zu einer Unterstützung der Aktivität eines Partners durch den jeweils anderen führt. Diese Lernregel lässt sich unmittelbar auf künstliche neuronale Netze übertragen, indem man sie dazu verwendet, die Veränderung der Gewichtungen von interneuronalen Verbindungen zu steuern. So wird das Verbindungsgewicht zwischen zwei zeitgleich und gleichartig aktiven und miteinander verschalteten künstlichen Neuronen erhöht, d.h. zukünftige Signale, die über diese Verbindung laufen, werden verstärkt, wohingegen die Gewichte zwischen unterschiedliche oder asynchron aktivierten Neuronen verringert werden.

(Besold & Kühnberger, 2013, S. 164)

Zwei wichtige Kennzeichen konnektionistischer Netzwerke ergeben sich aus der *parallel ablaufenden Verarbeitung von Informationen* und der *verteilten Natur von Reprä-*

*sentationen* (Besold & Kühnberger, 2013, S. 168). Künstliche neuronale Netzwerke führen in jeder Schicht mehrere Signalverarbeitungen gleichzeitig durch, was sie erheblich von klassischen Computerprogrammen unterscheidet, die demgegenüber Informationen sequentiell, d.h. schrittweise, verarbeiten. Auch werden Berechnungen im Netzwerk nicht auf Basis von Manipulationen diskreter Symbole vorgenommen. Repräsentationen im Netzwerk ergeben sich aus einer Vielzahl gleichzeitig aktivierter Knotenpunkte. Das Netzwerk bildet dabei semantisch ähnliche Einheiten auch syntaktisch ähnlich ab und zwar über den teilweise überschneidenden Gebrauch verteilter Repräsentationen (Clark, 2014, S. 73). Ein roter und ein gelber Tisch bspw. werden in einem Netzwerk durch ein ähnliches und teilweise überschneidendes Aktivitätsmuster repräsentiert. Im Ergebnis führt dies einerseits zu einem sehr sparsamen und effizienten Umgang mit den internen Ressourcen eines Netzwerks und andererseits ist es eine Art „[...] trick for forcing still more information into a system of encodings by exploiting even more highly structured syntactic vehicles than words“ (Clark, 2014, S. 73). Künstlichen neuronalen Netzwerken ist es damit möglich, semantische Eigenschaften auf *subsymbolischer* Ebene darstellen zu können, wodurch ihnen ein semantisch hohes Auflösungsvermögen eigen ist.

### 3.3 Charakteristika

Parallelität in der Informationsverarbeitung und die verteilte Natur von Repräsentationen führen in konnektionistischen Netzwerken zu folgenden Charakteristika: Einerseits verfügen solche Netzwerke über die Fähigkeit zu *spontanen Verallgemeinerungen* (*spontaneous generalizations*), „that is, the ability to spontaneously apply existing knowledge to different situations“ (Sun, 2014, S. 110). Für das Netzwerk neuartige Eingangssignale führen, so diese eine gewisse Ähnlichkeit zu bereits davor verarbeiteten Signalen aufweisen, zu einer diese Ähnlichkeit berücksichtigenden Reaktion. Dies ergibt sich daraus, dass zueinander ähnliche Signale auch ähnliche Aktivierungsmuster im Netzwerk hervorbringen (Clark, 2014, S. 73). Des Weiteren führen Beschädigungen im Netzwerks zu einem *bloß allmählichen Abbau* (*graceful degradation*) ihrer Funktionen (Sun, 2014, S. 110). „This is possible because the overall system now acts as a kind of pattern completer—given a large enough fragment of a familiar pattern, it will recall the whole thing“ (Clark,

2014, 73f.).

### 3.4 Kognitivismus vs. Konnektionismus

Insgesamt teilt der Konnektionismus die Sichtweise des Kognitivismus, wonach kognitive Prozesse ihrem Wesen nach informationsverarbeitende Prozesse über Repräsentationen sind. Über die genaue Auslegung dieser recht allgemeinen Überzeugung allerdings weichen konnektionistische und kognitivistische Ansätze teils erheblich voneinander ab und führten zu einer ersten grundsätzlichen Debatte innerhalb der Kognitionswissenschaft.

[c]ognitive science has employed a fundamental hypothesis, that thinking is produced by computational procedures operating on mental representations. However, there has been much controversy about what kind of representations and what kind of procedures are best suited to explain the many varieties of human thinking.

(Thagard, 2012, S. 52)

Worin liegen nun genau die Unterschiede zwischen diesen beiden Theorien? Kapitel 1 hat versucht, deutlich zu machen, dass die Hypothese über physikalische Symbolsysteme insgesamt bzw. die Repräsentationale Theorie des Geistes im speziellen, den menschlichen Geist in enger Analogie zu Computern betrachtet. Computer wiederum sind im Kern materialisierte formallogische Systeme. Formallogische Systeme manipulieren Symbolstrukturen anhand (eingeschriebener) syntaktischer Regeln und konservieren dabei auch den semantischen Gehalt dieser Strukturen. Gemäß des Kognitivismus ergibt sich intelligentes Verhalten also aus der regelgeleiteten Verarbeitung von Repräsentationen.

SHRDLU und SHAKEY sind zwei paradigmatische Beispiele für die Modellierung kognitiver Systeme im Sinne des Kognitivismus. Aufgabenstellungen werden von diesen Systemen rein formallogisch gelöst. Dafür besitzen sie eine eingeschriebene Wissensdatenbank, die sich aus einem Grundvokabular, Regeln zur Ableitung von Symbolstrukturen und einem inneren Weltmodell zusammensetzt. Werden SHRDLU oder SHAKEY

Aufträge erteilt, so setzen sie diese als Zielvorgabe und durchsuchen anhand eines Algorithmus einen abstrakten Lösungsraum in sequentiellen Schritten ab, der im Ergebnis eine Liste von Handlungen zur Zielerreichung liefern soll. Solche Systeme lernen aus eigenen Erfahrungen, indem sie entweder bereits existierende Regeln verallgemeinern oder aus existierenden Regeln neue Regeln ableiten (Thagard, 2012, S. 57).

Konnektionistische Modelle verfügen demgegenüber weder über eine interne Wissensdatenbank, noch über ein internes Modell eines Lösungsraums. Über große Datenmengen antrainierte, und mittels eines Lernalgorithmus feinjustierte Netzwerke, sind ihrem Wesen nach nichts mehr als „[...] a whole pattern of connections involving the input, output, and hidden layers“ (Thagard, 2012, S. 53). Dieses Muster beschreibt „a statistical connection between inputs and outputs that is often hard to characterize in words and is rarely replaceable by general if-then rules“ (Thagard, 2012, S. 53). Das Wissen eines Netzwerks ist also gesamtheitlich in den Verbindungsgewichtungen der Knotenpunkte abgespeichert, und Fähigkeiten zur Lösung von Problemen ergeben sich holistisch aus der Signalverarbeitung. „Komplexes Verhalten und mentale Phänomene werden hierbei als emergente Prozesse innerhalb von Netzwerken miteinander in Verbindung stehender, simpler Teilelemente verstanden“ (Besold & Kühnberger, 2013, S. 164). Wissen ist demzufolge also nicht über (sprachähnliche) diskrete Symbole realisiert, sondern wird vielmehr über verteilte Repräsentationen in das Netzwerk eingeschrieben. Lernen geschieht nicht wie bei regelbasierten Architekturen über explizit existierenden Regeln, sondern über „procedures that adjust the weights on the links between nodes“ (Thagard, 2012, S. 58).

Methodisch gesehen steht der Konnektionismus damit für einen *Bottom-Up-Ansatz* und nicht einen den Kognitivismus auszeichnenden *Top-Down-Ansatz*. Während kognitivistische Ansätze den Ausgangspunkt ihrer Forschungen in der Identifikation kognitiver Leistungen auf personaler Ebene sehen, und anschließend versuchen, einen diese kognitive Leistungen hervorbringenden Algorithmus zu entwickeln, setzen konnektionistische Untersuchungen bei der vom biologischen Gehirn inspirierten Konstruktion künstlicher neuronaler Netzwerke an. Diese Netzwerke werden in Bezug auf ein bestimmtes Aufgabengebiet mit Trainingsdaten gefüttert, wodurch sich mittels eines Lernalgorithmus die gewichteten Verbindungen zwischen den Knotenpunkten abändern und im Er-

gebnis zu einer strukturellen Veränderung des Netzwerks mit gesteigerten Fähigkeiten zur Aufgabenbewältigung führen.

### 3.5 Kritik

An dieser Stelle sollen nun zwei Einwände vorgestellt werden, die Vertreter symbolischer Architekturen konnektionistischen Theorien vorhalten.

Die erste Kritik wirft dem Konnektionismus vor, nicht erklären zu können, weshalb alltagspsychologische Überzeugungen *kausal wirksam* sind. Wird etwa eine Person dabei beobachtet, bei einsetzendem Regen ihr Schrittempo zu erhöhen, so erklären sich die ihn beobachtenden Menschen das i.d.R. dadurch, dass die Person (i) den Regen wahrgenommen hat, (ii) die Überzeugung gebildet hat, dass es regnet, die Person (iii) nicht nass werden möchte und (iv) aus diesem Grund das Schrittempo erhöht.

[c]ommonsense understanding of mind involves a crucial commitment to [...] ‚propositional modularity‘. This is the claim that the folk use of propositional attitude talk [...] implies a commitment to ‚functionally discrete, semantically interpretable states that play a causal role in the production of other propositional attitudes and ultimately in the production of behavior‘.

(Clark, 2014, S. 81)

Menschen erklären sich das Verhalten ihrer Mitmenschen indem sie diesen individuelle Überzeugungen unterstellen, die für ihr Verhalten handlungsleitend, d.h. kausal wirksam, sind. Die Repräsentationale Theorie des Geistes (RTG) liefert eine schlüssige Erklärung für dieses Phänomen. Die RTG behauptet ja gerade, dass solche individuellen Überzeugungen tatsächlich nicht nur auf personaler Ebene, sondern auch auf sub-personaler Ebene vorliegen, realisiert durch eine neuronal kodierte Sprache des Geistes, die als diskrete symbolische Repräsentationen zu sehen sind. Anhand algorithmischer Berechnungen setzen sie Handlungen. Die kausale Wirksamkeit alltagspsychologischer Erklärungen ergibt sich aus Sicht der RTG letztlich daraus, dass diese Erklärungen auf

zugrundeliegende neuronale Mechanismen des Gehirns verweisen. Der Konnektionismus kann demgegenüber die kausale Wirksamkeit alltagspsychologischer Erklärungen nicht verständlich machen, was ihn als unplausibel erscheinen lässt.<sup>1</sup>

Der Philosoph Andy Clark, dessen kognitionswissenschaftliche Theorie in Kapitel 5 eingehend vorgestellt wird, gilt als gemäßigter Vertreter des Konnektionismus und begegnet der Kritik mit dem Zugeständnis, dass (zumindest oberflächlich betrachtet) die RTG in Bezug auf die kausale Wirksamkeit alltagspsychologischer Erklärungen einen besseren Rahmen zu dessen Verständnis anbieten kann, als der Konnektionismus.

Konnektionistische Netzwerke verfügen über keine atomistischen Symbole, die auf Ebene des Gehirns eine mentale Sprache realisieren können. Eingabesignale in ein Netzwerk sind, mathematisch gesehen, Vektoren. Aufgabe dieses Netzwerkes ist es gewissermaßen anhand der Verbindungsgewichte eine Abbildungsmatrix zu erstellen, die diesen Inputvektor in einen adäquaten Outputvektor übersetzt (Clark, 2014, S. 82). Das Wissen eines Netzwerkes scheint holistisch bzw. gesamtheitlich in dieser Abbildungsmatrix gespeichert zu sein. Die Handlungen, die der vom Netzwerk erzeugte Outputvektor in Gang setzt, lassen sich deshalb auch nicht so einfach auf fein abgrenzbare innere Zustände des Netzwerkes rückführen. Repräsentationen künstlicher neuronaler Netzwerke sind ihrer Natur nach verteilte Repräsentationen, die sich auf viele verschiedene Verbindungsgewichte aufteilen. Dabei ist nicht offensichtlich, in welcher Beziehung diese über das gesamte Netzwerk verteilten Repräsentationen zu bestimmten, vom Netzwerk initiierten, Verhaltensweisen stehen. Dadurch scheint der Konnektionismus keinen sinnvollen Zusammenhang zwischen alltagspsychologischen Erklärungen auf personaler Ebene

---

<sup>1</sup>Es sei hier nur am Rande erwähnt, dass auf diese Herausforderung auf unterschiedliche Weise reagiert werden könnte. Die beiden Neurophilosophen Paul und Patricia Churchland, die konnektionistische Netzwerke ursprünglich in die Philosophie des Geistes eingebracht haben (Bickle, 2014, S. 237), stehen der Alltagspsychologie insgesamt recht kritisch gegenüber und meinen, dass es so etwas wie geistige Überzeugungen tatsächlich gar nicht gibt. Mit dieser als eliminativen Materialismus bezeichneten Theorie reagieren sie auf oben beschriebenen Einwand also dahingehend, als sie diesen Einwand schon allein deshalb als nicht stichfest erachten, weil die Vorstellung bzgl. der Existenz alltagspsychologischer Entitäten auf personaler Ebene falsch ist (Bickle, 2014, S. 238).

und neuronalen Mechanismen auf subpersonaler Ebene herstellen zu können.

Nun ist es allerdings so, dass zur Analyse konnektionistischer Modelle bestimmte mathematische Techniken entwickelt wurden. Sie erlauben im Kern eine „geometric analysis of connectionist-style representational systems“ (Clark, 2014, S. 83), die dazu in der Lage ist, Gemeinsamkeiten auf höherer Ebene, die zwischen dem Ausgabesignal des Netzwerks und der Signalverarbeitung besteht, abzubilden (Clark, 2014, S. 83). Mithilfe dieser statistischen Methoden ist es möglich, sogenannte *konzeptionelle Einheiten* zutage zu fördern. Diese sind mathematisch betrachtet ein Vektorpunkt in einem mehrdimensionalen Raum. Jeder Punkt in diesem Raum repräsentiert einen bestimmten Zustand, in dem sich das künstliche neuronale Netzwerk befinden kann. Damit ist es möglich, einem bestimmten Ausgangssignal eines Netzwerkes einen bestimmten Netzwerkzustand zuzuordnen. Dieser als mehrdimensionaler Vektor dargestellte Zustand eines Netzwerkes liefert die konzeptionelle Basis zur Erklärung eines bestimmten Netzwerkverhaltens. Andy Clark schlägt vor, diese konzeptionellen Einheiten als *verstreute Ursachen* (scattered causes) zu betrachten. *Verstreut* ist eine (als mehrdimensionaler Vektor verstandene) konzeptionelle Einheit deshalb, weil jeder Eintrag des Vektors einen anderen Aspekt des Netzwerks repräsentiert. Als *Ursache* für das Verhalten des Netzwerkes kann die konzeptionelle Einheit deshalb angesehen werden, weil sie einen statistischen Zusammenhang zu einem konkreten Ausgabesignal herstellt. Konnektionistische Netzwerke können also mittels mathematischer Analysen Ursachen für das Verhalten von Netzwerken angeben, die allerdings nicht wie beim Kognitivismus diskreter Natur sind, sondern über das Netzwerk verteilt, in den Verbindungsgewichtungen liegen. Damit ist es auch dem Konnektionismus möglich, das Verhältnis zwischen personaler und subpersonaler Ebene bestimmen zu können, wenn auch nicht auf solch direkte Art und Weise wie dies etwa die Repräsentationalistische Theorie des Geistes kann. Während die RTG individuelle (handlungsanleitende) Überzeugungen durch diskrete neuronale Strukturen realisiert sieht, und damit zwischen personaler und subpersonaler Ebene eine Art semantische Transparenz schafft, stellt der Konnektionismus diese Verbindung mittels eines subsymbolischen Ansatzes her: Die Ursache für eine handlungsanleitende Überzeugung liegt verstreut im als Netzwerk vorgestelltem Gehirn vor. Eine geometrische Analyse der Informationsverarbeitung kann verständlich machen, welche Teile des Netzwerkes in welcher

Art und Weise ein bestimmtes Verhalten realisieren. Damit etabliert der Konnektionismus eine *subsymbolische* Beschreibungsebene, weil hier nicht ein diskretes Symbol für ein bestimmtes Konzept, Objekt oder Ereignis einsteht, sondern mentale Entitäten aus mehreren Teilaspekten des Netzwerkes zusammen konstituiert werden. Diese werden durch einen mehrdimensionalen Vektor repräsentiert, der eine engmaschigere Beschreibungsebene bzw. ein höheres Auflösungsvermögen als diskrete Symbole aufweist.

Clark bezeichnet aus diesem Grund den Konnektionismus auch als einen *Mikrofunktionalismus* (Clark, 2014, S. 40). Er ist deshalb ein *Funktionalismus*, weil eine konnektionistische Sichtweise im Sinne Clarks kognitive Zustände als durch kausale Rollen realisiert sieht, nämlich durch verstreute Ursachen; und er ist deshalb ein *Mikrofunktionalismus*, weil verstreute Ursachen subsymbolische kausale Ursachen sind, d.h. auf einer feinmaschigeren Erklärungsebene operieren.

Neben der kausalen Wirksamkeit alltagspsychologischer Erklärungen konfrontiert der Kognitivismus konnektionistische Theorien oftmals mit dem sogenannten Problem der Systematizität. Eine plausible Theorie über menschliche Intelligenz muss diesem Problem zufolge erklären können, weshalb menschliches Denken allem Anschein nach recht systematisch strukturiert ist:

An ability to entertain one proposition is correlated with an ability to think other propositions. For example, someone who can entertain the thought that John loves Mary can also entertain the thought that Mary loves John.

(Rescorla, 2017)

Menschliche Gedanken scheinen demzufolge *kompositional* zusammengesetzt zu sein, und eine Theorie menschlichen Denkens muss verständlich machen können, weshalb dies so ist. Die RTG löst dieses Problem, indem es Gedanken als durch eine darunter liegende Sprache des Geistes realisiert sieht. Menschliches Denken ist allen voran sprachliches Denken. Die Sprache weist ein hohes Maß an *Systematizität* auf, und weil neuronale Prozesse des Gehirns bei Denkvorgängen ebenfalls über eine (mentale) Sprache arbeiten, geht die menschliche Fähigkeit kompositionale Gedanken zu bilden direkt daraus hervor.

Eine aus Sicht des Konnektionismus entwickelte und für den weiteren Verlauf der Arbeit noch wichtig werdende Replik von Andy Clark, stellt die Frage, ob menschliches Denken für sich genommen tatsächlich so systematisch abläuft, wie es das Problem der Systematizität voraussetzt (Clark, 2014, S. 86). Vielleicht, so Clark, erscheint menschliches Denken nur deshalb so systematisch, weil Denken sich (zumeist) sprachlich vollzieht, Denken somit strukturelle Eigenschaften von Sprache widerspiegelt und Sprache selbst, aufgrund ihres grammatikalischen Aufbaus, ein hohes Maß an Systematizität aufweist. „[H]uman thought may actually inherit such systemacity as it actually displays from the grammatical structure of human language itself“ (Clark, 2014, S. 85). Diesem Gedanken zufolge wäre also nicht das menschliche Denken selbst kompositional strukturiert, sondern hat sich durch die Einbettung in ein sprachliches Umfeld dahingehend kognitiv umtransformiert.

The presence of a public code in which real chunky external symbols are indeed widely recombinable and highly manipulable adds whole new dimensions to basic biological cognition [...] Perhaps [...] it is these new (and relatively recent) dimensions that give human thought the appearance of such deep systematicity. We possess a new tool—language—that sculpts and guides our thought in new ways.

(Clark, 2014, S. 86)

Sollte Clark Recht haben, dass die Systematizität menschlicher Denkprozesse eine Anpassung an ein sprachliches Umfeld darstellt, dann stellt das Problem der Systematizität keine Herausforderung für konnektionistische Theorien dar. Aus dieser Sichtweise wäre es möglich, dass jene neuronalen Aktivitäten, die menschliche Denkprozesse realisieren, nicht-systematischer Natur sind und allein aufgrund eines, von einem externen Umfeld geschaffenen, sprachlichen Rahmens den Anschein von Systematizität erwecken. Die Theorie künstlicher neuronaler Netzwerke und das Problem der Systematizität wären, in diesem Lichte betrachtet, also gut miteinander vereinbar.

### 3.6 Zusammenfassung

Insgesamt legt der Konnektionismus damit eine alternative Theorie über eine kognitive Architektur vor, die zwar am informationsverarbeitenden Paradigma des Kognitivismus festhält, nichtsdestotrotz aber eine völlig neue Interpretation über die Art und Weise der Informationsverarbeitung und der dabei verwendeten Repräsentationen liefert.

Konnektionistische Modelle besitzen heutzutage sowohl einen fixen Platz in der KI-Forschung (Deep Learning) als auch innerhalb der Kognitionswissenschaft (Sun, 2014, S. 108). Insbesondere in Bereichen wie der Gesichts- und Stimmerkennung, beim Spracherwerb oder zur Modellierung sensomotorischer Fähigkeiten konnten sie teilweise recht beachtliche Erfolge liefern.

Einen Schwachpunkt stellen allerdings höhere kognitive Fähigkeiten, wie deduktives Schlussfolgern, Planen und typische Aufgaben sequentiellen Denkens, dar, die der Mensch ohne Zweifel ebenfalls besitzt, aber nicht gut von neuronalen Netzwerken angelernt werden können. Hier erzielen regelbasierte Systeme weitaus bessere Ergebnisse.

Es wird deshalb mehrheitlich davon ausgegangen, dass zur Modellierung menschlicher Intelligenz hybride Architekturen entwickelt werden müssen, die sowohl symbolischer, als auch konnektionistischer Natur sind.

The current battle between rule-based and connectionist architecture is analogous to a previous episode in the history of science, the controversy between wave and particle theories of light. From the seventeenth through the nineteenth centuries, there was oscillation between the wave theory [...] and the particle theory [...] The battle was only settled in the twentieth century by the advent of quantum theories of light, according to which light consists of photons that exhibit properties of both particles and waves. Similarly, I think that the most reasonable conclusion from the current impasse of rule-based and connectionist architectures is that the mind is both a rule-based and a connectionist system, and that problem solving can sometimes be search through a space of rules and sometimes parallel constraint satisfaction.

(Thagard, 2012, S. 59)

Im nächsten Kapitel wird mit der *Theorie situierter Kognition* eine Sichtweise auf Intelligenz vorgestellt, die sich sowohl von kognitivistischen als auch konnektionistischen Ideen erheblich unterscheidet.

Während der Konnektionismus gegenüber dem Kognitivismus insbesondere die Rolle der Ebene der Implementation hervorhebt, will die situierte Kognition deutlich machen, dass eine ausschließlich am Gehirn fokussierte Untersuchung zur Erklärung kognitiver Prozesse nicht ausreichend ist.

In gewisser Hinsicht wird hier auf die Bedeutung der personalen Ebene verwiesen und dort insbesondere auf die *körperlichen* Eigenschaften kognitiver Agenten sowie deren *Einbettung* in eine Umwelt betont. Körperliche und umweltliche Aspekte dürfen zur Beurteilung kognitiver Phänomene nicht ausgeklammert werden, da sie ebenfalls einen wesentlichen Beitrag für intelligentes Verhalten liefern.

Damit verbindet sich auch eine Sichtweise, die sich vom informationsverarbeitenden Paradigma, dem sich bei aller Verschiedenheit sowohl Kognitivismus als auch Konnektionismus verschrieben haben, abwendet und stattdessen auf die Wichtigkeit *dynamischer* Prozesse zur Schaffung intelligenten Verhaltens hinweist.

# Kapitel 4

## Kontextualisierung III: Situierete Kognition

Mit der *Theorie über situierte Kognition* (TSK) hat sich in den 1990er Jahren eine weitere Strömung innerhalb der Kognitionswissenschaft etabliert. Diese allerdings bietet einen weniger klar umrissenen theoretischen Rahmen, wie etwa der Kognitivismus oder der Konnektionismus, an. Das ist insbesondere darauf zurückzuführen, dass es sich bei der situierten Kognition eher um einen Oberbegriff handelt, der viele unterschiedliche Sichtweisen in Bezug auf Kognition unter sich versammelt, die zwar in einigen grundlegenden Überzeugungen miteinander übereinstimmen, nichtsdestotrotz aber auch recht spezifische Ausformulierungen und Schlussfolgerungen dieser gemeinsamen Annahmen entwickeln. Situierete Kognition sollte man sich deshalb am ehesten als „[...] loose-knit family of approaches to understanding the mind and cognition“ (Wilson & Clark, 2009, S. 55) vorstellen.

Abhängig von der jeweils betrachteten Theorie, will die situierte Kognition entweder eine *Alternative*, mindestens aber eine erhebliche *Modifikation* zu klassischen Vorstellungen über das Wesen von Intelligenz anbieten. Unter Schlagworten wie *Verkörperung*, *Einbettung* oder *Situiertheit* will sie für eine *weitende* Perspektive auf kognitive Phänomene werben, die nicht allein auf das Gehirn und die dort stattfindenden neuronalen Prozesse blickt, sondern auch körperliche und umweltliche Aspekte miteinbezieht: „In a

nutshell, the basic idea is that we should think of situated cognition as a form of cognitive extension, or rather as a variety of forms that such cognitive extension can take“ (Wilson & Clark, 2009, S. 55), d.h. „[e]mbodiment and situatedness are taken to broaden, literally, the boundaries of the cognitive system“ (Shapiro, 2011, S. 125).

In Abgrenzung zu dieser Sichtweise sehen sie insbesondere das traditionelle Computerbild von Kognition, dass sie sowohl im Hinblick auf dessen Internalismus, als auch auf dessen Zentrierung auf das Gehirn kritisieren (Newen, Bruin & Gallagher, 2018, S. 4). Diese *individualistische Konzeption des Geistes* klammert zur Charakterisierung kognitiver Zustände und Strukturen die gesamte übrige Welt (inklusive des Körpers) aus (Wilson & Clark, 2009, S. 56), weil sie von einer ganz bestimmten Vorstellung angeleitet ist, die in Kapitel 1 bereits ausführlich und hier nochmals kurz umrissen wird:

[c]ognition takes place inside the head, wedged between perception (on the input side) and action (on the output side) [...] It involves the computational processing of mental representations, where these are language-like both in their constituents (concepts as words) and in their structural composition (mental representations as generated by, and decomposable in terms of, and underlying mental syntax). Such processing relies purely on formal or syntactic features of the symbols themselves and the rules by which they are manipulated, rather than on broader features of the individuals who instantiate them and the environments in which they operate.

(Wilson & Clark, 2009, S. 56)

Bereits der Konnektionismus hat, wie im vorherigen Kapitel gezeigt wurde, in vielerlei Hinsicht mit diesen Vorstellungen gebrochen, indem er die Bedeutung der Ebene der Implementation hervorgehoben hat und dabei Zweifel an der rein abstrakten, algorithmischen Beschreibungsebene, die die traditionelle Kognitionswissenschaft so hervorhebt, infrage stellt. Kognition, so der Einwand, findet nicht in einer *substratneutralen* Sphäre statt, sondern ist ihrer Funktionsweise nach eng an materielle Eigenschaften und der spezifischen Organisation des Gehirns gebunden.

Allerdings bleibt letztlich auch der Konnektionismus allein auf das Gehirn zentriert, und analysiert dieses in einem gewissermaßen körper- und weltlosen Vakuum: „As in typical computers, connectionist processing begins with settings to input nodes and ends with adjustments to output nodes. Understanding the activity in a connectionist net requires understanding features internal to the net [...]“ (Shapiro, 2011, 125f.).

Die situierte Kognition will alldem entgegenwirken und begreiflich machen, dass sämtliche kognitive Phänomene „essentially rely on the system’s body and its dynamical and reciprocal real-time interaction with its environment“ (Newen et al., 2018, S. 5). Kognitive Prozesse „[...] are in some sense all dependent on the morphological, biological, and physiological details of an agent’s body, an appropriately structured natural, technological, or social environment, and the agent’s active and embodied interaction with this environment“ (Newen et al., 2018, S. 5).

Im Ganzen betrachtet, ist die situierte Kognition dementsprechend von der Grundannahme geleitet, „dass sowohl die kognitiven als auch die geistigen Zustände und Prozesse von Lebewesen – insbesondere auch in Bezug auf den Menschen – intrinsisch verkörpert und als solche wesentlich in eine Umwelt eingebettet sind“ (Fingerhut, Hufendiek & Wild, 2013, S. 9). Wesentlich ist dahingehend die „[...] Idee, dass der intelligente Zugriff auf die Welt überhaupt – sei es durch Verhalten, Wahrnehmung oder begriffliches Denken – in viel höherem Maße auf körperlichen Fähigkeiten und der Beschaffenheit des Körpers und der Umwelt beruht, als zumeist angenommen wird“ (Fingerhut et al., 2013, S. 12).

Als wesentliche Inspirationsquelle dienen der situierten Kognition dabei die phänomenologischen Untersuchungen von Husserl, Heidegger und Merlau-Ponty (Fingerhut et al., 2013, S. 25–32). Sie alle verbindet die Vorstellung, dass der Mensch als solches nur unter Miteinbeziehung seines Leibes und seiner Umwelt verstanden werden kann. „Von der Phänomenologie gingen also die entscheidenden Anstöße für das verkörperte und eingebettete Bild des Geistes aus“ (Fingerhut et al., 2013, S. 32).

Die situierte Kognition kann demzufolge mit einer *holistischen*, d.h. ganzheitlichen, Sichtweise in Bezug auf die Entstehung von Intelligenz identifiziert werden. Selbst für den Fall, dass dem Gehirn eine kognitive Sonderstellung zukommen sollte — und selbst dies wird von manchen Ansätzen in Zweifel gezogen —, so ist es doch so, dass das Gehirn, seiner Entwicklungsgeschichte nach, aus einfachen, körperlichen Organismen her-

aus entstanden ist, die stets in eine spezifische Umwelt eingebettet waren. Ihrer Funktion nach entwickelten sich kognitive Architekturen dabei stets aus den Bedürfnissen und umweltlichen Herausforderungen eines *Gesamtorganismus* heraus.

Damit etablierte sich mit der situierten Kognition auch eine stärker an der Biologie und Ökologie ausgerichtete Betrachtungsweise von Kognition.

## 4.1 Zwei Merkmale situiertes Kognition

Zwei Merkmalen kommt dabei eine besonders zentrale Bedeutung zu: Erstens wurden kognitive Prozesse immer mehr aus einer *handlungszentrierten* Perspektive heraus untersucht, aus der sich ein radikal anderes Verständnis in Bezug auf Repräsentationen heraus entwickelte. Es wird davon ausgegangen, dass (i) die Existenz innerer Repräsentationen zur Erklärung bestimmter Verhaltensweisen überhaupt nicht angenommen werden muss und, (ii) im Falle der Existenz solcher Repräsentationen, diese ganz anders als bisher vorzustellen sind. In Bezug auf (i) wird argumentiert, dass Wahrnehmungen eines Organismus oftmals direkt, d.h. ohne einen vermittelnden Denkprozess, auf Handlungen wirken. Intelligentes Verhalten muss sich demzufolge also nicht über ein *sequentielles Stufenmodell der Informationsverarbeitung* ergeben, worunter zu verstehen ist, dass Prozesse der Wahrnehmung, des Planens und des Handelns strikt voneinander getrennt auftreten. Wahrnehmungsinhalte können demgegenüber auch direkt auf Handlungen wirken, was als *geschlossene sensomotorische Schleife* (auch: geschlossene Regelschleife) bezeichnet wird (Clark, 2014, S. 97). Selbst wenn Denkprozesse zwischengeschaltet sein sollten, so erscheint das sequentielle Stufenmodell der Informationsverarbeitung als problematisch. Das betrifft nun (ii), und zwar insbesondere mit Blick auf die RTG.

Perception is [...] continuously influenced by cognitive, contextual, and motor factors. It need not yield a rich, detailed, and action-neutral inner model awaiting the services of ‚central cognition‘ so as to deduce appropriate actions [...] In a certain sense, the brain is revealed not as (primarily) an engine of reason or quiet deliberation, but as an organ of environmentally situated control.

(Clark, 2014, 105f.)

Eine sequentielle Anschauung ist also deshalb problematisch, weil bereits auf die Inhalte der *Wahrnehmung* kognitive, kontextuelle und handlungsspezifische Faktoren wirken. Ebenfalls scheint es, in Bezug auf die vom Gehirn verwendeten Repräsentationen fragwürdig zu sein, diese als *objektiv* anzusehen: Die Fähigkeit zur Wahrnehmung eines Organismus besitzt nämlich nicht (notwendigerweise) die Funktion, dem Gehirn die Erstellung eines detailreichen und wahrheitsgemäßen inneren Weltmodells zu ermöglichen. Vielmehr ergibt sich aus einer handlungszentrierten Perspektive die Vorstellung, dass innere Repräsentationen ganz auf die *Handlungsmöglichkeiten* eines Organismus hin ausgerichtet sind. Sie sind demnach, weil handlungszentriert, auch *agentenzentriert*. Im Kern bedeutet dies, dass innere Repräsentationen die Aufgabe besitzen, allein jene Aspekte der Umwelt abzubilden, die für ein jeweiliges Lebewesen von Bedeutung sind. Was für einen jeweiligen Organismus von Bedeutung ist, hängt dabei sowohl von seiner spezifischen Umwelt, als auch von seinen körperlichen Eigenschaften ab.

Aus der handlungszentrierten Perspektive folgt demnach als zweites charakteristisches Merkmal situierter Kognition, nämlich ein Fokus auf die kognitive Bedeutung *körperlicher und umweltlicher Aspekte* im Hinblick auf einen spezifischen Organismus.

## 4.2 Methode

Aus den gerade eben erwähnten zwei Merkmalen ergibt sich dann allerdings nicht bloß eine neue Sichtweise auf Kognition, sondern auch ein anderer *methodischer* Zugang.

Nicht nur ist es so, dass innere Prozesse des Gehirns nun eine Funktion der körperlichen Beschaffenheit eines Organismus und dessen spezifischer Umwelt sind; auch folgen jene evolutionsbiologischen Mechanismen, aus denen heraus sich Lebewesen entwickelt haben, ganz anderen Prinzipien, als jenen, von denen sich Menschen bei der Konstruktion von Maschinen haben leiten lassen:

Biological evolution [...] is both constrained and liberated in ways we are not. It is constrained to build its solutions incrementally on top of simpler

but successful ancestral forms [...] More positively, biological evolution is liberated by being able to discover efficient but ‚messy‘ or unobvious solutions that may, for example, exploit environmental interactions and feedback loops so complex that they would quickly baffle a human engineer. Natural solutions [...] can exploit just about any mixture of neural, bodily, and environmental resources along with their complex, looping, and often nonlinear interactions. Biological evolution is thus able to explore a very different solution space (wider in some dimensions, narrower in others) than that which beckons to conscious human reason.

(Clark, 2014, S. 95)

Damit allerdings verliert das Computerbild des Geistes einiges an Plausibilität. Die Vorstellung, intelligentes Verhalten biologischer Organismen analog zur Architektur von Computersystemen modellieren zu können, wird sich nämlich eben nicht mit dem Lösungspfad decken, den die Evolution zur Entwicklung von Intelligenz eingeschlagen hat. Computer können, und zwar genau deshalb, weil Menschen sie entwickelt haben, fein säuberlich in ihre funktionalen Einzelbestandteile zerlegt und auf Basis der drei Ebenen von Marr hin analysiert werden (Clark, 2014, S. 109). Biologische Intelligenz allerdings ergibt sich aus einer engen Wechselwirkung zwischen Organisationsweise des Gehirns, körperlichen Eigenschaften des Organismus und umweltlichen Aspekten. Dies macht es, bereits auf der Rechenebene von Marr, in der die kognitiven Leistungen auf personaler Ebene identifiziert werden sollen, notwendig, körperliche und umweltliche Spezifika eines Organismus miteinzubeziehen. Theorien situierter Kognition stellen damit den seitens des Kognitivismus gegebenen, privilegierten Status der algorithmischen Ebene zum Verständnis von Kognition fundamental in Frage:

[r]estricting ourselves to the level of familiar kinds of algorithmic specification is a poor strategy if we hope to understand the way biological brains guide intelligent action. Instead, we must pay close and constant attention to the nature and properties of neural circuits and to the complex interactions among brain, body, and environment. The strategy of focusing attention on

a kind of disembodied ‚pure software‘ level, floating high above the messy material realm, works well when we confront computational systems of our own design [...] Understanding the intricate, unexpected, yet often stunningly efficient products of blind natural evolution calls for other techniques and ways of thinking.

(Clark, 2014, S. 111)

Anschaulich gemacht kann diese alternative Sichtweise auf Kognition anhand der sogenannten *Künstlichen-Leben-Forschung* (KL-Forschung) und der *Theorie Dynamischer Systeme* (TDS) werden. Beiden ist gemein, dass sie sich in enger Anbindung zu den Leitgedanken situiertes Kognition verstehen (Shapiro, 2011, S. 116).

KL-Forschung und Ansätze der TDS wählen dabei jeweils als Ausgangspunkt ihrer Betrachtungen relativ einfache kognitive Leistungen, etwa bestimmte motorische Fähigkeiten, bei denen dann allerdings versucht wird, sie ihrem vollen Umfang nach und biologisch plausibel umzusetzen (Bermúdez, 2014, S. 424).

### 4.3 Künstliches Leben

*Künstliches Leben* (artificial life) beschäftigt sich „[m]it der Nachbildung intelligenter biologischer Phänomene [...]“ (Rohde, 2013, S. 180). Ihre Kernthemen sind dabei adaptive Verhaltensweisen, Selbstorganisation und das Verhalten in der geschlossenen Regelschleife (closed sensorimotor loop), wobei von letzterer dann gesprochen wird, „[...] wenn ein System die Auswirkungen seines Verhaltens in der Umwelt unmittelbar registrieren und in Echtzeit darauf reagieren kann“ (Rohde, 2013, S. 180). Der KL-Forschung geht es demnach zuvorderst um die Konstruktion solcher KI-Systeme und Roboter, die eigenständige Lösungen auf umweltliche Problemstellungen entwickeln können und flexibel auf unerwartete Situationen reagieren können. Im Gegensatz zu klassischen KI-Systemen besitzen sie deshalb auch kein vollständig handlungsfestlegendes inneres Programm (Bermúdez, 2014, S. 424).

Als große Inspirationsquelle dienen ihr dabei biologische Organismen, die im Laufe ihrer Evolutionsgeschichte auf vielfältige Weise Mechanismen zur Lösung umweltlicher

Herausforderungen entwickelt haben. Insbesondere einfache Lebewesen, die aufgrund ihrer relativ primitiven Nervensysteme nur über recht eingeschränkte Kapazitäten zur Informationsverarbeitung verfügen, werden in den Mittelpunkt der Aufmerksamkeit gerückt.

Entwickelt hat sich dieser Forschungszweig aus der Tatsache heraus, dass klassische Ansätze der KI zunehmend an ihre Grenzen stießen: Obwohl sie gute Ergebnisse bei der Lösung logisch-komplexer Aufgabenstellungen (etwa beim Schach oder Turm von Hanoi) liefern konnten, bereitete ihnen bspw. die Echtzeitkontrolle von Robotern in veränderlichen Umgebungen erhebliche Schwierigkeiten (Rohde, 2013, S. 180).

Die KL-Forschung nahm sich dieses Problems an und fokussierte ihr Interesse auf einfache Lebewesen, die ebenfalls mit dem Problem der *Echtzeitkontrolle* konfrontiert sind. Bereits Insekten müssen sich schnell und souverän in ihrer Umwelt fortbewegen, um auf Feinde und Beute zeitgerecht und adäquat reagieren zu können. Zur Bewältigung dieser Herausforderungen können solch einfache Organismen allerdings nicht auf komplexe informationsverarbeitende Prozesse zurückgreifen, d.h. höheren kognitiven Fähigkeiten kann hierbei keinerlei Bedeutung zukommen. Vielmehr muss auf biologisch einfachen Stufen die Realisierung adaptiver, intelligenter Verhaltensweisen auf körperliche Eigenschaften und einfache geschlossene sensomotorische Schleifen rückzuführen sein (Rohde, 2013, S. 181).

Daraus ergeben sich für die KL-Forschung drei Themenbereiche von besonderem Interesse: Zum einen ein Interesse an der Konstruktion kognitiv bescheidener künstlicher Systeme, die relativ autonom und flexibel mit ihrer Umwelt interagieren können. Zweitens eine Auseinandersetzung mit den Beiträgen, die körperliche Merkmale, spezifische Handlungen und umweltlicher Kontext für adaptive Verhaltensweisen liefern können und schließlich eine besondere Aufmerksamkeit emergenten bzw. kollektiven Effekten gegenüber, die sich aus dem Zusammenwirken einfacherer System- und Umwelteigenschaften ergeben (Clark, 2014, S. 118).

Innerhalb der Robotik gilt der am Mobile Robot Laboratory des MIT wirkende Rodney Brooks als einer der Pioniere eines KL-Ansatzes (Clark, 1998, S. 12). Brooks befasst sich dort mit der Entwicklung *autonomer Roboter*, deren oberstes Ziel es ist, sich *flexibel* in einer realistischen und nicht (vollständig) vorhersehbaren Umwelt bewegen zu können

und dabei bestimmte Ziele zu verfolgen. In Anlehnung an echte biologische Organismen stellt Brooks dabei folgende vier Bedingungen an all seine Roboter: Erstens müssen sie dazu in der Lage sein, *adäquat* und *zeitgerecht* auf Veränderungen in der Umwelt reagieren können (Echtzeitkontrolle). Zweitens muss ihre Interaktion mit der Umwelt *robust* sein. Drittens sollen die Roboter *mehrere Ziele gleichzeitig* unterhalten können und viertens in der Welt *aktiv* sein, d.h. ihr Dasein soll auf einen bestimmten Zweck hin ausgerichtet sein.

Einer dieser von Brooks entwickelten Roboter ist Herbert. Er wurde entwickelt, um während des Laborbetriebs leere Getränkedosen einzusammeln, was alles andere als trivial war: „[...] the robot had to negotiate a cluttered and changing environment, avoid knocking things over, avoid bumping into people, and identify and collect the cans“ (Clark, 1998, S. 14).

Erreichen sollte er dies mithilfe einer sogenannten *Subsumptionsarchitektur*, die sich aus mehreren autonomen Schichten zusammensetzt: Jede dieser Schichten realisiert eine geschlossene sensomotorische Schleife zwischen wahrnehmenden Sensoren und motorischen Einheiten. „Each layer is [...] functionally equivalent to a simple whole robot, capable of performing just one task“ (Clark, 2014, S. 114). Innerhalb jeder Schicht wird, ausgehend von den Sensoren, ein einfaches Signal an die Motoreinheiten gesendet, d.h. es existieren keine internen Repräsentationen und damit auch keine informationsverarbeitenden Prozesse. Auch untereinander können die verschiedenen Schichten nur auf sehr simple Weise interagieren. Die Interaktion reduziert sich im Prinzip darauf, dass gewisse Eingangssignale eine Schicht dazu veranlassen, aktiv zu werden und dabei eine andere Schicht deaktivieren. Insofern arbeiten die Schichten relativ autonom und sind nur recht lose aneinander gekoppelt (Clark, 2014, S. 114).

Herbert verfügt über mehrere Schichten: Die unterste Schicht dient dazu, Herbert vor Auffahrunfällen zu schützen. Dafür sorgen Ultraschallsensoren, die, falls sie ein Hindernis registriert haben, Herbert direkt stoppen können, worauf der Roboter sich umorientiert. Die unterste Schicht aktiviert sich immer dann, wenn die Ultraschallsensoren ein relevantes Signal einspeisen, wodurch alle anderen Schichten deaktiviert werden. Immer dann, wenn die Ultraschallsensoren nichts registrieren ist die zweite Schicht aktiv, die Herbert zufällig im Raum herumbewegen lässt. Die dritte Schicht schaltet sich wieder-

um dann zu, wenn das einfache visuelle System von Herbert ein tischähnliches Objekt identifiziert, und dieses dann mit einem Laserstrahl und einer Kamera scannt. Eine andere Schicht aktiviert sich, wenn eine Getränkedose identifiziert wurde und Herbert sich so ausrichtet, dass die Dose zentral in seinem Blickfeld positioniert ist. Daraufhin fährt Herbert seinen mit primitiven Drucksensoren ausgestatteten Roboterarm aus und sammelt die Dose ein.

Die Subsumptionsarchitektur gewährleistet damit jene Bedingungen, die Brooks an seine Roboter stellt. Das Problem der Echtzeitkontrolle wird durch sensomotorische Regelschleifen realisiert, die auf keinerlei zeitintensive informationsverarbeitende Prozesse zurückgreifen müssen. Ebenfalls sorgen sie für ein robustes Verhalten gegenüber der Umwelt, da das gesamte Handlungsspektrum solcher Roboter sich aus der engen Kopplung mit umweltlichen Aspekten ergibt. Weil jede Schicht einer Subsumptionsarchitektur ein funktional vollständiges Subsystem etabliert, können mehrere Ziele gleichzeitig unterhalten werden. Der Daseinszweck solcher Roboter ergibt sich gewissermaßen aus dem emergenten Effekt, der durch das Zusammenwirken der verschiedenen Schichten erreicht wird. Des Weiteren ist das *inkrementelle* Organisationsprinzip der Subsumptionsarchitektur an den evolutionären Entwicklungsprozess eines biologischen Organismus angelehnt: Eine neu auf den Roboter gelegte Schicht ändert nicht die darunter liegenden Funktionsprinzipien des Roboters, wie auch neue Fähigkeiten bei Organismen nicht das gesamte Organisationsprinzip von diesem abändern (Bermúdez, 2014, S. 434).

Dieses inkrementelle Design und die in jeder Schicht vorhandenen direkten Input-Output-Schleifen sorgen dafür, dass solche Roboter auch keine zentrale Kontrollinstanz besitzen und auch nicht benötigen. „There is no clear dividing line between perception and cognition, no point at which perceptual inputs are translated into a central code to be shared by various onboard reasoning devices“ (Clark, 1998, S. 14). Stattdessen sorgt die enge und ununterbrochene Verbindung des Roboters mit seiner Umwelt für adaptive, intelligente Verhaltensweisen ohne informationsverarbeitende Prozesse (Shapiro, 2011, S. 141).

Wird zum Vergleich nochmals der in Kapitel 1 kurz vorgestellte Roboter SHAKY herangezogen, so ist ersichtlich, wie weit sich KL-Ansätze von klassischen KI-Ansätzen voneinander unterscheiden: Bei Herbert sind die zentralen Komponenten handlungser-

zeugende, eng an die Umwelt gekoppelte Subsysteme, die ohne innere Symbole miteinander kommunizieren. Es gibt kein sequentiell und zentralisiert arbeitendes Symbolsystem, das eine klare Trennung zwischen Wahrnehmung, Planung und Handlung vornimmt, wie das bei SHAKEY der Fall ist.

## 4.4 Theorie Dynamischer Systeme

Die *Theorie Dynamischer Systeme* (TDS) stellt ebenfalls eine alternative Betrachtungsweise kognitiver Phänomene dar und ist dabei eng an den sogenannten *Enaktivismus* angelehnt. Der Enaktivismus ist eine spezifische theoretische Ausrichtung innerhalb der situierten Kognition und betont insbesondere die Bedeutung *dynamischer Koppelungen*, die zwischen Gehirn, Körper und Umwelt bestehen (Newen et al., 2018, S. 3). Kognition wird als *emergentes* Phänomen vorgestellt, das sich aus der aktiven Auseinandersetzung eines Systems mit seiner Umwelt ergibt (Newen et al., 2018, S. 6). Kognition ist damit keine Eigenschaft, die aus den Aktivitäten biologischer Gehirne allein realisiert wird, sondern durch das wechselseitige Zusammenspiel von Umwelt, Körper und Gehirn.

Die TDS kann dabei als jene Disziplin betrachtet werden, die den methodischen Rahmen zur Modellierung einer solchen Sichtweise auf Kognition bereitstellt.

Um die grundlegenden Ideen der TDS verständlich machen zu können, werden nachfolgend einige wichtige Begriffe geklärt: Ganz allgemein wird unter einem *dynamischen System* jedes System verstanden, dass regelhaften, zeitlichen Entwicklungen unterworfen ist. „The solar system is a dynamical system. So are you and I. So is a dripping tap. And so, for that matter, are Turing machines and artificial neural networks“ (Bermúdez, 2014, S. 405). Um die zeitlichen Veränderungen dynamischer Systeme präzise beschreiben zu können, benutzt die TDS ein *System von Differentialgleichungen* (Shapiro, 2011, S. 116), in denen Beziehungen wichtiger, für das System relevanter Quantitäten, festgehalten sind (Bermúdez, 2014, S. 405). Diese Gleichungen legen auch den sogenannten *Zustandsraum* eines Systems fest, der angibt, wohin ein dynamisches System sich potentiell entwickeln kann. Ein *bestimmter* Zustand des Systems ergibt sich dabei, indem jeder in den Gleichungen enthaltenen Größen ein fester Wert zugeordnet wird. Jede Gleichung für sich genommen beschreibt einen Teilaspekt des Gesamtsystems, auch *Komponente* genannt.

Sogenannte *gekoppelte* Systeme treten dann auf, wenn die Entwicklung von zumindest einer Komponente des Gesamtsystems abhängig von der Entwicklung einer anderen Komponente ist. Diese wechselseitigen Abhängigkeiten mehrerer Komponenten voneinander wird unter dem Begriff der *Koppelung* gefasst (Clark, 2014, S. 147). Mit ihnen ist es möglich, Prozesse darzustellen, die auf mehrere Subsysteme einen *kontinuierlich-zirkulären kausalen Einfluss* (continuous circular causal influence) besitzen (Clark, 2014, S. 141).

Im Hinblick auf kognitive Prozesse postuliert die TDS, dass diese aus einem dynamischen System heraus hervorgehen. Dieses System setzt sich aus Gehirn, Körper und Umwelt zusammen, wobei insbesondere auch materiellen Aspekten des Systems eine bedeutende Rolle zuerkannt wird.

Instead of abstracting away from the physical details of how cognitive systems actually work, it suggests that those physical details can play all sorts of unsuspected but vitally important roles in determining how a cognitive system changes and evolves over time. Dynamical systems theory offers a new way of analyzing and predicting cognitive systems.

(Bermúdez, 2014, S. 420)

Die TDS kann damit als ein kognitionswissenschaftlicher Ansatz gesehen werden, der zunächst einmal den Anspruch erhebt, kognitivistischen und konnektionistischen Theorien gegenüber einen besseren Rahmen zur *Beschreibung* kognitiver Phänomene zur Verfügung zu stellen (Clark, 2014, S. 141). Informationsverarbeitende Prozesse (des Gehirns) finden hier keine Berücksichtigung; was zählt, sind die interaktiven (gekoppelten) Wechselwirkungen, die zwischen den Komponenten des Systems bestehen.

Hier sollen nun einige wichtige Charakteristika dynamischer Systeme vorgestellt werden. Neben der bereits erwähnten möglichen Existenz von Koppelungen, weisen sämtliche Dynamische Systeme die Eigenschaft von *Zeitsensitivität* auf. Damit ist gemeint, dass zeitliche Faktoren oftmals einen wichtigen Beitrag zur Erklärung im System ablaufender Prozesse liefern. Auch das unterscheidet sie stark von informationsverarbeitenden Ansätzen:

Computational descriptions have a beginning, a middle, and an end. An algorithm is essentially a recipe that makes explicit the steps involved in transforming an input into an output. However, apart from the ordering of the sequence of steps, timing is irrelevant. This is why a modern computer and a slowly plodding Turing machine may be said to be running the same algorithm despite the fact that the former might produce a solution millions of times faster than the latter.

(Shapiro, 2011, S. 122)

Ein weiterer erwähnenswerter Aspekt ist, dass die Grenzen dynamischer Systeme *beobachtungsabhängig* sind: Was ein System konstituiert, ist wesentlich von der gewählten Fragestellung bzw. dem Erkenntnisobjekt abhängig (Clancey, 2009, S. 14).

Nun allerdings begnügt sich die TDS nicht allein damit, dass der von ihr zur Verfügung gestellte methodische Rahmen eine bessere Beschreibung kognitiver Phänomene liefert (Clark, 2014, S. 151); auch erhebt die TDS den Anspruch, mit ihren Modellen eine bessere *Erklärung* für das Wesen von Kognition liefern zu können:

Dynamical systems theorist think that the explanations work because they track the basic design principles of the mind. They think not only that the mind is a dynamical system, but also that when we look at the relation between the organism and the environment what we see is a coupled system. The organism–environment complex is a system whose behavior evolves as a function of a small number of variables.

(Bermúdez, 2014, 411f.)

Kurzum wird seitens der TDS behauptet, dass das Bild von Kognition in Form eines dynamischen Systems eine bessere Vorstellung über dessen Natur liefert, als das Modell der Turingmaschine (Shapiro, 2011, S. 141).

Dafür sprechen aus Sicht der TDS folgende Tatsachen: Kognitive Prozesse sind zeit-sensitive Prozesse und eng an körperliche und umweltliche Aspekte gekoppelt, zwischen

denen wechselseitige kausale Beziehungen bestehen. Kognition kann deshalb nicht (ausschließlich) anhand informationsverarbeitender Prozesse verständlich gemacht werden. Kognition ist ein aus den Aktivitäten eines dynamischen Systems hervorgehendes, emergentes Phänomen (Shapiro, 2011, S. 126), zu dessen Entstehung keine algorithmischen Verarbeitungen über symbolhafte Strukturen notwendig sind.

In seiner radikalsten Form richtet sich die Theorie Dynamischer Systeme vollständig gegen das informationsverarbeitende Paradigma des Kognitivismus und Konnektionismus. Mit der *These über eine radikale Verkörperung von Kognition*, wonach „[s]tructured, symbolic, representational, and computational views of cognition are mistaken“ (Clark, 2014, S. 151), soll insbesondere die ablehnende Haltung gegenüber der Annahme der Existenz innerer symbolhafter Strukturen zum Ausdruck gebracht werden. Wenn nämlich Repräsentationen für konkrete Objekte eintreten, kognitive Agenten allerdings in interaktiven und kontinuierlichen Austausch zu diesen Objekten selbst stehen, dann benötigen solche Agenten schlichtweg keine inneren Abbildungen dieser Objekte (Shapiro, 2011, S. 118).

Zur Veranschaulichung einer solchen Sichtweise auf Kognition kann das von Thelen und Smith untersuchte Gehverhalten von Kleinkindern herangezogen werden (Clark, 1998, S. 40):

Wird ein Säugling vom Boden gehoben, so führt es koordinierte Schrittbewegungen aus, die allerdings ab dem 2. Lebensmonat aussetzen und erst ca. 8-10 Monate nach der Geburt wieder auftreten. Zu diesem Zeitpunkt können die Beine das eigene Gewicht tragen. Ab Erreichung des ersten Lebensjahres schließlich werden diese Schrittbewegungen vom Kleinkind selbst initiiert und lassen so erste eigenständige Gehbewegungen entstehen. Wodurch ist dieses Beobachtungsmuster zu erklären? „At one time, it was thought that the best explanation of this overall pattern would depict it as the expression of a prior set of instructions, complete with timing, encoded in (perhaps) a genetically specified central pattern generator“ (Clark, 2014, S. 144). Thelen und Smith allerdings konnten anhand ihrer Untersuchungen aufzeigen, dass auch bei Kleinkindern zwischen dem zweiten und achten Lebensmonat solche Schrittbewegungen hervorgerufen werden konnten und zwar dann, wenn sie in warmes Wasser gehalten wurden. Dies konnte damit erklärt werden, dass ab dem 2. Lebensmonat das Gewicht der Beine eines Kleinkinds so stark

zunimmt, dass die wie eine Feder wirkende Beinmuskulatur keine Schrittbewegung herstellen kann. Wird das Kind in warmes Wasser gehalten, kehrt sich dieses Verhältnis allerdings um. Werden Gewichte an den Beinen von Kindern befestigt, die sich gerade in einer Gehphase befinden, so blieben auch dort die Schrittbewegungen aus.

Ebenfalls konnte der Einfluss von umweltlichen Faktoren aufgezeigt werden. Kinder im Alter von 7 Monaten zeigten auf einem motorisierten Laufband Gehbewegungen.

Insgesamt wird also das Schrittverhalten bei Kindern durch ein komplexes Zusammenspiel zwischen körperlichen, neuronalen (Schrittbewegungen tauchen erst ab dem 1. Lebensjahr unabhängig auf) und umweltlichen Aspekten herbeigeführt.

Das Bewegungsverhalten bei Kleinkindern ist damit eine dynamische Eigenschaft eines von Gehirn, Körper und Umwelt etablierten gekoppelten Systems und „[...] highlights the complex and subtle interplay between intrinsic dynamics, organic change, and external task environment“ (Clark, 2014, S. 145).

## 4.5 Zusammenfassung

Insgesamt etablierte sich mit Theorien situiertes Kognition eine stärkere Hinwendung auf *körperliche* und *umweltliche* Aspekte zur Klärung intelligenten Verhaltens.

Damit in engem Zusammenhang stehend, verlagerte sich der Fokus von inneren Mechanismen kognitiver Systeme hin zu den Interaktionen eines Agenten mit seiner Umwelt. Mit dieser *handlungszentrierten* Sichtweise eng verbunden kristallisierte sich auch eine große Skepsis innerer repräsentationaler Strukturen gegenüber. Aus einem engen dynamischen Wechselspiel zwischen Agenten und Umwelt erscheint die Annahme innerer Repräsentationen nicht notwendig zu sein was insbesondere in der Theorie Dynamischer Systeme ersichtlich ist, aber auch etwa anhand der Subsumptionsarchitektur von Brooks verständlich gemacht werden kann.

Methodisch gesehen setzte sich mit dieser, auf das gesamte System und seine Umwelt gerichteten Perspektive, ein *holistischerer* Zugang in der Beschreibung und Modellierung intelligenten Verhaltens durch.

Fraglich dabei allerdings ist, inwiefern Theorien situiertes Kognition auch im Hinblick auf anspruchsvollere Denkprozesse fruchtbar zur Anwendung gebracht werden können.

Der Philosoph und Kognitionswissenschaftler Andy Clark setzt mit seinem *dynamischen Computationalismus*, der im Kern eine Synthese kognitivistischer, konnektionistischer und situierter Überlegungen darstellt, genau dort an und versucht damit, das gesamte kognitive Spektrum menschlicher Intelligenz erklären zu können.

Damit nun wird sich das nächste Kapitel beschäftigen.

## Kapitel 5

# Der dynamische Komputationalismus von Andy Clark

Bevor nun im nächsten Kapitel die These über erweiterte Kognition (TEK) vorgestellt wird, soll in diesem Kapitel der allgemeine Denkraum von Andy Clark präsentiert werden.

Insbesondere wird hierbei auf die von ihm 1997 veröffentlichte Monographie *Being There. Putting Brain, Body, and World Together Again* Bezug genommen werden. Sie ist damit gut ein Jahr vor Erscheinen des Artikels *The Extended Mind* veröffentlicht worden, den er zusammen mit David J. Chalmers verfasst hat und durch die die TEK erstmalig einer breiteren wissenschaftlichen Öffentlichkeit nähergebracht wurde.

*Being There* ist dabei eine umfassende Gesamtschau der von Clark vertretenen Ansichten, und umfasst dabei so gut wie alle theoretischen Ansätze, die in den vorangegangenen Kapiteln vorgestellt wurden. Auf originäre Art und Weise greift Clark die verschiedenen Vorstellungen aus den Bereichen des Kognitivismus, Konnektionismus und situierter Kognition auf, unterzieht diese einer kritischen Bewertung und leitet daraus Schlussfolgerungen ab, die zusammengenommen ein ganz eigenständiges Bild über das Wesen kognitiver Phänomene liefern.

## 5.1 Die evolutionsbiologisch-ökologische Perspektive

Als guten Ausgangspunkt für Clarks Überlegungen dient die Feststellung, dass sich (kognitive) Herausforderungen immer aus der Perspektive des gesamten Organismus ergeben. Eine von der Umwelt gestellte Aufgabe kann immer nur mit Blick auf ein spezifisches Lebewesen und dessen Bedürfnissen verständlich gemacht werden. Aus dieser Tatsache heraus entsteht die Forderung nach einer *agentenzentrierten Sichtweise auf Kognition*. Problemstellungen ergeben sich für den Organismus insgesamt, und zwar aus Sicht des Organismus, selbst. Im Hinblick auf Lösungsmöglichkeiten ist es dabei entscheidend zu begreifen, dass Lebewesen stets *verkörpert* und in eine spezifische Umwelt *eingebettet* sind. Daraus ergibt sich, dass zur Bewältigung einer Aufgabe sowohl neuronale, körperliche<sup>1</sup>, als auch *auf die Umwelt bezogene* Ressourcen herangezogen werden können.

Es ist demnach nicht notwendig, sämtliche Problemlösefähigkeiten im Gehirn selbst als verankert anzunehmen. Arbeiten aus der KL-Forschung zeigen eindrücklich, wie Probleme der Echtzeitkontrolle, vor die alle Lebewesen (also auch der Mensch) gestellt werden, biologisch plausibel modelliert werden können, ohne dabei rechenintensive, informationsverarbeitende Prozesse zugrunde legen zu müssen. Der in Kapitel 4 vorgestellte Roboter Herbert, mit der dort implementierten Subsumptionsarchitektur, liefert dafür ein gutes Beispiel.

Daraus ergibt sich für Clark folgende erste Moral: „[a]void excessive world modeling, and [...] gear such modeling as is required to the demands of real-time, behavior-producing systems“ (Clark, 1998, S. 23). Im Kern steckt hier eine erste Kritik an klassischen KI-Ansätzen zur Modellierung kognitiver Phänomene, die eine zu starke Fixierung auf die algorithmische Ebene aufweist und dabei völlig außer Acht lässt, dass auch dezentrale, viel sparsamere Lösungen, existieren. Ein top-down-Ansatz wird Problemen der Echtzeitkontrolle schon deshalb nicht gerecht, weil diese bereits bei sehr einfachen Lebewesen auftreten, die sie mit sehr bescheidenen Ressourcen zu lösen haben. Besser

---

<sup>1</sup>Körperlich meint hier die *nicht-neuronalen* körperlichen Eigenschaften eines Lebewesens.

ist es hier den Ansätzen der KL-Forschung zu folgen, die ihren Ausgangspunkt bei der grundlegenden Anatomie einfacher Lebewesen nimmt, dort insbesondere auch materiellen Eigenschaften oder der genauen Platzierung der Wahrnehmungseinheiten Beachtung schenkt. Auch ist für einen solchen bottom-up-Ansatz entscheidend, die spezifische Umwelt eines Organismus unter die Lupe zu nehmen. Evolutionsbiologisch betrachtet, haben sich Organismen immer aus einer spezifischen ökologischen Nische heraus entwickelt, woraus sich für Clark eine zweite wichtige Moral für Kognition ergibt: „[f]ind very close fits between the needs and lifestyles of specific systems [...] and the kinds of information-bearing environmental structures to which they will respond“ (Clark, 1998, 23f.).

Als wichtige Inspirationsquellen gelten Clark hierbei die Schriften von Jakob von Uexküll, einem der Wegbereiter der Ökologie, der den Begriff der *Umwelt* erstmals in den biologischen Diskurs einbrachte. Uexküll erkannte, dass jeder Organismus auf eine für ihn besondere ökologische Nische hin ausgerichtet ist, die er als *effektive Umwelt* bezeichnet. Abhängig von dieser und den eigenen Bedürfnissen, entwickelt jedes Lebewesen eine *nischenabhängige Wahrnehmung*, d.h. es wird (insbesondere) das wahrgenommen, was für den Fortbestand eines Organismus von Relevanz ist. Daraus ergibt sich die wichtige Erkenntnis, dass evolutionsbiologische Prozesse zu hoch selektiven kognitiven Leistungen einer biologischen Gattung führen können: „[i]t can sensitize an organism to whatever (often simple) parameters reliably specify states of affairs that matter to the specific life form“ (Clark, 1998, S. 25).

In diesem Zusammenhang ist es auch wichtig festzustellen, dass die Wahrnehmung eines spezifischen Organismus nicht nur *nischenabhängig* ist, sondern auch *handlungszentriert*. Ökologische Herausforderungen ergeben sich innerhalb jener Nische, in die ein Organismus eingebettet ist; gelöst können diese immer nur vor dem Hintergrund der eigenen Handlungsmöglichkeiten eines Lebewesens. Eine agentenzentrierte Sichtweise auf Kognition schließt also sowohl ein nischenabhängiges, als auch ein handlungszentriertes Wahrnehmungsvermögen des Organismus mit ein.

Welche kognitiven Lösungsstrategien können vor diesem Hintergrund erwartet werden? Für Clark wäre es falsch zu denken, dass ein verkörperter, und in eine Umwelt eingebetteter Organismus, nicht auch körperliche und umweltliche Ressourcen zur Bewäl-

tigung seiner Aufgaben heranziehen würde. Dies hat Clark bereits in einem früheren Werk dazu veranlasst, das sogenannte *007-Prinzip* zu formulieren, dass in weiterer Folge als *Prinzip repräsentionaler Sparsamkeit* bezeichnet werden soll.

In general, evolved creatures will neither store nor process information in costly ways when they can use the structure of the environment and their operations upon it as a convenient stand-in for the information-processing operations concerned. That is, know only as much as you need to know to get the job done.

(Clark, 1998, S. 46)

Das Prinzip geht also davon aus, dass aus Sicht eines Organismus die interne Speicherung und Verarbeitung von Information sehr kostenintensiv ist, und deshalb evolutionäre Prozesse günstigeren Lösungen den Vorzug geben. Wenn also umweltliche Strukturen selbst zur Lösung eines Problems beitragen können bzw. mittels der durch den Organismus vorgenommenen Manipulation dieser Strukturen Probleme vereinfacht werden können, so wird dies in der Regel zur Generierung einer Lösung auch miteinbezogen werden.

In Bezug auf menschliche Intelligenz identifiziert Clark insgesamt drei kognitive Mechanismen, die diesem Prinzip unterliegen: (i) Die *Aufteilung der Last* einer umweltlichen Herausforderungen auf Umwelt, Körper und Gehirn; (ii) Ökologische Probleme werden mittels der *Selbststrukturierung von Informationen* rekonfiguriert und damit vereinfacht; (iii) ein Lebewesen erweitert die eigenen kognitiven Fähigkeiten mittels *bioexterner Ressourcen* (Clark, 2008, S. 196). Letztgenannter Mechanismus verweist hierbei auf die Theorie des erweiterten Geistes und wird im nächsten Kapitel eingehend dargestellt werden. Im nächsten Abschnitt werden demgegenüber die erstgenannten kognitiven Mechanismen genauer beschrieben.

### 5.1.1 Mechanismen intelligenten Verhaltens

Eine Aufteilung der Last kognitiver Herausforderungen ist für Clark immer dann gegeben, wenn die Mechanismen zu dessen Lösung auf nichttriviale Art und Weise kausal auf

Gehirn, Körper und Umwelt aufgeteilt wurde: „Nontrivial causal spread occurs whenever *something we might have expected to be achieved by a certain well-demarcated system turns out to involve the exploitation of more far-flung factors and forces*“ (Clark, 2008, S. 7).

Dies versucht Clark am Beispiel zweier Ansätze der Robotik zu veranschaulichen: Klassische KI-Ansätze nutzen zur Bewegungskoordination eines Roboters rechenintensive Kontrollstrukturen, die zentralisiert die verschiedenen Bewegungseinheiten eines Roboters ansteuern. Der von Honda entwickelte Roboter Asimo ist für Clark ein paradigmatischer Fall eines solchen Ansatzes. Nun ist es allerdings so, dass Roboter wie Asimo sich zwar durchaus effektiv in ihrer Umwelt fortbewegen können, die dafür aufgewendete Energie allerdings so hoch ist, dass Fortbewegungsmechanismen wie diese kein biologisch plausibles Modell liefern. Tatsächlich besitzen biologische Lebewesen einen um ein Vielfaches niedrigeren Energieaufwand, als dies etwa Asimo tut und erreichen dies, indem sie sich zahlreiche sogenannte *passive Dynamiken* zunutze machen (Clark, 2008, S. 5). Passive Dynamiken ergeben sich dabei aus der spezifischen Organisation eines Bewegungsapparates und daraus resultierenden Bewegungseigenschaften. Innerhalb der KL-Forschung werden Roboter entwickelt, die über solche passive Dynamiken verfügen und diese auch zur eigenen Fortbewegung verwenden können. Solchen Robotern werden etwa gekrümmte Füße mit einer nachgiebigen Ferse implementiert, den Beinen zur Nachahmung von Muskeln, Sehnen und Bändern mechanische Federn zur Verfügung gestellt oder die beiden Arme mechanisch miteinander verbunden (Clark, 2008, 5f.). Selbst materielle Eigenschaften oder die Platzierung von (wahrnehmenden) Sensoren werden hierbei berücksichtigt. Im Ergebnis besitzen solche Roboter sehr ausgefeilte und elegante Bewegungsabläufe und erreichen dies mit einem sehr niedrigen Energieeinsatz. Die Ausnutzung passiver Dynamiken sorgt dafür, dass viele herausfordernde Aspekte der Fortbewegung von diesen passiven Dynamiken selbst erledigt werden. Damit ist es nicht notwendig, auf ein rechenintensives internes Kontrollsystem zurückgreifen zu müssen, das sich all dieser Herausforderungen anzunehmen hat. Insofern sorgen solch passive Dynamiken für eine effiziente Aufteilung der kognitiven Last:

One of the big lessons of contemporary robotics is that the coevolution of morphology (which can include sensor placement, body plan, and even the

choice of basic building materials, etc.) and control yields a truly golden opportunity to spread the problem-solving load between brain, body, and world.

(Clark, 2008, S. 8)

Dass in diesem Zusammenhang auch die interne Signalverarbeitung eines Roboters schnell und kostengünstig erfolgen kann, zeigt etwa die in Kapitel 4 vorgestellte Subsumptionsarchitektur, die zur Steuerung des Roboters Herbert verwendet wurde.

Ökologische Lösungsstrategien greifen somit nicht zwangsläufig auf neuronale Prozesse zur Bewältigung von Aufgaben zurück. Ein Organismus besitzt auch körperliche und umweltliche Ressourcen zur Lösung seiner Probleme. Der Rückgriff auf solche führt oftmals zu evolutionär stabilen Verhaltensweisen. Zentrale Aufgabe des Gehirns ist es dann *nicht*, Mikromanagement zu betreiben, d.h. alle Details, die bei der Bewältigung einer Aufgabe anfallen, zentral zu koordinieren; vielmehr kommt ihm hierbei die Rolle eines ökologischen Kontrollsystems zu:

[a]n ecological control system is one in which goals are not achieved by micromanaging every detail of the desired action or response but by making the most of robust, reliable sources of relevant order in the bodily or worldly environment of the controller.

(Clark, 2008, 5f.)

Die *Selbststrukturierung von Information zur Rekonfiguration von Aufgabenstellungen* bewirkt demgegenüber eine Vereinfachung der ursprünglichen Aufgabenstellung. Die zentrale Idee ist hierbei, dass verkörperte Organismen direkt mit ihrer Umwelt interagieren können und so die eigenen Wahrnehmungsinhalte zusätzlich anreichern bzw. verändern können. Studien konnten etwa belegen, dass die Ausbildung von Konzepten und Kategorielernen bei Kleinkindern unterstützt wird, wenn diese direkt mit der Umwelt interagieren, d.h. Objekte angreifen, bewegen oder drehen. Solche Interaktionen erzeugen einen reichen Fluss zeitlich abgestimmter, multimodaler Wahrnehmungssignale.

Davon inspiriert wurde ein als BABYBOT bezeichneter Roboter entwickelt, der mittels eines Roboterarms Objekte seiner Umwelt manipulieren kann und dabei über die Grenzen, die zwischen einem Objekt und der übrigen Umwelt bestehen, Wissen ansammelt. Wie auch bei Kleinkindern beobachtet werden konnte, war es hier auch BABYBOT möglich, durch die Selbststrukturierung von Informationen neues Wissen über die Welt zu erhalten (Clark, 2008, S. 18). Im Ergebnis sieht Clark hier ein hybrides kognitives Phänomen, das sowohl dem informationsverarbeitenden Paradigma, als auch Theorien über situierte Kognition unterliegt:

Here, then is a nice example of an approach that combines a bedrock computational and information-processing perspective with a potent functional role for timing and environmentally coupled action [...] Such work depicts intelligent response as grounded in processes of information extraction, transformation, and use, while recognizing the key roles, *in those very processes*, played by timing, action, and coupled unfolding.

(Clark, 2008, S. 19)

Tatsächliche Handlungen in der Welt führen dabei also zu einer kognitiven Transformation der ursprünglichen Aufgabenstellung. Das Gehirn ist nicht passiv darauf angewiesen, aus einer statischen Position heraus Informationen aus der Umwelt zu erhalten, sondern kann durch die Initiierung von Handlungen die eigenen Wahrnehmungsinhalte aktiv anreichern bzw. vereinfachen. Solche von Clark als *Handlungsschleifen* (action-loops) bezeichnete Mechanismen finden bspw. auch bei Puzzlespielen Anwendung, indem etwa Puzzleteile vorsortiert, gedreht oder einfach ausprobiert werden: „[p]ure thought leads to actions which in turn change or simplify the problems confronting pure thought.“ This is probably the simplest kind of example of the phenomena known as *action loops*“ (Clark, 1998, S. 36).

In Bezug auf solche Handlungsschleifen, die hier die Rolle der Rekonfiguration einer kognitiven Aufgabenstellung besitzen, nimmt Clark nun auch eine prinzipielle Unterscheidung zwischen *pragmatischen* und *epistemischen Handlungen* vor:

Pragmatic action is action undertaken because of a need to alter the world to achieve some physical goal [...] Epistemic action, in contrast, is action whose primary purpose is to alter the nature of our own mental tasks. In such cases, we still act on the world, but the changes we impose are driven by our own computational and information-processing needs.

(Clark, 1998, S. 64)

Eine noch größere Bedeutung werden solche epistemischen Handlungen dann zur Stützung der These des erweiterten Geistes bekommen. Allerdings sei bereits an dieser Stelle vorweggenommen, dass Clark epistemische Handlungen auch als Merkmal der Existenz eines *physikalisch-informationalen Raumes* (physico-informational space) begreift: „The world [...] can function as much more than just external memory. It can provide an arena in which special classes of external operations systematically transform the problems posed to individual brains“ (Clark, 1998, S. 66). Im Kern würden physikalisch-informationale Räume den Schluss nahelegen, dass das informationsverarbeitende Paradigma der traditionellen Kognitionsforschung unter eine erweiternde Perspektive gebracht werden könnte. Informationsverarbeitung wird dann zu keiner rein internen Angelegenheit neuronaler Prozesse, sondern weitet sich auch auf nicht-neuronale Eigenschaften des Körpers und umweltliche Strukturen aus. Beide der in diesem Abschnitt vorgestellten kognitiven Mechanismen legen für Clark folgenden Schluss nah: So biologische Lebewesen als verkörperte und in eine ökologische Nische eingebettete Wesen ernst genommen werden, ergibt sich daraus eine völlig neue Sichtweise auf die Natur kognitiver Phänomene. Die Lösung ökologischer Herausforderungen befindet sich nicht allein in den neuronalen Strukturen des Gehirns, sondern ist ebenfalls auf körperliche und umweltliche Aspekte aufgeteilt, und auch die Handlungen des Organismus tragen zur Bewältigung kognitiver Aufgaben bei.

Dies alles bedeutet für Clark nicht, dass dem Gehirn auch weiterhin die zentrale Rolle im Hinblick auf kognitive Leistungen zuerkannt werden sollte; allerdings muss der spezifische Aufgabenbereich des menschlichen Gehirns zur Hervorbringung intelligenten Verhaltens etwas anders gedacht werden. Dies soll im nächsten Abschnitt genauer ausgeführt werden.

### 5.1.2 Die Rolle des Gehirns

Gemäß dem Prinzip repräsentionaler Sparsamkeit führt Clark ein, dass die Natur eine Vorliebe für sogenannte *Soft-Assembly-Lösungen* zu haben scheint. Darunter ist zu verstehen, dass sich Organismen zur Bewältigung ihrer Herausforderungen externe Strukturen und, aus Sicht des Gehirnes, dezentralisierte Lösungen anstreben (Clark, 1998, S. 45).

Neuronale Strukturen werden nicht für umfassende Informationsverarbeitungen verwendet, sondern sparsam eingesetzt. Das Gehirn fungiert eher als ökologisches Kontrollsystem zur Feinjustierung bzw. Modulation bestimmter Parameter, die neuronale, körperliche und umweltliche Aspekte orchestriert (Clark, 1998, S. 45).

Solche Soft-Assembly-Lösungen reichern das Gehirn damit mit sogenannten *Teilprogrammen* (partial programs) zur Erbringung kognitiver Leistungen an (Clark, 1998, S. 159). Jene Aspekte einer Problemstellung, die etwa durch passive Dynamiken oder der durch epistemische Handlungen hervorgebrachten Selbststrukturierung von Informationen gelöst werden können, sind nicht in neuronalen Strukturen verankert und können damit auch nicht im Gehirn selbst verortet werden. Die kognitive Funktion des Gehirns besteht nicht in der passiven, abbildungstreuen, inneren Spiegelung der äußeren Welt, sondern liefert dem Organismus ein nischenabhängiges und handlungszentriertes Teilprogramm, das verlässliche Strukturen des Körpers und der Umwelt zwar implizit annimmt, dabei allerdings nicht nochmals explizit repräsentiert.

The idea here is that the brain should not be seen as primarily a locus of inner *descriptions* of external states of affairs; rather, it should be seen as a locus of inner *structures* that act as operators upon the world via their role in determining actions.

(Clark, 1998, S. 47)

Dieses im biologischen Gehirn verankerte Teilprogramm besitzt also die evolutionsbiologische Funktion, den Gesamtorganismus zuverlässig in einer unsicheren Welt navigieren zu lassen, und adäquate Handlungen zur Abwehr von Gefahren und zur Befriedigung von Bedürfnissen, zu initiieren. Wenn die Aufgabe allerdings auch unter Rückgriff auf körperliche und umweltliche Aspekte erledigt werden kann, so muss ein solches

Teilprogramm eben nicht vollständig sein: Solche Gehirne „[w]ill surely develop computational resources that are complementary to the actions and interventions they control. [...] Its task is to provide computational processes (such as powerful pattern completion) that the world, even as manipulated by us, does not usually afford“ (Clark, 1998, S. 68).

Neuronale Strukturen sollen demzufolge diejenigen informationsverarbeitenden Aufgaben erledigen, die nicht an körperliche oder ökologische Ressourcen ausgelagert werden können. Im Ergebnis entsteht ein vollständiges Programm nur unter Einbeziehung dieser körperlichen und umweltlichen Aspekte.

Every thought is had by a brain. But the *flow* of thoughts and the adaptive success of reason are now seen to depend on repeated and crucial interactions with external resources. The role of such interactions [...] is clearly computational and informational: it is to transform inputs, to simplify search, to aid recognition, to prompt associative recall, to offload memory, and so on. In a sense, then, human reasoners are truly *distributed* cognitive engines: we call on external resources to perform specific computational tasks, much as a networked computer may call on other networked computers to perform specific jobs.

(Clark, 1998, 68f.)

Das neuronal realisierte Teilprogramm wird dabei auch keine internen Repräsentationen zur neutralen bzw. objektiven Beschreibung der Welt liefern. Das Gehirn unterliegt funktional dem Gesamtorganismus, und dieser muss in einer spezifischen ökologischen Nische mit einem ihm eigenen Körper sein Überleben sichern. „The internal representations the mind uses to guide actions may thus be best understood as actions-and-context-specific control structures rather than as passive recapitulations of external reality“ (Clark, 1998, S. 51).

Clark will damit aber nicht die aus seiner Sicht außerordentliche Rolle des Gehirns zur Lösung von Problemstellungen herunterspielen. Vielmehr geht es ihm darum, zu zeigen, dass in Bezug auf das Gehirn vielerlei Fehlvorstellungen — vor allem seitens des Kognitivismus — existieren. Nur wenn diese beseitigt sind, kann eine ehrliche Würdigung neuronaler Informationsverarbeitung stattfinden.

### 5.1.3 Intelligentes Verhalten als emergentes Phänomen

Intelligentes Verhalten ergibt sich für Clark (in manchen Fällen) also aus einem komplexen Zusammenspiel von Gehirn, Körper und Umwelt. Dabei ist dann allerdings zu beachten, dass „[t]o the extent that brain, body, and world can at times be joint participants in episodes of dense reciprocal causal influence, we will confront behavioral unfoldings that resist explanation in terms of inputs to and outputs from a supposedly insulated individual cognitive engine“ (Clark, 1998, S. 165).

Clark stellt sich hier vor, dass intelligentes Verhalten, in enger Anlehnung zur Theorie Dynamischer Systeme, *emergent* aus einem komplizierten interaktiven Wechselspiel zwischen Körper, Gehirn und Umwelt hervortritt. Steht der Organismus vor einem Problem, so aktiviert sich gewissermaßen das neuronale Teilprogramm und verzahnt sich implizit mit jenen körperlichen und umweltlichen Aspekten, vor dessen Hintergrund es sich (evolutionär) entwickelt hat. Daraus ergibt sich dann ein, nicht allein auf neuronale Prozesse beschränktes, *gekoppeltes* dynamisches System. Ein solches System besteht nun aus zahlreichen positiven Feedback-Schleifen, d.h. der Entwicklungsverlauf von Teilkomponenten des Systems ist auch von der Entwicklung anderer Komponenten abhängig.

Dies allerdings wird die Kognitionswissenschaften vor neue *methodische* Herausforderungen stellen. Wie, so die Frage, können solche emergenten intelligenten Verhaltensweisen, die nicht rein neuronaler Natur sind, bestmöglich identifiziert und wissenschaftliche beschrieben werden? Die klassische Rekonstruktion kognitiver Leistungen, die, auf Basis der in Kapitel 1 beschriebenen drei Ebenen von Marr, einen top-down-Ansatz verfolgt, ist Clark (Clark, 1998, S. 81) zufolge dafür nicht gewappnet und zwar gleich aus mehreren Gründen. Einerseits ist sie zu stark auf die algorithmische Ebene fixiert, arbeitet also zu intensiv an der Entwicklung regelbasierter und symbolischer Informationsverarbeitung, und die hier interessierenden Lösungsstrategien besitzen bloß neuronale Teilprogramme. Auch ist es so, dass der Kognitivismus überhaupt alle nicht-neuronalen Aspekte ausklammert und daher das Wesen epistemischer Handlungen überhaupt nicht erkennen kann. Ebenfalls wurde im Kapitel über situierte Kognition darauf hingewiesen, dass Menschen bei der Suche nach Lösungen einen ganz anderen Lösungsraum aufsuchen (können), als dies evolutionsbiologischen Prozessen möglich ist. Schließlich müssen evolutionäre Lö-

sungen immer auch auf Pfadabhängigkeiten eines Organismus Rücksicht nehmen.

Eine adäquate Methode zur Beschreibung emergenten intelligenten Verhaltens muss für Clark (Clark, 1998, 81f.) folgende drei Aspekte umfassen: (i) Die kognitive Aufgabenstellung muss in Begriffen, die auf Ebene des Organismus Relevanz besitzen, gefasst werden. Als Input gelten hier physikalische Quantitäten, und Outputs sind Handlungen des Organismus. Zusätzlich muss das Verhalten des Organismus unter Angabe biologisch realistischer Zeiteinheiten erfolgen. (ii) Ein Bewusstsein für dezentralisierte Lösungen, d.h. die Existenz innerer (neuronaler) Pläne mit hohem Detailgrad, darf nicht automatisch vorausgesetzt werden. Stattdessen muss ein Fokus auf intelligente Verhaltensweisen gelegt werden, die sich als Produkt vieler einfacher Interaktionen zwischen Individuum, dessen Komponenten und der Umwelt darstellen lassen. (iii) Eine erweiterte Vorstellung von Kognition und Berechnung, die nicht allein auf das Gehirn beschränkt sein muss, sondern auch bioexterne Ressourcen umfassen kann.

### 5.1.4 Methodischer Pluralismus

Um diese drei Aspekte methodisch sicherzustellen, ist es für Clark wichtig, drei unterschiedliche methodische Rahmen zur Beschreibung kognitiver Phänomene heranzuziehen. Keiner dieser drei Methoden besitzt dabei eine bevorzugte Stellung, sondern nur eine zusammengenommene Integration dieser Ansätze zu einem methodischen Gesamtbild lässt es möglich erscheinen, emergentes intelligentes Verhalten adäquat abbilden zu können. Damit ist es für Clark notwendig, folgende drei gleichsam wichtigen Typen von Erklärungen und Beschreibungen liefern zu können (Clark, 1998, S. 126): Erstens muss das grobe Gesamtverhalten des, gut auf seine Umwelt eingestellten, Lebewesens abgebildet werden. Zweitens ist eine Darstellung notwendig, die die verschiedenen Komponenten identifiziert, deren kollektive Eigenschaften das grobe Gesamtverhalten ergeben. Hier müssen in weiterer Folge auch daran beteiligte *neuronale* Komponenten und deren Interaktionen identifiziert werden. Drittens muss die informationsverarbeitende Rolle der zuvor identifizierten Komponenten entschlüsselt werden und zwar insbesondere mit Blick auf die nicht-neuronalen Komponenten.

Wie kann das erreicht werden? Clark plädiert hier im Kern für einen methodischen

Pluralismus, der sowohl die Theorie Dynamischer Systeme, als auch die algorithmische Ebene des Kognitivismus miteinbezieht und darüber hinaus, gewissermaßen als vermittelndes erklärendes Element, eine *verteilte funktionale Dekomposition* (distributed functional decomposition) umfasst.

Die Theorie Dynamischer Systeme wurde bereits in Abschnitt 4.2. eingehender beleuchtet, weshalb hier nur kurz erläutert werden soll, worin genau Clark hier die große Stärke dieses Ansatzes sieht. Wenn Intelligenz (in manchen Fällen) tatsächlich durch eine aktive Auseinandersetzung mit der Umwelt entsteht, dann findet aus Sicht von Clark eine Koppelung von zwei komplexen Systemen, nämlich dem Agenten und der Umwelt, statt und das dort hervorgebrachte Wechselspiel erzeugt schließlich die Problemlösung. Die Interaktion zwischen dem Körper eines Lebewesens und der Umwelt sorgt dafür, dass nicht sämtliche relevante Informationen seitens des Organismus internalisiert werden müssen (Clark, 1998, S. 98). Die Sprache der Theorie Dynamischer Systeme eignet sich hierbei perfekt zur Beschreibung der Interaktionen zwischen Körper, Gehirn und Umwelt, weil es alle Aspekte dieser Teilbereiche in denselben Begriffen auszudrücken weiß. „As a result, it becomes especially easy and natural to characterize adaptive behavior in terms of complex couplings of brains, bodies, and environment“ (Clark, 1998, S. 102). Darüber hinaus ist die TDS besonders gut dazu geeignet, solche Phänomene abzubilden, die sich aus zeitlich eng abgestimmten Interaktionen ergeben. Diese Sensibilität gegenüber *zeitlichen Faktoren* zur Hervorbringung kognitiver Lösungsstrategien ist die zweite, methodisch herausragende Eigenschaft dieses Ansatzes.

Ist das grobe Gesamtverhalten einer intelligenten Leistung einmal über die TDS eingefangen, so ist es Aufgabe der *verteilten funktionalen Dekomposition*, die dort involvierten Komponenten zu identifizieren. Das Ziel dieser Methode ist es „to display some target performance as the outcome of an interacting multitude of unintelligent („mechanical“) interactions and effects but to do so relative to a larger organizational whole [...] Such approaches recognize the important contributions that embodiment and environmental embedding can make to the solution of a problem [...]“ (Clark, 2008, S. 14). Hier wird in gewisser Weise der Versuch unternommen, dass eng gekoppelte System in seine funktionalen Bestandteile zu zerlegen. Der Rahmen bleibt auch hier noch interaktiv, weil ein Teil der Analyse, insbesondere auch auf die Interaktionen zwischen den Komponen-

ten, Rücksicht nimmt. Diese Interaktionen werden mechanistisch vorgestellt, also noch nicht als informationsverarbeitende Prozesse. Allerdings soll aus dem (mechanischen) Interaktionsmuster zwischen den Komponenten gefolgert werden, welche funktionalen Rollen diese Komponenten besitzen und verdeutlichen, welchen spezifischen Stellenwert eine Komponente für das System besitzt.

Sobald diese Komponenten identifiziert wurden ist es Aufgabe der *Komponentenanalyse*, die innere Funktionsweise dieser Komponenten offenzulegen. „To explain the functioning of a complex whole by detailing the individual roles and the overall organization of its parts is to engage in componential explanation“ (Clark, 1998, S. 104). Damit beschreibt die Komponentenanalyse im Kern also das Aufgabenfeld der algorithmischen Ebene von Marr. Clark will hier ja auch explizit die informationsverarbeitende Rolle der jeweiligen, auch nicht-neuronalen Komponenten, herausgearbeitet wissen.

Diesen, TDS, verteilte funktionale Dekomposition und Komponentenanalyse vereinigenden methodischen Ansatz bezeichnet Clark als *dynamischen Computationalismus* (dynamical computationalism) „in which the details of the flow of information are every bit as important as the larger scale dynamics and in which some dynamical features lead a double life as elements in an information-processing economy“ (Clark, 2008, S. 27). Daraus ergibt sich, im Hinblick auf die an kognitiven Leistungen beteiligten Ressourcen, eine Art Doppelrolle. Einerseits fungieren sie innerhalb eines dynamischen gekoppelten Systems als Komponenten, aus deren Zusammenspiel zeitlich eng abgestimmte Interaktionsmuster inklusive positiver Feedbackschleifen hervorgehen; andererseits offenbart die Innenansicht der Komponenten eine informationsverarbeitende Funktion. Die von außen empfangenen Signale (mancher) Komponenten werden nicht bloß unmoduliert ins System rückgeschleust, sondern internen informationsverarbeitenden Prozessen zugeführt und umtransformiert.

Dieser doppelte Charakter, der an kognitiven Leistungen beteiligten Komponenten ist dabei durchaus nicht widersprüchlich. „The fact that a cognitive system can be modeled at one level without any explicit mention of information processing does not rule out viewing it at another level as an information-processing system“ (Bermúdez, 2014, S. 421). Wie im Abschnitt über die Theorie Dynamischer Systeme angemerkt wurde, ist jedes System, das regelgeleiteten, zeitlichen Änderungen unterworfen ist, einer dynamischen

Beschreibungsebene zugänglich. Dies gilt insbesondere etwa auch für Computer oder künstliche neuronale Netzwerke.

Dynamical systems models are perfectly compatible with information-processing models of cognition. Dynamical systems models operate at a higher level of abstraction. They allow cognitive scientists to abstract away from details of information-processing mechanisms in order to study how systems evolve over time. But even when we have a model of how a cognitive system evolves over time we will still need an account of what makes it possible for the system to evolve in those ways.

(Bermúdez, 2014, S. 420)

Clark geht davon aus, dass anspruchsvolle kognitive Leistungen biologischer Lebewesen genau diese dynamisch-informationsverarbeitende Charakteristik aufweisen: „The systems have distinct and functionally specialized neural components, but the complex and nonlinear interactions (feedback and feedforward relations) between these components may be crucial determinants of most intuitively ‚psychological‘ phenomena“ (Clark, 1998, S. 114).

Ein dynamischer Computationalismus ist die methodische Konsequenz, die aus den Schlüssen über situierte Kognition und Kognitivismus gezogen werden sollte. Beide Ansätze heben wichtige Eigenschaften kognitiver Phänomene hervor und sind damit für die Kognitionswissenschaft gleichermaßen relevant. Doch auch jeder der Ansätze besitzt gewisse Schwachpunkte. Während der Kognitivismus die komplexen, zeitlich eng abgestimmten Interaktionen die sich zwischen Gehirn, Körper und Umwelt ergeben theoretisch nicht fassen kann, fehlt es Theorien über situierte Kognition, insbesondere der TDS, informationsverarbeitende Prozesse adäquat zu beschreiben. „A successful cognitive science, I shall argue, will thus study both the larger dynamics of agent/environment systems and the computational and representational microdynamics of real neural circuitry“ (Clark, 1998, S. 103).

### 5.1.5 Die Bedeutung von Repräsentationen

Was bereits im vorherigen Abschnitt in Bezug auf Repräsentationen angeklungen ist, soll an dieser Stelle nun explizit gemacht werden.

Während Theorien über situierte Kognition, sowohl was die KL-Forschung als auch die TDS betrifft, eine große Skepsis gegenüber der Annahme interner Repräsentationen besitzen, weisen symbolische Repräsentationen im Kognitivismus einen zentralen Stellenwert zur Vorstellung von Intelligenz auf. Auch der Konnektionismus geht über das subsymbolische Paradigma theoretisch davon aus, dass in den Verbindungsgewichtungen eines Netzwerkes verteilte Repräsentationen existieren, anhand dessen ein Netzwerk auf zielgerichtete Weise Signale verarbeiten kann.

Zur Einordnung über die Existenz und Relevanz innerer repräsentationaler Strukturen geht Clark von der im Kapitel über situierte Kognition bereits vorgestellten These der radikal verkörperten Kognition aus, die hier nochmals in Erinnerung gerufen werden soll: „Structured, symbolic, representational, and computational views of cognition are mistaken. Embodied cognition is best studied by means of noncomputational and non-representational ideas and explanatory schemes involving, e.g., the tools of Dynamical Systems theory“ (Clark, 1998, S. 148).

Mit Blick auf den im vorherigen Abschnitt entwickelten methodischen Ansatz von Clark, dem dynamischen Computationalismus, scheint klar, dass Clark mit dieser These nicht übereinstimmt. Seinen Vorstellungen entsprechend widersprechen sich die Ansichten über situierte Kognition und die Annahme der Existenz repräsentationaler Strukturen auch nicht: „Minds may be essentially embodied and embedded and *still* depend crucially on brains which compute and represent“ (Clark, 1998, S. 143).

Eine dynamische Sichtweise auf kognitive Phänomene kann ohne weiteres auch Repräsentationen annehmen, und manchmal muss sie diese auch annehmen. Das biologische Gehirn, verstanden als ökologische Kontrollstruktur, besitzt die Funktion, einen Organismus zuverlässig und erfolgreich durch eine unsichere Welt zu navigieren. Clark hat klar gemacht, dass es dafür nicht notwendig ist, ein umfassendes inneres Abbild der äußeren Umwelt zu erzeugen, sehr wohl aber eine Art neuronales Teilprogramm zur Verfügung stellen muss, das gewisse körperliche und umweltliche Gegebenheiten implizit

voraussetzt und, darauf aufbauend, erfolgreiche Handlungen initiieren kann. Diese körperlichen und ökologischen Bedingtheiten werden nicht auf explizite Weise repräsentiert, wohl aber so in die neuronale Architektur integriert sein, dass sie neuronalen Subsysteme einen Möglichkeitsraum zu solchen informationsverarbeitenden Prozessen liefern, die auf die Erfordernisse des Gesamtorganismus abgestimmt sind.

Representation talk gets its foothold, I suggest, when we confront inner states that [...] exhibit a systematic kind of coordination with a whole space of environmental contingencies. In such cases it is illuminating to think of the inner states as a kind of code that can express the various possibilities and which is effectively ‚read‘ by other inner systems that need to be informed about the features being tracked.

(Clark, 1998, S. 147)

Darüber hinaus ist es auch so, dass nicht alle Formen kognitiver Leistungen stark auf körperliche und umweltliche Ressourcen zurückgreifen können. Clark gibt hier zu bedenken, dass es auch *repräsentationshungrige Aufgaben* gibt, die insbesondere bei komplexeren Gedankengängen auftreten. Zwei wichtige Formen solch komplexer Denkmuster sind (Clark, 2014, S. 128): (i) Die Koordination eigener Handlungen und Entscheidungen unter Berücksichtigung hypothetischer, kontrafaktischer, oder nicht direkt wahrnehmbarer Sachverhalte. Die Planung eines Familienurlaubs, die Prognose der Konsequenzen vorgestellter Handlungen oder die gedankliche Ausführung arithmetischer Operationen sind Beispiele dafür. (ii) Die Koordination eigener Handlungen und Entscheidungen unter Berücksichtigung umweltlicher Parameter, deren physikalische Manifestationen komplex und unregelmäßig erscheinen. Die Erfassung einer gefährlichen Situation beinhaltet das Konzept der Gefahr, und dieses kann nicht direkt aus den Wahrnehmungssignalen gewonnen werden, sondern benötigt demgegenüber neuronale Mechanismen, die tiefer liegende, d.h. abstraktere, Eigenschaften aus den Wahrnehmungssignalen herausfiltern können.

Wenn also Theorien über situierte Kognition die Existenz innerer Repräsentationen insgesamt ablehnen, dann führt dies, so Clark, zu einem sehr unangemessenen Verständnis der Rolle des Gehirns zur Hervorbringung kognitiver Leistungen.

Demgegenüber macht sich allerdings der Kognitivismus schuldig, repräsentationalen Strukturen eine zu starke Rolle in Bezug auf intelligentes Verhalten zuzuschreiben. Viele kognitiven Phänomene ergeben sich aus den komplexen Interaktionsmustern, die körperliche, umweltliche und neuronale Komponenten gemeinsam hervorrufen. Dafür ist es nicht notwendig, vorauszusetzen, dass das Gehirn eine vollständige und detaillierte innere Abbildung der äußeren Welt mittels objektiver Repräsentationen erstellt. Vielmehr realisieren biologische Gehirne neuronale Teilprogramme auf Basis nischenabhängiger und handlungszentrierter Repräsentationen und integrieren dabei körperliche und umweltliche Aspekte implizit in die Gesamtarchitektur.

Die Synthese aus Gedanken über situierte Kognition und Kognitivismus lässt Clark für einen *minimalen Repräsentationalismus* eintreten, der einerseits intelligentes Verhalten, als durch körperliche, neuronale und umweltliche Ressourcen dynamisch hervorgerufenen, emergentes Phänomen ernst nimmt. Gleichzeitig anerkennt er aber auch die herausragende Rolle neuronaler Prozesse. Methodisch gesehen ergibt sich daraus für ihn ein dynamischer Komputationalismus, der einen minimalen Repräsentationalismus miteinschließt.

### 5.1.6 Das konnektionistische Gehirn

Dynamischer Komputationalismus und minimaler Repräsentationalismus zur Erklärung emergenten intelligenten Verhaltens, bilden zwei fundamentale Säulen innerhalb der Theorie von Clark. Dabei ist immer wieder auf die besondere Rolle des Gehirns hingewiesen worden. Hier nun soll erläutert werden, in welcher Weise sich Clark das biologische Gehirn vorstellt.

Grundsätzlich hält Clark den Konnektionismus für einen vielversprechenden theoretischen Ansatz zur Klärung des Wesens neuronaler Prozesse. Mit dem Konnektionismus stellt sich für Clark das Gehirn im Kern als leistungsfähige Assoziationsmaschine dar, die über wirkmächtige Mechanismen zur Mustererkennung verfügt (Clark, 1998, S. 53). Sensorische Interaktionen mit der Umwelt werden also über ein neuronales Netzwerk konnektionistischer Natur vermittelt, das Berechnungen in Form einer iterativen Serie einfacher Mustervervollständigungen durchführt. Clark hält künstliche neuronale Netz-

werke insbesondere deshalb für biologisch plausibel, weil diese, über eine dem Menschen ähnliches, Stärken- und Schwächenprofil verfügen.

They are resistant to local damage. They are fast. And they excel at tasks involving the simultaneous integration of many small cues or items of information—an ability that is essential to real-time motor control and perceptual recognition. These benefits accrue because the systems are, in effect, massively parallel pattern completers.

(Clark, 1998, S. 59)

In zweiten Kapitel dieser Arbeit wurden diese Charakteristika neuronaler Netzwerke bereits eingehender beschrieben, weshalb an dieser Stelle nicht nochmals darauf eingegangen werden soll. Die parallelisierte Ausführung von Berechnungen, die Fähigkeit zu spontanen Verallgemeinerungen und der langsame, nur allmählich stattfindende, Abbau der Funktionsfähigkeit aufgrund lokaler Beschädigungen des Netzwerks, ergibt insgesamt ausgeprägte Fähigkeiten in der Gesichts- und Stimmerkennung, beim Auslesen handgeschriebener Texte und der Koordination von Bewegungsaktivitäten. Schlecht schneiden solche Netzwerke bei Aufgaben ab, die sequentielle, schrittweise Lösungsstrategien erfordern, die starken Gebrauch von logischen und planerischen Prinzipien machen.

Für Clark ergibt sich hier ein Stärke- und Schwächeprofil, das sehr eng an dasjenige des Menschen erinnert und hält deshalb konnektionistische Ansätze für biologisch überaus plausibel. Nicht eine Turingmaschine, sondern die netzwerkartige Struktur konnektionistischer Modelle, liefert die bessere Vorstellung über die neuronale Architektur des Gehirns. Menschliches Denken selbst sollte sich demzufolge als Mustererkennung und nicht als klassisch-logisches Schlussfolgern vorgestellt werden (Clark, 1998, S. 60).

Mit ihrem subsymbolischen Paradigma, das für Clark eine Art Mikrofunktionalismus begründet (siehe Kapitel 3), zeigt der Konnektionismus auf, wie ein „[p]hysical system like the brain might encode and exploit information and knowledge“ (Clark, 1998, S. 58).

Clark ist deshalb davon angetan, neben einer handlungszentrierten und nischenabhängigen Interpretation repräsentationaler Strukturen, diese auch als *subsymbolisch* und

*verteilt* anzusehen. Die Fähigkeiten eines Netzwerkes liegen verstreut in den gewichteten Verbindungen zwischen den Knotenpunkten und können nur mittels mathematisch-statistischer Techniken als konzeptionelle Einheiten sichtbar gemacht werden. Diese konzeptionellen Einheiten oder verstreuten Ursachen besitzen allerdings durchaus einen repräsentationalen Charakter, auch wenn dies hier anders als seitens eines (diskreten) Symbolismus vorgestellt wird. Die Bezeichnung als Mikrofunktionalismus soll dabei zum Ausdruck bringen, dass die repräsentationalen Strukturen konnektionistischer Netzwerke einen höheren Auflösungsgrad, als diskrete, symbolische Strukturen besitzen, wie dies etwa bei der Repräsentationalen Theorie des Geistes vorgestellt wird.

Clark erkennt hierbei allerdings auch an, dass der Konnektionismus, für sich allein genommen, kognitive Leistungen nicht adäquat abbilden können. Dieser ist, wie auch der Kognitivismus, einer individualistischen Konzeption des Geistes verhaftet, klammert also alles nicht-neuronale aus seinen Betrachtungen aus und unterliegt dabei einem fehlerhaften, gehirnzentrierten Internalismus. Die Lehren aus der Existenz emergenter intelligenter Verhaltensweisen benötigen eben eine agentenzentrierte Sichtweise auf Kognition, die den Fokus auch auf die leiblichen und umweltlichen Aspekte eines intelligenten Organismus legt.

[c]ognitive science can no longer afford simplifications that take the real world and the acting organism out of the loop—such simplifications may obscure the solutions to ecologically realistic problems that characterize active embodied agents such as human beings. Cognitive science’s aspirations to illuminate real biological cognition may not be commensurate with a continuing strategy of abstraction away from the real-world anchors of perception and action. [...] One central theme which has already emerged is that abstracting away from the real-world poles of sensing and acting deprives our artificial systems of the opportunity to simplify or otherwise transform their information-processing tasks by the direct exploitation of real-world structure. Yet such exploitation may be especially essential if we hope to tackle sophisticated problem solving using the kinds of biologically plausible pattern-completing resources that artificial neural networks provide, as we shall now

see.

(Clark, 1998, 59f.)

### **5.1.7 Fazit**

Emergentes intelligentes Verhalten: Aufteilung der Aufgabenlast auf Gehirn, Körper und Umwelt; Selbststrukturierung zur Rekonfiguration und Vereinfachung von Aufgabenstellungen; Theorie über den erweiterten Geist. Dynamischer Computationalismus, der sowohl dynamische, zeitsensible Aspekte, als auch informationsverarbeitende Prozesse berücksichtigen kann. Minimaler Repräsentationalismus, der die Aufgabe des Gehirns in der Bereitstellung eines neuronalen Teilprogramms sieht und dabei gemäß dem Prinzip repräsentationaler Sparsamkeit, achtsam in der Verwendung von Repräsentationen umgeht, und dies als nischenabhängig und handlungszentriert interpretiert. Die Struktur des Gehirns im Bilde eines konnektionistischen Netzwerkes, das Repräsentationen als verteilt deutet und damit ein subsymbolisches Paradigma gemäß eines Mikrofunktionalismus mit hohem Auflösungsgrad etabliert.

Die inneren Kontrollstrukturen dürfen nicht als „highly detailed, complete neural instruction sets“ vorgestellt werden, sondern als ein Muster formendes, sich selbst organisierendes System, das keine objektiv und beobachterunabhängigen Repräsentationen besitzt, sondern handlungszentrierte und auf den Agenten zentrierte Repräsentationen. Gehirn, Körper und Umwelt sind zur Lösung von Problemen aufs tiefste miteinander verwoben.

## **5.2 Die kulturelle Perspektive**

Es stellt sich nun in weiterer Folge die Frage, inwiefern Clarks Theorie auch im Hinblick auf repräsentationshungrige Aufgabenstellungen verständlich gemacht werden kann. Ausgangspunkt ist also der von ihm eingeführte dynamische Computationalismus, der einen minimalen Repräsentationalismus mit nischenabhängigen und handlungszentrierten Repräsentationen umfasst, und dem biologischen Gehirn eine konnektionistische

Funktionsweise zuschreibt. Dieses theoretische Gerüst soll jene kognitiven Phänomene beschreiben und erklären können, die sich aus der Nutzbarmachung neuronaler, körperlicher und umweltlicher Ressourcen in emergenter Weise ergeben. Zwei Varianten solcher Phänomene sind die Verteilung der Aufgabenlast ökologischer Herausforderungen und die Eigenstrukturierung von Informationen zur Vereinfachung kognitiver Problemstellungen.

Ein von der Idee einer starken Kontinuität zwischen Leben und Geist angeleiteter biologischer Inkrementalismus, der sich innerhalb der Theorien über situierte Kognition großer Beliebtheit erfreut, hält Clark für nicht plausibel. Eine rein dynamische Perspektive auf kognitive Leistungen wird der Sonderstellung des Gehirns nicht gerecht:

[t]here is a principled *distinction* between knowledge-based and merely physical-causal systems. [...] Isn't there a (morally and scientifically) crucial distinction between systems that select actions for reasons and on the basis of acquired knowledge, and other (often highly complex) systems that do not display such goal-oriented behavior?

(Clark, 2014, S. 157)

Genau das ist auch der Grund, weshalb Clark den methodischen Ansatz des dynamischen Computationalismus so ausgestaltet, dass auch die algorithmische Ebene des Kognitivismus Berücksichtigung findet, und also explizit das informationsverarbeitende Paradigma miteingeschlossen ist. Die Formulierung eines minimalen Repräsentationalismus und die Identifikation des biologischen Gehirns mit einer konnektionistischen Netzwerkarchitektur bekräftigt diese Sichtweise ein weiteres Mal.

Eine andere Möglichkeit wäre es, sich das Gehirn als gewissermaßen *hybride* Struktur vorzustellen. Einer solchen Sichtweise zufolge stellt das Gehirn einerseits, in Form neuronaler Teilprogramme, die Möglichkeit zur Koppelung an körperliche und umweltliche Ressourcen her, aus deren Interaktionsmuster heraus sich ökologische Herausforderungen des Organismus bewältigen lassen; andererseits verfügt das Gehirn darüber hinaus *zusätzlich* über ein physikalisches Symbolsystem, kraft dessen sequentiell-logische Denkvorgänge realisiert werden können.

Doch auch dies erscheint Clark nicht überzeugend zu sein. Während der biologische Inkrementalismus eine zu starke Kontinuität zwischen elementaren Lebensprozessen und kognitiven Leistungen befürwortet, verortet eine hybride Struktur biologische Gehirne in zu großer Isolation anderen organismischen Prozessen gegenüber.

Die von Clark präferierte Erklärung bedient sich einer, wie er es ausdrückt, subtil modifizierten Kombination eben beschriebener Ansätze „[t]o depict much of advanced cognition as rooted in the operation of the same basic kinds of capacity used for online, adaptive response, but tuned and applied to the special domain of *external and/or artificial cognitive aids—the domain, as I shall say, of wideware or cognitive technology*“ (Clark, 2014, S. 167). In der *aktiven* Auseinandersetzung mit der eigenen Umwelt und unter Zuhilfenahme *kognitiver Technologien* sieht Clark also den Schlüssel zum Verständnis der menschlichen Fähigkeit, repräsentationshungrige Aufgaben lösen zu können. Der Mensch als verkörpertes Wesen besitzt eine ganz spezielle, in den Worten Uexkülls, *effektive Umwelt*, nämlich eine solche, die durch kulturelle Aktivitäten auf die Bedürfnisse des Menschen hin umstrukturiert wurde. In Bezug auf kognitive Problemstellungen des Menschen existieren speziell entworfene *kognitive Nischen*, die den Menschen bei deren Lösung zur Seite stehen. In der interaktiven Auseinandersetzung mit jenen umweltlichen Ressourcen, die (von Menschen kreierte) kognitive Nischen speziell für den Menschen bereithalten, entsteht eine spezielle Klasse emergenter intelligenter Verhaltensweisen, die die *kognitive Sonderstellung* des Menschen begründen und selbst repräsentationshungrige Aufgabenstellungen erfolgreich bewältigen lassen.

Höhere Kognition, und das ist die zentrale Idee von Clark, „[a]rises at the *productive collision points* of multiple factors and forces—some bodily, some neural, some technological, and some social and cultural“ (Clark, 2014, S. 167). Während nicht-menschliche Lebewesen ebenfalls körperliche und umweltliche Aspekte zur Lösung von Problemen heranziehen, besitzen Menschen zusätzlich technologische, soziale und kulturelle Artefakte und Institutionen, mittels derer sie die eigene kognitive Leistungsfähigkeit in unermessliche Höhen haben steigern können.

Welchen Beitrag leisten diese kognitiven Technologien hierbei? „One way to understand the cognitive role of many of our self-created cognitive technologies is as affording complementary operations to those that come naturally to biological brains“ (Clark,

2014, S. 168). Menschen haben kognitive Technologien also deshalb entwickelt, weil sie genau jene *komplementären* Operationen übernehmen können, die das menschliche Gehirn selbst nicht gut meistern kann. Kognitive Nischen sind also eine künstliche, weil bewusst adaptierte, effektive Umwelt des Menschen, die dessen neuronal realisierten Denkprozesse wie ein Stützgerüst umschließen, und so mächtigere kognitive Leistungen hervorbringen können, als dies das Gehirn, für sich alleine genommen, schaffen könnte.

### 5.2.1 Kognitive Nischenbildung

Was genau zählt Clark zu den kognitiven Technologien? Grundsätzlich all jenes, das Menschen bei kognitiven Aktivitäten unterstützt und extra dafür entwickelt wurde. Landkarten, Taschenrechner, Lineale, materielle und virtuelle Modellierungen von Aspekten der Welt, Computer und ähnliche Gegenstände können hier etwa hinzugezählt werden. Das vielleicht wichtigste kognitive Werkzeug ist für Clark die als öffentlich bereitstehender Code zur Verfügung stehende *Sprache*, die die gesamte Umwelt, insbesondere auch kognitive Nischen, des Menschen durchdringt und damit eine Art *metakognitive* Nische bildet. Die Fähigkeit zur Sprache wurde allmählich beim Menschen ausgebildet und seitdem systematisch weiterentwickelt und verfeinert. Sprache wird damit zu einer „[a]nimal-built structure that productively transforms our cognitive capacities“ (Clark, 2014, S. 182).

Was macht Sprache so besonders? Einerseits lässt die externe Verwendung sprachlicher Symbole, bspw. mittels Stift und Papier oder am Computer, menschliche Gedanken systematisch abbilden, und diese, je nach Bedarf, auch weiter auszubauen bzw. umzustrukturieren. Die extern abgebildeten, sprachlich formulierten Gedanken des Menschen fungieren dabei allerdings nicht ausschließlich als externer Speicher; vielmehr können Gedanken nun auch räumlich strukturiert werden oder zu sequenziellen Abfolgen aneinandergereiht werden.

Clark spekuliert ebenfalls, dass die Sprachfähigkeit des Menschen zu sogenannten *kognitiven Dynamiken zweiter Ordnung* geführt hat, d.h. der kognitiven Fähigkeit, Gedanken selbst als Objekte wahrzunehmen und sich so Gedanken über Gedanken machen zu können.

It is easy to see, in broad outline, how this might come about. As soon as we formulate a thought in words (or on paper), it becomes an object for ourselves and for others. As an object, it is the kind of thing we can have thoughts about. In creating the object, we need have no thoughts about thoughts—but once it is there, the opportunity immediately exists to attend to it as an object in its own right. The process of linguistic formulation thus creates the stable structure to which subsequent thinkings attach.

(Clark, 1998, S. 209)

Dies also erklärt für Clark die kognitive Sonderstellung des Menschen.

We, more than any other creature on the planet, deploy nonbiological wieware (instruments, media, notations) to complement our basic biological modes of processing, creating extended cognitive systems whose computational and problem-solving profiles are quite different from those of the naked brain.

(Clark, 2014, S. 179)

### 5.2.2 Die Mangrovenbaum-Metapher

Wenn menschliche Sprachfähigkeit also den Schlüssel zum Verständnis seiner kognitiven Sonderstellung liefert, so stellt sich allerdings die Frage, ob hier nicht eine Form zirkulärer Argumentation vorliegt. Ist es nicht genau die *Intelligenz* des Menschen, die seine Fähigkeit zur Sprache erklärt und, falls ja, kann ebenjene Sprachfähigkeit zur Erklärung menschlicher Intelligenz herangezogen werden?

Clark sieht hier ein Henne-Ei-Problem vorliegen und bemüht zu dessen Lösung die von ihm kreierte *Mangrovenbaum-Metapher* (Clark, 1998, 207f.). Beim Anblick eines Baumes, der sich auf einer Insel befindet, ist es für gewöhnlich richtig anzunehmen, dass dieser deshalb entstehen konnte, weil ihm die bereits vor seinem Dasein existierende Insel die dafür notwendigen Mittel zur Verfügung stellt. Bei Mangrovenbäumen ist dies aber

genau umgekehrt. Deren Samen schwimmen frei in Gewässern und bilden dort in seichten, schlammigen Zonen vertikale Wurzeln aus. Allmählich sammelt sich zwischen den vertikal schwimmenden Wurzeln Schlamm und anderes Material und lässt alsbald eine Insel entstehen, die nunmehr den Eindruck entstehen lässt, der Baum habe sich von Beginn an auf dieser Insel entwickelt. Tatsächlich allerdings wäre es hier richtiger zu sagen, dass sich Baum und Insel gleichzeitig, und zwar wechselseitig zueinander, heranbilden.

So in etwa stellt sich Clark auch die sprachliche Entwicklungsgeschichte beim Menschen vor. In einem Bootstrapping-Prozess haben *protosprachliche* Formen der menschlichen Kommunikation Einfluss auf die neuronale Struktur des Menschen genommen, und diese Veränderungen des Gehirns haben wiederum allmählich sprachliche Ausdrücke verfeinert. Ist dieser Prozess einmal angestoßen, ist der Grundstein zur Entwicklung kognitiver Technologien gelegt und ein weiterer Bootstrapping Prozess initiiert.

[o]nce the process of cultural and technological evolution is underway, the explanation of our contemporary human achievements lies largely in a kind of iterated bootstrapping in which brains and (first-generation) cognitive technologies cooperate so as to design and create the new, enriched technological environments in which (new) brains and (second-generation) cognitive technologies again conspire, producing the third-generation environment for another set of brains to learn in, and so on.

(Clark, 2014, S. 180)

### 5.2.3 Turingmaschine reloaded

Für Clark ergibt sich die kognitive Sonderstellung des Menschen damit insbesondere aus dem wechselseitigen Zusammenspiel des, als konnektionistische Assoziationsmaschine vorgestellten, Gehirns und externer kognitiver Technologie:

One powerful hypothesis [...] is that we transcend these limits, in large part, by combining the internal operation of a connectionist, pattern-completing device with a variety of external operations and tools that serve to redu-

ce the complex, sequential problems to an ordered set of simpler pattern-completing operations of the kind our brains are most comfortable with.

(Clark, 2014, S. 168)

In diesem Zusammenhang ist es interessant, nochmals auf die Arbeiten von Turing, die in Abschnitt 2.1.1. vorgestellt wurden, zu verweisen. Dort wurde ausgeführt, dass Alan Turing mit der von ihm entwickelten Turingmaschine im Kern ein mathematisches Modell baute, das die *Tätigkeiten* einer Person abbildet, während diese einen Algorithmus ausführt. Diese Tätigkeiten wurden dabei ursprünglich von einem textitverkörperten und in eine effektive Umwelt *eingebetteten* kognitiven Agenten ausgeführt. Die Bearbeitung logischer Problemstellungen seitens des Menschen wurden also erst durch das Modell der Turingmaschine ihrer leiblichen und umweltlichen Aspekte beraubt. Clark (2001, S. 140) verweist hier auf den kognitiven Anthropologen Ed Hutchins der in seinem Werk *Cognition in the Wild* aus alledem den Schluss zieht, dass das Modell der Turingmaschine nicht das menschliche Gehirn selbst beschreiben wollte. „It was a model [...] not, in the first instance, of what goes on inside the individual head, but of the kinds of serial, symbol-matching problem-solving that we engage in using pen, paper and other external props and artifacts“ (Clark, 2001, S. 140). Die Turingmaschine bildet damit also denjenigen Prozess mathematischer und logischer Tätigkeiten ab, die eine Person unter Zuhilfenahme kognitiver Technologien betreibt. Die Fähigkeit von Turingmaschinen bzw. Computern zur Lösung logisch-sequentieller Aufgabenstellungen verweist damit also *nicht* auf die neuronalen Fähigkeiten menschlicher Gehirne, sondern bildet demgegenüber jene kognitiven Leistungen des Menschen ab, die dieser im Zusammenspiel mit körperlichen und umweltlichen Ressourcen zusammengenommen realisiert.

Damit allerdings ergibt sich für die Theorie des Kognitivismus, dass dieser sich mit dem Computerbild des Geistes und der Hypothese über physikalische Symbolsysteme von einer falschen Vorstellung über das Wesen des menschlichen Gehirns hat anleiten lassen. Das traditionelle Bild der Kognitionswissenschaft verweist auf die Fähigkeiten eines *erweiterten* Systems, bestehend aus Gehirn, Körper und kognitiven Technologien, verortet dieses aber irrtümlich in den neuronalen Prozessen des biologischen Gehirns.

The image of serial, clunky, symbol manipulation as the deep explanation of the mechanical possibility of reason is thus plausibly taken as a repressed image of the *distributed mechanisms of environmentally situated reason*, rather than as a revealing image of the operation of the inner-biological aspects of that extended system.

(Clark, 2001, 140f.)

Dieses erweiterte System realisiert sich Clarks Ansicht zufolge aus der neuronalen Einbindung von Operationen kognitiver Technologien, die der Mensch in Form *sensomotorischer Schleifen* aktiv hervorbringt.

Vor diesem Hintergrund stellt Clark sich die Frage, ob ein solch erweitertes System nicht als erweitertes kognitives System angesehen werden muss bzw. inwiefern solche erweiterten Systeme erweiterte kognitive Prozesse realisieren, die auch nicht-neuronale bzw. bioexterne Ressourcen umfassen. Damit ist der Ausgangspunkt der These über den erweiterten Geist gelegt, der nun im nachfolgenden Kapitel genauer betrachtet werden soll.

### 5.3 Zusammenfassung

Inspiziert durch Theorien über situierte Kognition wählt Clark als Ausgangspunkt seiner Überlegungen eine *agentenzentrierte* Sichtweise auf kognitive Herausforderungen. Es erscheint plausibel, dass umweltliche Aufgabenstellungen sich stets für den Gesamtorganismus stellen, der verkörpert und in eine spezifische ökologische Nische eingebettet ist und charakteristische Bedürfnisse zur Lebenssicherung besitzt. Verkörperung und effektive Umwelt legen den Schluss nahe, dass die Wahrnehmungsinhalte eines Organismus nischenabhängig und handlungszentriert abgebildet werden. Von einem Lebewesen wahrgenommene Problemstellungen können immer unter Verwendung neuronaler, körperlicher und umweltlicher Ressourcen bewältigt werden. Aus dem Prinzip repräsentationaler Sparsamkeit ergibt sich, dass Organismen dies i.d.R. auch tun werden. Empirische Belege dafür liefert etwa der Nachweis über die Existenz emergenter intelligenter

Verhaltensweisen, die sich bspw. aus der Aufteilung der Aufgabenlast von Problemstellungen oder aus der Selbststrukturierung von Informationen zur Vereinfachung von Problemen ergeben. Der methodisch richtige Rahmen zur adäquaten Beschreibung solcher kognitiven Phänomene ist ein dynamischer Computationalismus, der sowohl dynamische Interaktionen, die zwischen neuronalen, körperlichen und umweltlichen Aspekten bestehen abbilden kann, als auch spezifisch informationsverarbeitende Prozesse einzelner Komponenten kognitiver Leistungen berücksichtigt. Die Fähigkeit des Menschen, sich hypothetische oder kontrafaktische Situationen vorstellen zu können, spricht zumindest für einen minimalen Repräsentationalismus, d.h. die situierte Kognition mit ihrer grundlegenden Skepsis Repräsentationen gegenüber, schießt hier über das Ziel hinaus. Allerdings ist es mit Blick auf das Stärken- und Schwächenpotential bei Menschen so, dass nicht die Turingmaschine, sondern demgegenüber konnektionistische Netzwerke ein plausibleres Bild für das biologische Gehirn liefern. Die kognitive Sonderstellung des Menschen ergibt sich weniger aus der Annahme eines biologischen Inkrementalismus oder einer hybriden Architektur des Gehirns, sondern aus einer fruchtbaren Interaktion zwischen Menschen und den von ihnen entwickelten kognitiven Technologien. Dieses erweiterte System scheint auch eher jene Fähigkeiten zu besitzen, die der Kognitivismus Menschen zwar ebenfalls zuordnet, diese allerdings fälschlicherweise allein neuronal im Gehirn realisiert sieht. Das Modell der Turingmaschine bezog sich von Anfang an auf die aktiven, gekoppelten Tätigkeiten des Menschen inklusive kognitiver Technologien.

# Kapitel 6

## These der erweiterten Kognition

Gemäß dem traditionellen informationsverarbeitenden Paradigma, das sowohl den Kognitivismus als auch den Konnektionismus eint, sind kognitive Prozesse *exklusiv* neuronale Tätigkeiten des biologischen Gehirns. Im vorangegangenen Kapitel konnte allerdings mittels den theoretischen Überlegungen von Clark aufgezeigt werden, dass manche kognitive Phänomene nicht ausschließlich vom Gehirn realisiert werden, sondern auch körperliche und umweltliche Ressourcen miteinbeziehen. Solch emergente, intelligente Verhaltensweisen finden sich bspw. in Situationen, wo die kognitive Last ökologischer Herausforderungen auf Gehirn, Körper und Umwelt aufgeteilt wird, oder in Fällen der Selbststrukturierung von Information zur Rekonfiguration und Vereinfachung kognitiver Problemstellungen. Der von Clark vorgeschlagene dynamische Komputationalismus soll diesen Phänomenen methodisch Rechnung tragen.

Insbesondere in Bezug auf den Menschen, der seine effektive Umwelt in vielen Fällen selbstständig zu kognitiven Nischen hin umtransformiert hat, in denen eine Vielzahl kognitiver Technologien existieren, scheint eine individualistische Sichtweise auf kognitive Leistungen fragwürdig zu sein.

Die von Clark vorgelegte alternative Sichtweise betrachtet die kognitiven Fähigkeiten des Menschen als durch ein enges Wechselspiel zwischen kognitiven Technologien und neuronalen Prozessen realisiert, die, durch Tätigkeiten des Körpers vermittelt, eine höhere kognitive Kraft entfalten, als dies der Mensch mit seinem biologischen Gehirn allein bewerkstelligen könnte. Die enge Koppelung zwischen Mensch und solchen

Technologien wäre demzufolge als erweitertes kognitives System zu verstehen, das tatsächlich auch jene logisch-sequentielle Berechnungskraft besitzt, wie dies der Kognitivismus dem Menschen für sich genommen unterstellt. Mathematisch-logische Arbeiten des Menschen mittels Zettel und Stift offenbaren dann auch den rechten Platz für diskrete symbolische Strukturen. Die schriftliche Multiplikation ist tatsächlich eine Verarbeitung diskreter Symbolstrukturen und zwar mittels eines konnektionistischen Gehirns und verkörperter Handlungen im Medium Papier.

Die Frage, die sich nun stellt, ist, ob in solchen Fällen nicht tatsächlich ein *erweitertes* kognitives System vorliegt, das kognitive Prozesse hervorbringt, die sich über den Organismus hinaus erstrecken und auch bioexterne Ressourcen miteinschließen. Vor diesem Hintergrund also lassen sich die Anliegen, die Andy Clark in seinem gemeinsam mit David J. Chalmers verfassten und 1998 erschienen Artikel *The Extended Mind*, adäquat nachvollziehen.

In weiterer Folge sollen nun die wesentlichen Inhalte des Artikels vorgestellt werden, wobei hierbei die Einschränkung gilt, dass ausschließlich auf die *These der erweiterten Kognition* (TEK) eingegangen werden wird. Der Artikel selbst formuliert darüber hinaus auch eine stärkere These, nämlich die *These des erweiterten Geistes* (TEG). Bei der These der erweiterten Kognition handelt es sich im Kern um eine Aussage über die *physische Realisierung kognitiver Prozesse*. Sie ist damit eine These über das Wesen der an kognitiven Prozessen beteiligten *Vehikel*, d.h. über die *materiellen Träger* kognitiver Prozesse.

Demgegenüber bezieht sich die These des erweiterten Geistes auf die *mental*en Gehalte, die solche kognitiven Prozesse bei einem kognitiven Agenten hervorbringen (Lyre, 2010, S. 190).

In der angelsächsischen Literatur werden unter dem Schlagwort *extended mind* sowohl die TEK als auch die TEG ausverhandelt.

## 6.1 These der erweiterten Kognition

Mit der These der erweiterten Kognition (TEK) wird im Wesentlichen versucht, eine Art *erweiterten Funktionalismus* im Hinblick auf kognitive Prozesse zu etablieren. „Es scheint sogar, als ob das Konzept erweiterter Kognition nichts anderes, als einen konsequent zu

Ende gedachten Funktionalismus darstellt“ (Lyre, 2010, S. 197). In Abschnitt 1.2. der vorliegenden Arbeit wurde bereits eingehend auf das Wesen funktionalistischer Theorien, in Bezug auf mentale Prozesse, eingegangen. Mentale Prozesse werden hier nicht als materielle Prozesse vorgestellt, sondern als durch solche *realisiert* verstanden. Prozesse sind dann als mental anzusehen, wenn diese innerhalb eines Gesamtsystems eine spezifische kausale Rolle besitzen. In der klassischen Variante (und in Bezug auf biologische Lebewesen) ist ein kognitiver Prozess ein neuronaler Prozess, der eine *informationsverarbeitende Leistung* für das Gehirn erfüllt.

### 6.1.1 Ausgangspunkt

Den Ausgangspunkt des Artikels (Clark & Chalmers, 1998, S. 8) bildet ein Gedankenexperiment, bei dem sich drei Personen vor unterschiedlichen Computerbildschirmen befinden, auf denen zweidimensionale geometrische Objekte und verschiedene Passformen abgebildet sind. Sie haben dabei die Aufgabe herauszufinden, welche Objekte welchen Passformen zugeordnet werden können.<sup>1</sup>

Zur Erledigung dieser Aufgabe kann die erste Person nur auf ihre neuronale Grundausstattung, also das Gehirn, zurückgreifen und dabei versuchen, die geometrischen Figuren in Gedanken rotieren zu lassen und so der entsprechenden Passform zuzuweisen. Die zweite Person kann zusätzlich mittels einer Tastatur die geometrischen Objekte am Bildschirm selbst rotieren lassen. Die dritte Person schließlich besitzt ein neuronales Implantat, das die Rotation der Figuren elektronisch durchführt und das Ergebnis an entsprechende neuronale Areale des Gehirns sendet.

Die Autoren stellen sich diesbezüglich die Frage, wie viel Kognition zur Lösung der Aufgabe von den jeweiligen Personen eingesetzt werden musste. Sie kommen zu dem Schluss, dass sich die Situation der ersten Person tatsächlich nicht sonderlich von jener Situation unterscheidet, vor die die dritte Person gestellt wurde. In beiden Fällen haben

---

<sup>1</sup>Das Gedankenexperiment ist vom Computerspiel Tetris inspiriert

innere, d.h. im Kopf stattfindende, Prozesse zur Bewältigung der Aufgabe beigetragen; dass die dritte Person dabei auf einen neuronal integrierten Computerchip zurückgegriffen hat, ändert nichts daran, dass der Prozess *insgesamt* ein *kognitiver* ist. Wird dieser Sichtweise zugestimmt, so stellt sich die Frage, ob es einen wesentlichen Unterschied zwischen den Situationen gibt, in denen sich die zweite Person gegenüber der dritten befindet. Während die dritte Person auf ein elektronisches Hilfsmittel zurückgreift, das sich im Gehirn befindet, greift die zweite Person, mittels von ihr initiierten Handlungen, auf Berechnungen eines Computers zurück. Clark und Chalmers halten es hierbei für wenig plausibel, den neuro-elektronischen Aktivitäten des Gehirn-Chip-Systems einen kognitiven Status zu verleihen, wenn dieser nicht auch den Aktivitäten des Gehirn-Computer-Systems zugestanden wird. Schließlich handelt es sich auch hierbei um (sensomotorisch koordinierte) neuro-elektronische Aktivitäten. Einzig die *räumliche Verteilung* der verschiedenen Ressourcen begründet einen Unterschied zwischen Person zwei und Person drei, der allerdings im Hinblick auf die Verleihung eines kognitiven Status als nebensächlich erscheint. Aus einem unvoreingenommener Blickwinkel betrachtet, der kognitive Prozesse nicht schon im Vorhinein mit *neuronalen* Prozessen gleichsetzt, scheint ein Verweis auf die räumliche Verteilung der Ressourcen willkürlich zu sein: „We cannot simply point to the skin/skull boundary as justification, since the legitimacy of that boundary is precisely what is at issue“ (Clark & Chalmers, 1998, f.).

### **6.1.2 Schlussfolgerung**

Für Clark und Chalmers etabliert dieses Gedankenexperiment nicht bloß die theoretische Möglichkeit, dass nicht-neuronale bioexterne Ressourcen in kognitiven Prozessen integriert sein können; vielmehr heben sie hervor, dass es dafür eine Fülle empirischer Belege gibt. Beipielsweise die schriftliche Durchführung einer langen Multiplikation, die Umsortierung von Scrabble-Teilchen, die Verwendung sprachlicher Symbole oder die Erstellung von Diagrammen: „In all these cases the individual brain performs some operations, while others are delegated to manipulations of external media“ (Clark & Chalmers, 1998, S. 8).

Dadurch ergibt sich ein dynamisches Wechselspiel zwischen neuronalen und nicht-

neuronalen Komponenten: „[t]he human organism is linked with an external entity in a two-way interaction, creating a coupled system that can be seen as a cognitive system in its own right“ (Clark & Chalmers, 1998, S. 8).

Gekoppelte Systeme werden durch *epistemische Handlungen* des Menschen initiiert, deren Zweck darin besteht, Teile der äußeren Welt, zur Unterstützung und Anreicherung kognitiver Prozesse, abzuändern. Daraus ergibt sich in weiterer Folge die Forderung, all jene *epistemischen Leistungen*, die ein auf solche Weise etabliertes gekoppeltes System hervorgebracht hat, gerecht auf die daran beteiligten (auch nicht-neuronalen) Komponenten aufzuteilen (Clark & Chalmers, 1998, S. 8).

Daraus ergibt sich nun die *These der erweiterten Kognition*. Ihr zufolge sind kognitive Systeme „keineswegs auf den lokalen Verarbeitungsapparat, die neuronale Maschinerie, beschränkt, sondern erstrecken sich, über die traditionellen Systemgrenzen hinaus, in den Körper, die Umgebung und externe kognitive Werkzeuge“ (Lyre, 2010, S. 190). Oder prägnanter auf den Punkt gebracht: „Cognitive processes ain't (all) in the head!“ (Clark & Chalmers, 1998, S. 8).

Zur Feststellung, ob bei einer bestimmten kognitiven Leistung auch erweiterte kognitive Prozesse beteiligt sind, haben Clark und Chalmers das sogenannte *Paritätsprinzip* entwickelt, das wie folgt lautet:

If, as we confront some task, a part of the world functions as a process which, were it done in the head, we would have no hesitation in recognizing as part of the cognitive process, then that part of the world *is* (so we claim) part of the cognitive process.

(Clark & Chalmers, 1998, S. 8)

Das Paritätsprinzip (PP) bedient sich dabei einer *funktionalistischen* Sichtweise: „If a coupled process has the relevant functionality [...] then it doesn't matter whether that process is partly, or indeed, mostly external. Focus on the function, not the location, is the purpose of the PP“ (Menary, 2010, S. 6).

Es stellt dabei eine *Heuristik* zur Identifikation erweiterter kognitiver Prozesse dar und soll vor einem biologischen Chauvinismus bewahren. Es dient gewissermaßen als eine

Art Schleier metabolischen Nichtwissens im Sinne von Rawls und stellt „[a]n den Vertreter einer konservativen internalistischen Position die Frage, was denn das biologische Vehikel und seine Grenzen so besonders macht und auszeichnet“ (Lyre, 2010, S. 195).

Erweiterte kognitive Prozesse werden also über gekoppelte Systeme hervorgebracht, deren daran beteiligten Komponenten sich zu einem neuartigen und erweiterten kognitiven System integrieren und in komplexen kausalen Wechselbeziehungen zueinander stehen:

Although we can identify the relevant components, and factorize them into internal and external components, the nature of reciprocal coupling makes it difficult to study the components as separate systems because they are continuously influencing and responding to one another. They are coordinating with one another to produce behavior. Insofar as brain, body, and world can be shown to be reciprocally coupled in this way, we can consider them to be a coupled system.

(Menary, 2010, S. 4)

*Kontinuierliche reziproke Kausalitäten* sind dabei in Anlehnung an den Begriff der *Koppelung* aus der Theorie Dynamischer Systeme zu verstehen. Sie sind überall dort gegeben, wo der Entwicklungsverlauf einer Komponente eines Gesamtsystems von der Entwicklung anderer Komponenten bzw. des Gesamtsystems abhängig ist. In diesen Fällen ist eine reine Analyse der jeweiligen Komponenten in Isolation zu den anderen Komponenten methodisch nicht zielführend, weil dadurch die komplexen kausalen Wechselbeziehungen zwischen den Komponenten nicht adäquat abgebildet werden können. (Clark, 2008, S. 24).

Insgesamt legt die These der erweiterten Kognition damit eine externalistische Sichtweise auf kognitive Phänomene nahe: Kognitive Prozesse finden nicht allein im biologischen Gehirn statt, sondern dehnen sich auch auf bioexterne Ressourcen aus. Clark und Chalmers bezeichnen diese Sichtweise als *aktiven Externalismus* der postuliert, dass „[...] if we retain internal structure but change the external features, behaviour may change completely. The external features here are just as causally relevant as typical internal

features of the brain“ (Clark & Chalmers, 1998, S. 9). Sofern epistemische Handlungen gekoppelte Systeme realisieren, ist es einem aktiven Externalismus zufolge so, dass die körperlichen Aktivitäten des Menschen über bioexterne Ressourcen weniger den Charakter einer Handlung besitzen, sondern demgegenüber als Teil eines Denkprozesses anzusehen sind (Clark & Chalmers, 1998, S. 10).

Ein aktiver Externalismus betont also die aktive Rolle umweltlicher Ressourcen und sehen diese deshalb als für kognitive Prozesse *konstitutiv* an. „It is not simply that the external features, to which the organism is interactively linked, have a causal influence on the cognitive processing of the organism; rather, the interactive link *is* the cognitive processing“ (Menary, 2010, S. 2). „Die Erweiterungsthese sieht hier keine kausale Abhängigkeit, sondern eine Konstitutionsbeziehung dergestalt, dass kognitive Prozesse [...] durch hybride, sich über die Grenzen des Gehirns hinaus in die Umwelt erstreckende, Prozesse realisiert sind“ (Walter, 2013, S. 193).

Aus diesem Grund wird der aktive Externalismus manchmal auch als *Vehikel-Externalismus* bezeichnet. Dieser will hervorheben, dass die *physikalischen Realisierer* kognitiver Prozesse nicht allein in den neuronalen Strukturen des Gehirns vorzufinden sind, sondern auch nicht-neuronale Ressourcen umfassen. Der Begriff des *Vehikels* bezieht sich dabei also auf das materielle Substrat, über das kognitive Prozesse physikalisch realisiert werden. So wie Sprache etwa in Form von Schallwellen zum Ausdruck gebracht wird, sind auch kognitive Prozesse auf einen materiellen Träger angewiesen. „Die Erweiterungsthese stellt eine These über die Lokalisierung dessen auf, was als ›Realisierer‹, ›materielles Substrat‹, ›Implementierung‹ oder ›Vehikel‹ kognitiver Prozesse bzw. Systeme bezeichnet wird“ (Walter, 2013, S. 193).

Ebenfalls stellen Clark und Chalmers drei Bedingungen auf, die festlegen, in welchen Fällen ein gekoppeltes kognitives System vorliegt.

All the components in the system play an active causal role, and they jointly govern behaviour in the same sort of way that cognition usually does. If we remove the external component the system's behavioural competence will drop, just as it would if we removed part of its brain. Our thesis is that this sort of coupled process counts equally well as a cognitive process, whether

or not it is wholly in the head.

(Clark & Chalmers, 1998, 8f.)

Ein erweitertes kognitives System liegt also dort vor, wo (i) allen Komponenten eine aktive kausale Rolle zukommt und diese gemeinsam das Verhalten dieses Systems so festlegen, wie dies nicht-erweiterte kognitive Prozesse üblicherweise tun; (ii) muss gelten, dass die Entfernung einer externen Komponente die kognitiven Fähigkeiten des Systems verringert, wie dies auch der Fall wäre, wenn Teile des biologischen Gehirns entfernt würden. Darüber hinaus müssen (iii) externe Komponenten *zuverlässig* arbeiten und für den Menschen leicht *zugänglich* sein (Lyre, 2010).

### 6.1.3 Einordnung

Bereits im vorhergehenden Kapitel wurden zwei Varianten emergenter kognitiver Phänomene vorgestellt, nämlich einerseits die Verteilung der Aufgabenlast auf Gehirn, Körper und Geist, und andererseits die Selbststrukturierung von Informationen zur Vereinfachung von Aufgabenstellungen. Mit der TEK liegt nun eine dritte Form eines emergenten kognitiven Phänomens vor. Hierbei bedient sich der Mensch bioexterner Ressourcen zur Erweiterung der eigenen kognitiven Kompetenzen.

Insgesamt legen diese drei Formen emergenter kognitiver Leistungen für Clark den Schluss auf die *Hypothese kognitiver Objektivität* (hypothesis of cognitive impartiality) nahe, die wie folgt lautet:

Hypothese kognitiver Objektivität: Our problem-solving performances take shape according to some cost function or functions that, in the typical course of events, accord no special status or privilege to specific types of operation (motoric, perceptual, introspective) or modes of encoding (in the head or in the world).

(Clark, 2008, S. 121)

Zur Generierung kognitiver Lösungen gibt es demgemäß keine Bevorzugung in der Verwendung neuronaler Strukturen; vielmehr ist es so, dass eine Kostenfunktion, die Faktoren wie Zeit, Energieeinsatz u.ä. beinhaltet, festlegt, welche Ressourcen zur Bewältigung einer Aufgabe herangezogen werden. Aus der Hypothese kognitiver Objektivität folgt für Clark in weiterer Folge das *Prinzip ökologischer Fertigung* (principle of ecological assembly), das folgendes besagt:

Prinzip ökologischer Fertigung: [...] *the canny cognizer tends to recruit, on the spot, whatever mix of problem-solving resources will yield an acceptable result with a minimum of effort* [...].

(Clark, 2008, S. 13)

In Bezug auf diese merkt Clark an: „It is important that, according to the PEA, the recruitment process marks no special distinction among neural, bodily, and environmental resources except insofar as these somehow affect the total effort involved“ (Clark, 2008, S. 13).

#### 6.1.4 Gehirn und Körper

Welche biologischen Ursachen haben nun solche Lösungsstrategien entwickelt? Clark liefert dafür sowohl einen evolutionsbiologischen, als auch einen ontogenetischen Erklärungsansatz.

Aus *evolutionsbiologischer* Sichtweise wendet Clark sich explizit gegen ein dualistisches Verständnis im Hinblick auf die menschliche *Biologie* und menschliche *Kultur*:

Such a picture [...] invites us to believe in something like a basic biological human nature, gradually co-opted and obscured by the trappings and effects of culture and society [...] Instead we humans are, by nature, products of a complex and heterogenous developmental matrix in which culture, technology and biology are pretty well inextricably intermingled.

(Clark, 2001, 137f.)

Wie im vorangegangenen Kapitel ausführlich dargestellt, sieht Clark also die natürliche Entwicklung hin zum modernen Menschen in einem Gleichschritt mit kulturellen und technologischen Innovationen realisiert. Somit hat der Mensch nicht nur seine eigene effektive Umwelt abgeändert, sondern diese abgeänderte Umwelt hat selbst wiederum Einfluss auf den Entwicklungsverlauf des Menschen genommen. Im Ergebnis ist der moderne Mensch ein tief in die eigene Umwelt eingebettetes Wesen, dessen biologische Eigenschaften ebenso integrale Wesensmerkmale darstellen, wie dies auch kulturelle und technologische Ressourcen tun. Das menschliche Gehirn hat dieses tiefliegende Wechselspiel aus biologischen und bioexternen Ressourcen ebenfalls antizipiert:

Ours are (by nature) unusually plastic brains whose biologically proper functioning has always involved the recruitment and exploitation of non-biological props and scaffolds. More so than any other creature on the planet, we humans are *natural-born cyborgs*, factory tweaked and primed so as to participate in cognitive and computational architectures whose bounds far exceed those of skin and skull.

(Clark, 2001, S. 138)

*Ontogenetisch* betrachtet, hebt Clark des Weiteren die überaus lange Lernphase beim Menschen hervor, womit die Kindheit gemeint ist. Die sich evolutionsbiologisch entwickelte neuronale Plastizität menschlicher Gehirne lässt Clark, mit Blick auf die Kindheitsphase, eine *neurokonstruktivistische* Sichtweise vertreten. Der Neurokonstruktivismus charakterisiert den menschlichen Cortex, insbesondere den Neocortex und den präfrontalen Cortex, als plastische Struktur, deren wichtigste Aufgabe darin besteht, in Abhängigkeit zu gemachten Lernerfahrungen, neuronale Verbindungen zu konstruieren, die auch verlässliche Regularitäten der Außenwelt miteinbeziehen.

One upshot is that the learning device *itself* changes as a result of organism-environmental interactions—learning does not just alter the knowledge base for a fixed computational engine, it alters the internal computational architecture itself.

(Clark, 2001, S. 137)

Welche Rolle spielt dabei nun der menschliche Körper? Clark sieht den menschlichen Körper als Ausgangspunkt intentionaler Handlungen, die den Berührungspunkt zwischen Organismus und Welt herstellen: Er ist ein „locus of sensing and action“, der ein Tor für die Auslagerung kognitiver Prozesse öffnet, die als mobile Brücke fungiert, die eine Verbindung zwischen internen und externen informationsverarbeitenden Ressourcen herstellt und als primäres Werkzeug zur intelligenten Nutzung umweltlicher Strukturen anzusehen ist (Clark, 2008, 206f.).

Der menschliche Körper übernimmt dabei jene Rolle, die der Corpus callosum für das Gehirn darstellt:

Both are key physical structures whose cognitive role is in part to allow distinct sets of resources to engage in highly integrated forms of problem-solving activity [...] But I am inclined to go further and to assert not just that this is what the body *does* but that this [...] is what, at least for all cognitive scientific purposes, the body *is*.

(Clark, 2008, S. 207)

## 6.2 Kritik

Wie gezeigt wurde, postuliert die Theorie der erweiterten Kognition genau bei jenen kognitiven Prozessen ihre Erweiterungsthese, wenn sich neuronale Tätigkeiten des Gehirns eng an bioexterne Ressourcen koppeln und damit ein eigenständiges dynamisches System etablieren: „These are the cases when we confront a recognizably cognitive process, running in some agent, that creates outputs [...] that, recycled as inputs, drive the cognitive process along“ (Clark, 2008, S. 131). Der Körper stellt dabei jene sensomotorischen Aktivitäten zur Verfügung, die neuronale und umweltliche Prozesse integrieren. Nach Sichtweise der TEK sollen solche kognitiven Prozesse als *erweiterte* kognitive Prozesse angesehen werden.

### 6.2.1 Der Kopplungs-Konstitutions-Fehlschluss

Einer der wohl bekanntesten Einwände, in Bezug auf das Argument gekoppelter kognitiver Prozesse, wurde von den beiden Philosophen Fred Adams und Kenneth Aizawa erhoben. „When Clark makes an object cognitive when it is connected to a cognitive agent, he is committing an instance of a ‚coupling-constitution fallacy.‘ This is the most common mistake the extended mind theorists make“ (Adams & Aizawa, 2010, 67f.). Dieser *Kopplungs-Konstitutions Fehlschluss* geht dabei aus folgendem Irrtum hervor: „The pattern of reasoning here involves moving from the observation that process  $X$  is in some way causally connected (coupled) to a process  $Y$  of type  $\phi$  to the conclusion that  $X$  is part of a process of type  $\phi$ “ (Adams & Aizawa, 2009, S. 81). Adams und Aizawa wollen damit zum Ausdruck bringen, dass ein, für sich allein betrachteter nicht-kognitiver und bioexterner Prozess, nicht nur dadurch den Status des kognitiven bekommt, weil er an einen kognitiven und neuronalen Prozess gekoppelt ist „[T]he fact that object or process  $X$  is coupled to object or process  $Y$  does not entail that  $X$  is part of  $Y$ “ (Adams & Aizawa, 2010, S. 68). Es ist falsch, anzunehmen, dass die Koppelung eines umweltlichen Prozesses an einen kognitiven Prozess, diesen ebenfalls in den Status des Kognitiven erhebt. Überspitzt bringen sie Clark’s Denkfehler auf folgende Weise zum Ausdruck: „Question: Why did the pencil think that  $2 + 2 = 4$ ? Clark’s answer: Because it was coupled to the mathematician“ (Adams & Aizawa, 2010, S. 67).

Clark entgegnet diesem Vorwurf mit einer ähnlich absurden Fragestellung und will damit ein Missverständnis aufzeigen, dem Adams und Aizawa selbst unterliegen: „Question: Why did the V4 neuron think that there was a spiral pattern in the stimulus? Answer: Because it was coupled to the monkey“ (Clark, 2008, S. 86). Für sich genommen besitzt, und darauf will Clark hinweisen, auch ein einzelnes Neuron keine kognitiven Eigenschaften, weil es an das kognitive System eines Primaten gekoppelt ist. Der Kopplungs-Konstitutions Fehlschluss bringt also etwas Wahres zum Ausdruck. Adams und Aizawa interpretieren dabei allerdings das Argument der Koppelung falsch, welches in keiner Art und Weise das Ziel hat, bioexterne Ressourcen für sich genommen als kognitiv zu identifizieren: „The appeal to coupling is not intended to make any external object cognitive [...] Rather, it is intended to make some object [...] into a *proper part of some*

*cognitive routine*“ (Clark, 2008, S. 87). Damit ist gemeint, dass, aus bestimmten gekoppelten Prozessen insgesamt, erweiterte kognitive Prozesse entstehen, nicht aber, dass aus der Koppelung jene isolierten Prozesse, aus denen sich der erweiterte Prozess zusammensetzt, selbst kognitive Eigenschaften erhalten.

Let us first be clear then about the precise role of the appeal to coupling in the arguments for the extended mind. The appeal to coupling is not intended to make any external object ‚cognitive‘ [...] Rather, it is intended to make some object, which in and of itself is not usefully [...] thought of as *either cognitive or noncognitive*, into a *proper part of some cognitive system*, such as a human agent. It is intended [...] to ensure that the putative part is poised to play the kind of role that *itself* ensures its status as part of the agent’s cognitive routines.

(Clark, 2010a, S. 83)

Sofern bioexterne Ressourcen innerhalb erweiterter Prozesse eine adäquate funktionale, d.h. kognitive, Rolle erfüllen, werden sie gewissermaßen durch ihren eigenen Beitrag, den sie für den erweiterten kognitiven Prozess erfüllen, zu einem echten Bestandteil bzw. Konstituenten eines kognitiven Prozesses.

Die eigentliche Frage, die sich aus der Erweiterungsthese ergibt, ist also, in welchen Fällen physikalische Objekte oder Prozesse Teil einer größeren kognitiven Routine werden und *nicht*, unter welchen Umständen eine isolierte Betrachtung dieser externen Konstituenten ein kognitiver Status zugeschrieben werden kann (Clark, 2008, 87f.). Teile eines kognitiven Prozesses müssen nicht selbst kognitiv sein, sowie Teile einer sportlichen Aktivität nicht selbst wiederum eine sportliche Aktivität sein müssen.

## 6.2.2 Ein Kennzeichen für Kognition

Ein weiterer Einwand von Adams und Aizawa will, in Hinblick auf die TEK, darauf aufmerksam machen, dass zur Bewertung, ob bestimmte Prozesse einen kognitiven Status besitzen oder nicht, ein Kriterium zur Identifizierung von Kognition notwendig ist.

[i]f the fact that an object or process X is coupled to a cognitive agent does not entail that X is part of the cognitive agent's apparatus, what does? *The nature of X of course*. One needs a theory of what makes a process a cognitive process. One needs a theory of the ,mark of the cognitive.'

(Adams & Aizawa, 2010, S. 68)

Ein solches allerdings wird von Clark und Chalmers nicht geliefert, weshalb die Frage nach der Existenz erweiterter kognitiver Prozesse damit nicht beantwortet werden kann.

Inwiefern die TEK aber ein Kennzeichen für Kognition liefern muss, um erweiterte kognitive Prozesse hinreichend identifizieren zu können, ist fraglich. Zwar liefern Clark und Chalmers tatsächlich kein eindeutiges Kriterium zur Bestimmung des Kognitiven; nichtsdestotrotz enthält das Paritätsprinzip aber einen nicht unwesentlichen Hinweis auf erweiterte kognitive Prozesse. Formuliert wurde es wie folgt:

„If, as we confront some task, a part of the world functions as a process which, were it done in the head, we would have no hesitation in recognizing as part of the cognitive process, then that part of the world *is* (so we claim) part of the cognitive process.“

Das Paritätsprinzip wurde demnach so formuliert, dass es innerhalb der Kognitionswissenschaft hinreichend klar ist, wodurch kognitive Prozesse sich anderen Phänomenen gegenüber auszeichnen. In weiterer Folge plädiert das Prinzip dafür, externe Prozesse, die solch eine funktionale Rolle erfüllen, dass sie im Falle einer neuronalen Realisation zweifelsohne als Teil eines kognitiven Prozesses identifiziert würden, ebenfalls als Teil eines kognitiven Prozesses anzusehen sind. Das Paritätsprinzip umgeht damit elegant der Notwendigkeit einer Definition von Kognition. Damit bleibt also offen,

„[o]b sich dadurch überhaupt ein irgendwie gearteter Einwand gegen die erweiterte Kognition ergibt. Denn die EC-These beansprucht ja nicht aus sich heraus, eine Definition dessen zu geben, was Kognition ihrem Wesen nach ist. Für die Anwendung des Paritätsprinzips geht man von einem bereits gegebenen kognitiven System aus, und fragt danach, ob ein bestimmter kognitiver Zustand oder Prozess auf einem anderen, als dem traditionell internalistischen Weg, realisiert werden kann. Man bezieht sich also direkt auf die Möglichkeit der multiplen Realisierung kognitiver Zustände und Prozesse.

(Lyre, 2010, S. 197)

### 6.2.3 Intrinsische Inhalte

Sofern Adams und Aizawa zugestanden würde, dass ein eindeutiges Kriterium zur Identifikation kognitiver Phänomene notwendig ist, kann eine weitere Kritik der TEK gegenüber erhoben werden.

Als plausibles Kennzeichen von Kognition gilt den beiden das Vorhandensein *nicht-abgeleiteter* bzw. *intrinsischer Inhalte*: „[C]ognition involves particular kinds of processes involving non-derived representations“ (Adams & Aizawa, 2001, 52f.).

Der Begriff *intrinsisch* bringt dabei zum Ausdruck, dass repräsentationale Zustände im Gehirn von Lebewesen ihre Inhalte gewissermaßen kraft ihrer selbst besitzen und nicht anderswo abgeleitet wurden. Wörter der menschlichen Sprache besitzen hingegen ihre inhaltliche Bedeutung auf abgeleitete Weise; die Bedeutung eines Wortes leitet sich, (bspw.) aus dem Gebrauch dieses Wortes innerhalb einer Sprachgemeinschaft, ab. Insofern verfügen Wörter also über extrinsische Inhalte. Folgt aus dieser Bedingung nach der Existenz intrinsischer Inhalte in Bezug auf kognitive Prozesse nun aber auch, dass sämtliche Komponenten darüber verfügen müssen?

Zur Bewertung dieser Fragestellung legt Clark ein hypothetisches und ein reales Beispiel vor. Einerseits stellt er die hypothetische Frage, wie wir auf die Entdeckung einer Spezies auf dem Planeten Mars reagieren würden, die folgende Eigenschaft besitzt: „What if we found Martians whose biological routines stored *bitmapped images* of printed words that they could later access (and interpret) via bitmapped signals sent to visual cortex“ (Clark, 2010b, S. 48)? Clark meint, dass es hier wohl offensichtlich ist, dass jene Prozesse, die für die punktweise Abspeicherung von Bildern zuständig sind, als Teile des kognitiven Systems der Marsianer anzusehen sind. Diese Bilder allerdings müssen selbst erst von einem anderen Teil des kognitiven Systems *interpretiert* werden, d.h. für sich genommen besitzen diese Bilder *keine intrinsischen* Inhalte. Die inneren Bilder der Marsianer besitzen denselben repräsentativen Gehalt, wie die von uns *extern* auf Computern punktweise abgespeicherten Grafiken. Auch diese erhalten ihren Gehalt erst durch die Interpretation eines kognitiven Betrachters (Clark, 2008, S. 91).

Als zweites Beispiel betrachtet Clark den Fall eines Menschen, sich vor seinem inneren Auge ein Venn-Diagramm zur Lösung einer kognitiven Aufgabenstellung vergegenwärtigt. Die mengentheoretische *Bedeutung* sich zweier überschneidender Kreise in einem Venn-Diagramm ist dabei ebenfalls nicht *intrinsischer* Natur, sondern eine bloß mathematische *Konvention*. „Yet this image can clearly feature as part of a genuinely cognitive process“ (Clark, 2010b, S. 48). Daraus schließt Clark: „From the requirement [...] that every truly cognitive agent trade in states that bear intrinsic contents, it cannot follow that every proper part of the cognitive system of an agent must trade (and trade solely) in such contents“ (Clark, 2008, 91f.).

Von diesen Beispielen offensichtlich beeindruckt, versuchen Adams und Aizawa ihren Einwand zu präzisieren: „Clearly, we mean that if you have a process that involves no intrinsic content, then the condition rules that the process is non-cognitive“ (Adams & Aizawa, 2010, S. 70).

Dadurch allerdings, dass ein kognitiver Prozess, nur als *Ganzes* betrachtet, nicht-abgeleitete Inhalte aufweisen muss, löst sich der Einwand gegenüber erweiterten kognitiven Prozessen in Luft auf, da diese, mittels ihrer neuronalen Bestandteile, stets über intrinsische Inhalte verfügen werden.

#### 6.2.4 Natural Kinds

In Bezug auf ein von Adams und Aizawa eingefordertes Kriterium zur Identifizierung von Kognition, gibt es auch noch einen weiteren Einwand gegen die Theorie der erweiterten Kognition. Dieser stellt zunächst einmal fest, dass es Aufgabe der verschiedenen Wissenschaften ist, ihren Gegenstandsbereich klar zu definieren und dabei möglichst so vorzugehen, dass in ihm ein gewissermaßen von der Natur vorgegebener Phänomenbereich abgebildet ist. Dies sollte auch in Bezug auf die Kognitionswissenschaften gelten: „[t]he cognitive must be discriminated on the basis of underlying causal processes“ (Adams & Aizawa, 2001, S. 52). Für Adams und Aizawa stellt sich der Stand der Dinge innerhalb der traditionellen Kognitionswissenschaft allerdings so dar, als wäre dies bereits gelungen. Interne Prozesse kognitiver Agenten lassen ein Untersuchungsgebiet entstehen, das sich unter einem einheitlichen Ansatz heraus erforschen lässt. Das Hinzufügen

bioexterner Ressourcen würde die Einheit dieses Ansatzes in Gefahr bringen:

Tools do not constitute a natural kind; tools are, after all, artifacts. It is for this reason that, a would-be brain-tool science would have to cover more than just a multiplicity of causal processes. It would have to cover a genuine motley. A brain-tool science would not have to cover a mere disjunction of things; it would have to cover an open disjunction.

(Adams & Aizawa, 2010, S. 76)

Führt die Akzeptanz der TEK also dazu, dass sich der Fokus der Kognitionswissenschaft in einer unübersichtlichen Vielzahl unterschiedlichster kognitiver und kognitiv-erweiterter Phänomene verliert, und dabei der Blick für das Wesentliche aus den Augen gerät? Braucht es, wie Adams und Aizawa vorschlagen, ein Bekenntnis zu einer Basis bestimmter kausaler Prozesse im Hinblick auf Kognition, und ergibt sich daraus eine Ablehnung erweiterter kognitiver Prozesse gegenüber?

Für Clark ist dies alles andere als klar. Einerseits weist er darauf hin, dass eine Sichtweise, die Intelligenz ausschließlich mit Prozessen des Gehirns identifiziert, sowohl *anthropozentrisch* als auch *neurozentristisch* ist. Sie ist deshalb *anthropozentrisch*, weil nicht ausgeschlossen werden sollte, dass im Universum auch ganz andere Formen von Intelligenz realisiert sein könnten, auf die Forschungen zur Intelligenz Antworten geben müssten. Weiters ist diese Sichtweise *neurozentristisch*, weil damit die wichtige Rolle körperlicher und bioexterner Ressourcen verkannt bleibt. „To demand identity of fine-grained causal role is surely to set the cognitive bar too high and way too close to home“ (Clark, 2008, S. 93).

Auch ist nicht offensichtlich, weshalb der Gegenstandsbereich der Kognitionswissenschaft nicht einfach all jenes umfassen sollte, das der Hervorbringung intelligenten Verhaltens dient. Damit sollte sich die Kognitionswissenschaft nicht, wie von Adams und Aizawa vorgeschlagen, Kognition auf Basis einer Menge ihr zugrunde liegender kausaler Prozesse definieren, sondern überall dort von Kognition sprechen, wo intelligentes Verhalten beobachtbar ist.

That is to say, we should individuate the cognitive by its characteristic effects, not by its characteristic causes. The notion of a cognitive process, if that is correct, is best unpacked as the notion of a process that supports certain kinds of behavior (actual and counterfactual) [...] Why ask for more?

(Clark, 2010a, S. 93)

Sollte daraus folgen, dass die Kognitionswissenschaft mehrere *unterschiedliche* Ansätze zur Klärung der ganzen Bandbreite kognitiver Phänomene benötigt, so liegt dies dann eben im Wesen von Kognition selbst. Gemeinsam wäre diesen allerdings ein geteiltes Ziel, namentlich die Erklärung intelligenter Verhaltensweisen (Clark, 2008, S. 95). Darüber hinaus wendet Clark auch ein, dass überhaupt nicht feststeht, ob mit der These der erweiterten Kognition nicht auch ein allen intelligenten Verhaltens zugrunde liegender Mechanismus identifiziert werden könnte. Clark hat mit seinem dynamischen Komputationalismus gerade auch darauf aufmerksam gemacht, dass er sich kognitive Phänomene immer auch als *informationsverarbeitende* Prozesse vorstellt. Nur schließt dies nicht automatisch aus, wie dies üblicherweise vonseiten des Kognitivismus oder Konnektivismus getan wird, dass Verarbeitungen von Informationen auch von nicht-neuronalen Ressourcen übernommen werden können:

Traditionally, the sorts of computation that govern cognition have been thought to begin and end at the skull. Computationalism has thus been viewed as entailing an individualistic view of cognition. But [...] why think that the skull constitutes a magic boundary beyond which cognitively relevant computation ends and mere causation begins? We are creatures embedded in informationally rich and complex environments. The computations that occur inside the head are an important but nonexhaustive part of the corresponding computational systems. A wide or extended computational perspective opens up the possibility of exploring computational units that include the brain together with aspects of the brain's beyond-the head environment. Wide computational systems thus literally extend beyond the confines of the skull into the world.

(Wilson & Clark, 2009, S. 60)

In diesem Zusammenhang wird auch angemerkt, dass die Vorstellung von erweiterten informationsverarbeitenden Systemen außerhalb der Kognitionswissenschaft völlig unstrittig ist.

Computational processes occur within discrete entities – whether they be biological cells, computer chips, or larger entities comprised of these units – but they can also occur between such units. Typically, such extended computational processes constitute a larger computational system, but that should not obscure the fact that, with respect to those discrete, metal-shell bounded units, genuine computational processes physically extend beyond the boundaries of those same units. There is nothing ersatz, for example, about the computations that flow across a grid system or a local area network.

(Wilson & Clark, 2009, S. 60)

Selbst innerhalb von Teilen der Kognitionswissenschaft, besonders in der KI-Forschung, erzeugt eine neurozentristische Sichtweise auf Berechnung ein Spannungsverhältnis zum informationsverarbeitenden Paradigma. Der Berechnungsbegriff wird dort recht weit ausgelegt und ist von der Idee angeleitet, „[t]hat pretty much any kind of processing or encoding can form part of an information-based system for flexible adaptive response, just as long as it is properly located in some larger ongoing web of activity“ (Clark, 2008, S. 106).

### **6.2.5 Die These der eingebetteten Kognition**

Schließlich existiert auch eine vom Philosophen Robert D. Rupert entwickelte Kritik gegenüber der TEK. Diese Kritik nimmt dabei auf zwei Dinge Bezug: Erstens hält Rupert, wie auch Adams und Aizawa, die Beiträge interner und externer Ressourcen für so unterschiedlich, dass sie sich nicht in einer einzigen Wissenschaft sinnvoll integrieren lassen.

The extended approach seems to offer developmental psychologists no more reason to be interested in, for example, the series of temporal segments we

normally associate with Sally from ages two to six rather than to be interested in, say, Sally, aged two, together with a ball she was bouncing on some particular day, Johnny, aged five, together with the book he was reading on some particular afternoon, and Terry, aged seven, plus the stimulus item he has just been shown by the experimenter. It is simply not clear how one should proceed after giving up the traditional method.

(Robert D. Rupert, 2010, S. 330)

Rupert will damit zum Ausdruck bringen, dass eine solchermaßen erweiterte Erforschung von Kognition ein *methodischer* Albtraum wäre, oder zumindest nahelegen würde, erweiterte kognitive Systeme begrifflich strikt von konventionellen kognitiven Systemen abzugrenzen.

Ob Rupert damit tatsächlich Recht hat, soll an dieser Stelle nicht weiter ausgeführt werden. Mit Verweis auf den vorherigen Abschnitt sei allerdings dazu kurz angemerkt, dass ein erweitertes Verständnis von Berechnung sehr wohl die Möglichkeit eines einheitlichen Forschungsprogramms begründen könnte.

Als zweiten Einwand macht Rupert geltend, dass erweiterte gekoppelte Systeme zu meist außerordentlich kurzlebig sind. Dies allerdings führt dazu, dass sie außerhalb des Erkenntnisinteresses psychologischer und kognitionswissenschaftlicher Forschung geraten:

We want to understand how and why the capacities and abilities of individual persisting systems change over time, eventually taking a stable form. If the systems to be investigated were relatively short-lived coupled systems, developmental inquiry would seem incoherent. We want to be able to explain why, for example, the child categorizes on the basis of appearance at age two but pays more attention to insides at age five.

(Robert D. Rupert, 2010, S. 330)

Dabei ist allerdings fraglich, ob erweiterte kognitive Systeme tatsächlich so kurzlebig sind, wie Rupert dies behauptet. Die Existenz von Sprache, Stift und Zettel, Smartphone

und Computersystem scheint in (manchen) menschlichen Gesellschaften allgegenwärtig zu sein. Die Wahrscheinlichkeit, dass Menschen auf gewisse erweiterte kognitive Prozesse mittels eines Smartphones nicht zugreifen können, ist in heutigen Zeiten vermutlich nicht höher, als wenn sie aufgrund von Krankheit, medikamentösen Einflüssen o.ä. auf spezifische neuronale Ressourcen nicht zugreifen könnten.

Insgesamt veranlasst Rupert dies zur Formulierung einer eigenen, im Vergleich zur TEK *abgeschwächten* These. Diese als *These der eingebetteten Kognition* lautet folgendermaßen:

Cognitive processes depend very heavily, in hitherto unexpected ways, on organismically external props and devices and on the structure of the external environment in which cognition takes place.

(Robert D. Rupert, 2004, S. 393)

Rupert meint — und Clark stimmt ihm in dieser Hinsicht zu —, dass seine These alle kognitiven Phänomene miteinbezieht, die auch die TEK geltend machen würde. Auch wäre die These der eingebetteten Kognition dazu imstande, eine Erklärung solcher Phänomene anzubieten. Sie hätte allerdings den Vorteil gegenüber der TEK, dass sie eine klare begriffliche Trennung zwischen konventionellen und eingebetteten kognitiven Phänomenen vornimmt, und den Begriff der Kognition auf neuronale Prozesse einschränkt.

### 6.3 Zusammenfassung

Mit der Theorie der erweiterten Kognition wird in den dynamischen Computationalismus von Andy Clark nun also auch explizit eine These erweitert, die kognitive Prozesse des Menschen in manchen Fällen als in die Umwelt hinein ausgeweitet betrachtet. In solchen Fällen scheint die feste Grenze zwischen Mensch und Umwelt aufgehoben zu sein und ein erweitertes kognitives System hervorzubringen, dessen Intelligenz auch durch nicht-neuronale Konstituenten realisiert wird. Mit dem Paritätsprinzip wurde der Gedanke formuliert, dass informationsverarbeitende Prozesse kognitiver Aktivitäten nicht

ausschließlich mit Leistungen des Gehirns in Verbindung gebracht werden sollte. Richtig betrachtet, zeichnet sich das Wesen der Informationsverarbeitung durch eine funktionale Rolle aus, die es innerhalb eines kognitiven Prozesses spielt, um selbst Teil dieses kognitiven Prozesses dasein. Ob diese Rolle von neuronalen Strukturen, körperlichen Merkmalen oder umweltlichen Prozessen realisiert wird, ist dabei unerheblich.

Eine abschließende Bewertung der TEK, insbesondere im Hinblick auf die Theorie der eingebetteten Kognition, soll nun im nachfolgenden und abschließenden Kapitel dieser Arbeit gegeben werden.

# Kapitel 7

## Fazit

In Bezug auf die TEK wird in manchen Fällen behauptet, dass sie im Grundsatz eine rein *empirische* These über kognitive Prozesse zum Ausdruck bringt:

[d]ie kognitiven Neurowissenschaften sind es, die bestimmen müssen, welche ‚Komponenten [...] zur Durchführung, Aufrechterhaltung und Stabilisierung kognitiver Fähigkeiten und Aktivitäten benötigt werden‘. Die These bedarf also des Nachweises durch die entsprechende empirische Forschung und Modellbildung. Falls sie sich behauptet — wofür eine Menge spricht — führt sie zu einer erheblichen Abänderung unserer Auffassung von kognitiven Systemen und ihrer Dynamik.

(Lyre, 2010, S. 194)

Ob dies tatsächlich so ist, wird aber schon durch die Theorie der eingebetteten Kognition infrage gestellt. Ob nun nicht-neuronale Prozesse zur Hervorbringung kognitiver Leistungen als Teil eines kognitiven Prozesses selbst anzusehen sind, oder demgegenüber neuronal realisierte kognitive Prozesse bloß stark unterstützend zur Seite gestellt sind, wird empirisch nicht zu klären sein.

Die Debatte, ob die Theorie der eingebetteten Kognition oder der TEK der Vorzug gegeben werden sollte, läuft hier Gefahr, zu einer reinen Auseinandersetzung über Definitionen zu verkommen (Shapiro, 2011, S. 194).

Auch Clark scheint an manchen Stellen eher *pragmatische* Gründe dafür zu liefern, weshalb ein Standpunkt gemäß der Erweiterungsthese eingenommen werden sollte: Einerseits sieht er damit die Gefahr abgewendet, neuronalen Prozessen eine allzu mystische Kraft beizumessen. „The HEC reminds us that the neural goings-on are not blessed with some intrinsic property that makes them alone suitable to act as the circuitry of mind and intelligence“ (Clark, 2008, S. 136). Andererseits beugt sie auch der Fehlvorstellung vor, im Kopf des Menschen einen inneren Homunkulus zu verorten: „The HEC reminds us that there is no single, all-powerful, hidden agent inside the brain whose job is to do *all the real thinking* and which is able to intelligently organize all those teams of internal and external supporting structure“ (Clark, 2008, S. 136).

Ich persönlich denke daher, dass die Theorie von Clark bei aller wissenschaftlichen Plausibilität auch eine stark *metaphysische* Schlagrichtung besitzt. Nochmals auf die von Rupert vertretene Theorie der eingebetteten Kognition verweisend, hat Clark folgendes Gedankenexperiment entwickelt.

Dafür wird ein neu entdecktes, biologisches Wesen vorgestellt, dessen Lebenszyklus sich durch komplexe morphologische Veränderungen auszeichnet. Im Laufe seines Lebens durchlebt es eine Vielzahl struktureller Veränderungen seines Körpers, die sich durch das Hinzukommen oder den Verlust von Gliedmaßen, Augen, Ohren oder Flügel bemerkbar machen. Dies geschieht in Abhängigkeit zu den ökologischen Herausforderung, vor die dieses Wesen gestellt wird. Dabei ist es in der Lage, neue körperliche Strukturen hinzuzugewinnen, indem es Teile der äußeren Umwelt an sich reißt und anschließend, mittels biochemischer Vorgänge, in den eigenen Körper integriert: „Bits of tree, metal, sap, plastic, neoprene, and stone, are all fair game as seedcore for the newly emerging bodily forms and structures, which then persist or decay according to need, use, and the vagaries of enabling metabolism“ (Clark, 2014, S. 208). Passend zu dieser Charakteristik, bezeichnet Clark dieses Wesen als *Metamorpho*.

Auch wenn dieses Wesen zunächst seltsam anmuten mag, so gibt Clark in weiterer Folge zu bedenken, dass Metamorpho eine vielleicht größere Ähnlichkeit zu Eigenschaften des Menschen aufweist, als vermutet. Ist es nicht so, dass Menschen eine Art mentaler Metamorpho sind? „We might be like Metamorpho (Metamento?) in that bits of the encountered and designed world become more repeatedly and deeply incorporated into our

cognitive routines, persisting or decaying according to need, use, and the vagaries of our enabling sociotechnological cocoon“ (Clark, 2014, S. 209).

Was ist das *physiologische* Wesen von Metamorpho? Was ist das *kognitive* Wesen des Menschen? Das Gedankenexperiment will zeigen, dass es im Hinblick auf die Physiologie von Metamorpho zwei Möglichkeiten der Deutung gibt: Einerseits könnte die Ansicht vertreten werden, dass allein jene, stets unveränderlich bleibenden Teile des Körpers von wissenschaftlichem Interesse sind. Andererseits könnte die Wissenschaft aber auch den Standpunkt vertreten, dass sich die physiologische Natur von Metamorpho genau seine morphologische Veränderlichkeit ist. Welche Vorstellung würde diesbezüglich plausibler klingen?

Clark selbst vertritt die Überzeugung, dass darauf keine Antwort gegeben werden müsste, da sowohl die eine als auch die andere Sichtweise legitim ist und also gar nicht miteinander in Konkurrenz stehen. Ebenso sollte dies auch in Bezug auf die Kognitionswissenschaft gelten.

Inwiefern dieses Gedankenexperiment nun allerdings Aufschluss darüber gibt, ob nun die TEK der Theorie der eingebetteten Kognition gegenüber der Vorzug eingeräumt werden sollte oder nicht, bleibt zweifelhaft. Letztere stellt ja gerade eben *nicht* in Zweifel, dass kognitive Phänomene auch durch nicht-neuronale Prozesse gestützt werden können und deshalb auch in den Fokus kognitionswissenschaftlicher Untersuchungen rücken sollte; vielmehr will sie eine grundlegende Skepsis in Bezug auf den kognitiven Status nicht-neuronaler Ressourcen zum Ausdruck bringen.

Vielleicht verhält es sich so, dass mit dem Gedankenexperiment eher eine grundlegende *metaphysische* Überzeugung von Clark offengelegt wird. Er, so scheint sich die tiefe Natur des Menschen als eine im Bilde von Metamorpho vorzustellen.

Noch deutlicher tritt diese Wesensbestimmung des Menschen in seinem Werk *Natural-Born Cyborgs. Minds, Technologies, and the Future of Human Intelligence* hervor. Dort nämlich wird gleich zu Beginn offengelegt, dass Clark hier an der Entwicklung einer Art *philosophischen Anthropologie* gelegen ist: „[T]his is not primarily a book about new technology. Rather, it is about us, about our sense of self, and about the nature of the human mind“ (Clark, 2003, S. 7).

Clark vertritt hier die Überzeugung, dass Menschen über eine Art *weiches Selbst* verfü-

gen, dass sie besonders dafür eignet, mit einer Vielzahl bioexterner Ressourcen tiefliegende Verflechtungen einzugehen. Das überaus hohe Maß an neuronaler Plastizität menschlicher Gehirne sorgt dafür, auch *außerorganismische* Strukturen in das innere Selbstmodell des Menschen integrieren zu können. Das Gehirn besitzt die biologische Funktion einer *ökologischen Kontrollstruktur* und diese bezieht all jene Prozesse, über die sie hinreichend verfügen kann, in das eigene Selbstbild mit ein. In Bezug auf den Gebrauch von (kognitiven) Werkzeugen bedeutet dies bspw., dass „[i]f the links are sufficiently rich, fluid, bidirectional, fast, and reliable, then the interface between the conscious user and the tool is liable to become transparent, allowing the tool to function more like a proper part of the user“ (Clark, 2003, S. 103).

Neuronale Verbindungen sind ihrer funktionalen Rolle nach nicht bedeutsamer als visuelle oder siliziumbasierte Verbindungen. Was zählt, ist der Grad an Integration, den eine spezifische Verbindung zwischen zwei Einheiten hervorbringt. Clark denkt, dass in manchen Fällen auch sensomotorisch vermittelte Verknüpfungen zwischen neuronalen und bioexternen Prozessen so robust sind, dass es eigentlich passender wäre, das Leib-Seele-Problem in *Leib-Seele-Umwelt-Problem* umzubenennen:

It is the problem of understanding how human thought and reason is born out of looping interactions between material brains, material bodies, and complex cultural and technological environments. We create these supportive environments, but they create us too. We exist, as the thinking things we are, only thanks to a baffling dance of brains, bodies, and cultural and technological scaffolding. Understanding this evolutionarily novel arrangement is crucial for our science, our morals, and our self-image both as persons and as a species.

(Clark, 2003, S. 11)

Neben erweiterten *kognitiven* Prozessen führt all dies auch dazu, dass der menschliche Körper selbst, zu etwas höchst *Aushandelbaren* wird: „The body image itself is highly negotiable, and the brain is plastic enough to learn to exploit whole new kinds of feedback loop and action-potential“ (Clark, 2003, S. 111), weshalb auch den *biologisch fixierten* Merkmalen des Körpers kein privilegierter Status zugedacht wird: „The biological

skin-bag has no special significance here. It is the flow that counts“ (Clark, 2003, S. 114). Insofern also Teile der Umwelt hinreichend in neuronale Kontrollstrukturen integriert sind und gut abgestimmt auf die Aktivitäten des biologischen Körpers sind, kann der Mensch sich auch im Bilde eines Metamorpho begreifen:

Our sense of bodily presence is always constructed on the basis of the brain’s ongoing registration of correlations. If the correlations are reliable, persistent, and supported by a robust, reliable causal chain, then the body image that is constructed on that basis is well grounded. It is well grounded regardless of whether the intervening circuitry is wholly biological or includes nonbiological components.

(Clark, 2003, S. 105)

Insbesondere mit Blick auf zukünftige Technologien verhält es sich so, dass diese ganz neue Dimensionen von Mensch-Technologie-Verschmelzungen ermöglichen werden: „As technology becomes portable, pervasive, reliable, flexible, and increasingly personalized, so our tools become more and more a part of who and what we are“ (Clark, 2003, S. 10).

Dafür ist es für Clark gar nicht notwendig, nicht-biologische Technologien auf operative Weise dem biologischen Körper zuzuführen. Hinreichend responsive technologische Geräte werden auch durch sensomotorische Schleifen ausreichend ins menschliche Selbstmodell integriert werden können. Google-Glasses wäre hier etwa ein Beispiel für eine solche Technologie, die das Sehfeld ihres Trägers mit virtuellen Objekten und Informationen anreichert.

Menschen denken dank der neuronalen Plastizität des Gehirns *mit* und durch kognitive Technologien *hindurch*: Menschen sind *natural-born cyborgs*.

Pretty soon, and still without the need for wires, surgery, or bodily alterations we shall all be [...] cyborgs not in the merely superficial sense of combining flesh and wires but in the more profound sense of being human-technology symbionts: thinking and reasoning systems whose minds and selves are spread across biological brain and nonbiological circuitry.

(Clark, 2003, S. 3)

Aus alledem geht meines Erachtens hervor, dass Clark die TEK gegenüber der Theorie der eingebetteten Kognition auch deshalb verteidigen möchte, weil mit ihr auch ein implizites Verständnis über das Wesen des Menschen verbunden ist.

Zweifellos ist es so, dass die Theorie der erweiterten Kognition etwas Wesentliches über kognitive Prozesse zum Ausdruck bringt; inwiefern ihr allerdings gegenüber der Theorie der eingebetteten Kognition der Vorzug zu geben ist, scheint insbesondere auch eine Frage nach der anthropologischen Bestimmung des Menschen zu sein.

# Literaturverzeichnis

- Abrahamsen, A. & Bechtel, W. (2012). Introduction. In W. M. R. Keith Frankish (Hrsg.), *The Cambridge Handbook of Cognitive Science* (S. 9–28). Cambridge: Cambridge University Press.
- Adams, F. & Aizawa, K. (2001). The Bounds of Cognition. *Philosophical Psychology*, 14(1), 43–64.
- Adams, F. & Aizawa, K. (2009). Why the Mind is Still in the Head. In P. Robbins & M. Aydede (Hrsg.), *The Cambridge Handbook of Situated Cognition* (S. 78–95). Cambridge: Cambridge University Press.
- Adams, F. & Aizawa, K. (2010). Defending the Bounds of Cognition. In R. Menary (Hrsg.), *The Extended Mind* (S. 67–80). Cambridge: The MIT Press.
- Barker-Plummer, D. (2016). *Turing Machines*. In E. N. Zalta (Hrsg.), *The Stanford Encyclopedia of Philosophy* (Winter 2016 Edition). Zugriff unter <https://plato.stanford.edu/archives/win2016/entries/turing-machine/>. (letzter Zugriff: 31. Januar 2020)
- Bermúdez, J. L. (2005). *Philosophy of Psychology. A Contemporary Introduction*. New York: Routledge.
- Bermúdez, J. L. (2014). *Cognitive Science. An Introduction to the Science of the Mind* (2. Aufl.). New York: Cambridge University Press.

- Besold, T. R. & Kühnberger, K.-U. (2013). Konnektionismus, neuronale Netze und parallel distributed processing. In A. Stephan & S. Walter (Hrsg.), *Handbuch Kognitionswissenschaft* (S. 164–169). Stuttgart: J.B. Metzler.
- Bickle, J. (2014). The Neurophilosophies of Patricia and Paul Churchland. In A. Bailey (Hrsg.), *Philosophy of Mind. Key Thinkers*. (S. 237–258). London: Bloomsbury.
- Buckner, C. & Garson, J. (2019). *Connectionism*. In E. N. Zalta (Hrsg.), *The Stanford Encyclopedia of Philosophy* (Fall 2019 Edition). Zugriff unter <https://plato.stanford.edu/entries/language-thought/>. (letzter Zugriff: 31. Januar 2020)
- Chalmers, D. J. (2019). Extended Cognition and Extended Consciousness. In M. Colombo, E. Irvine & M. Stapleton (Hrsg.), *Andy Clark and his Critics* (S. 9–20). New York: Oxford University Press.
- Clancey, W. J. (2009). Scientific Antecedents of Situated Cognition. In P. Robbins & M. Aydede (Hrsg.), *The Cambridge Handbook of Situated Cognition* (S. 11–34). Cambridge: Cambridge University Press.
- Clark, A. (1998). *Being There. Putting Brain, Body, and World Together Again*. Cambridge, Massachusetts: The MIT Press.
- Clark, A. (2001). Reasons, Robots and the Extended Mind. *Mind & Language*, 16(2), 121–145.
- Clark, A. (2003). *Natural-Born Cyborgs. Minds, Technologies, and the Future of Human Intelligence*. New York: Oxford University Press.
- Clark, A. (2008). *Supersizing the Mind. Embodiment, Action, and Cognitive Extension*. New York: Oxford University Press.

- Clark, A. (2010a). Coupling, Constitution, and the Cognitive Kind: A Reply to Adams and Aizawa. In R. Menary (Hrsg.), *The Extended Mind* (S. 81–100). Cambridge: The MIT Press.
- Clark, A. (2010b). Memento's Revenge: The Extended Mind, Extended. In R. Menary (Hrsg.), *The Extended Mind* (S. 43–66). Cambridge: The MIT Press.
- Clark, A. (2014). *Mindware. An Introduction to the Philosophy of Cognitive Science* (2. Aufl.). New York: Oxford University Press.
- Clark, A. & Chalmers, D. J. (1998). The Extended Mind. *Analysis*, 58(1), 7–19.
- Collins, J. (2011). Chomsky. In B. Lee (Hrsg.), *Philosophy of Language: The Key Thinkers* (S. 159–178). London: Continuum International Publishing Group.
- Crane, T. (2003). *The Mechanical Mind. A Philosophical Introduction to Minds, Machines and Mental Representation* (2. Aufl.). New York: Routledge.
- Davis, M. (2011). *The Universal Computer. The Road from Leibniz to Turing* (2. Aufl.). Boca Raton: CRC Press.
- Figdor, C. (2019). The Rise of Cognitive Science in the 20th Century. In A. Kind (Hrsg.), *Philosophy of Mind in the Twentieth and Twenty-First Centuries. The History of the Philosophy of Mind, Volume 6* (S. 280–302). Abingdon: Routledge.
- Fingerhut, J., Hufendiek, R. & Wild, M. (2013). Einleitung. In J. Fingerhut, R. Hufendiek & M. Wild (Hrsg.), *Philosophie der Verkörperung. Grundlagentexte zu einer aktuellen Debatte* (S. 9–102). Berlin: Suhrkamp.
- Freidin, R. (2013). Noam Chomsky's Contribution to Linguistics: A Sketch. In K. Allan (Hrsg.), *The Oxford Handbook of the History of Linguistics* (S. 439–468). Oxford: Oxford University Press.

- Gallagher, S. (2009). Philosophical Antecedents of Situated Cognition. In P. Robbins & M. Aydede (Hrsg.), *The Cambridge Handbook of Situated Cognition* (S. 35–51). Cambridge: Cambridge University Press.
- Garnham, A. (2009). Cognitivism. In J. Symons & P. Calvo (Hrsg.), *The Routledge Companion to Philosophy of Psychology* (S. 99–110). Abingdon: Routledge.
- Heil, J. (2013). *Philosophy of Mind. A Contemporary Introduction* (3. Aufl.). New York: Routledge.
- Lyre, H. (2010). Erweiterte Kognition und mentaler Externalismus. *Zeitschrift für philosophische Forschung*, 64(2), 190–215.
- Menary, R. (2010). Introduction: The Extended Mind in Focus. In R. Menary (Hrsg.), *The Extended Mind* (S. 1–26). Cambridge: The MIT Press.
- Newen, A., Bruin, L. D. & Gallagher, S. (2018). 4E Cognition: Historical Roots, Key Concepts, and Central Issues. In A. Newen, L. D. Bruin & S. Gallagher (Hrsg.), *The Oxford Handbook of 4E Cognition* (S. 3–18). Oxford: Oxford University Press.
- Rescorla, M. (2017). *The Computational Theory of Mind*. In E. N. Zalta (Hrsg.), *The Stanford Encyclopedia of Philosophy* (Spring 2017 Edition). Zugriff unter <https://plato.stanford.edu/archives/spr2017/entries/computational-mind/>. (letzter Zugriff: 31. Januar 2020)
- Rescorla, M. (2019). *The Language of Thought Hypothesis*. In E. N. Zalta (Hrsg.), *The Stanford Encyclopedia of Philosophy* (Summer 2019 Edition). Zugriff unter <https://plato.stanford.edu/entries/language-thought/>. (letzter Zugriff: 31. Januar 2020)
- Rohde, M. (2013). Evolutionäre Robotik, organic computing und Künstliches Leben. In A. Stephan & S. Walter (Hrsg.), *Handbuch Kognitionswissenschaft* (S. 180–183). Stuttgart: J.B. Metzler.

- Rupert, R. D. [Robert D.]. (2004). Challenges to the Hypothesis of Extended Cognition. *The Journal of Philosophy*, 101(8), 389–428.
- Rupert, R. D. [Robert D.]. (2010). Representation in Extended Cognitive Systems: Does the Scaffolding of Language Extend the Mind? In R. Menary (Hrsg.), *The Extended Mind* (S. 325–354). Cambridge: The MIT Press.
- Shapiro, L. (2011). *Embodied Cognition*. New York: Routledge.
- Sprevak, M. (2015). Philosophy of the Psychological and Cognitive Science. In P. Humphreys (Hrsg.), *The Oxford Handbook of Philosophy of Science* (S. 92–114). London: Oxford University Press.
- Sun, R. (2014). Connectionism and Neural Networks. In W. M. R. Keith Frankish (Hrsg.), *The Cambridge Handbook of Artificial Intelligence* (S. 108–127). Cambridge: Cambridge University Press.
- Thagard, P. (2012). Cognitive Architectures. In W. M. R. Keith Frankish (Hrsg.), *The Cambridge Handbook of Cognitive Science* (S. 50–70). Cambridge: Cambridge University Press.
- Thagard, P. (2019). *Cognitive Science*. In E. N. Zalta (Hrsg.), *The Stanford Encyclopedia of Philosophy* (Spring 2019 Edition). Zugriff unter <https://plato.stanford.edu/archives/spr2019/entries/cognitive-science/>. (letzter Zugriff: 31. Januar 2020)
- Walter, S. (2013). Einleitung. In A. Stephan & S. Walter (Hrsg.), *Handbuch Kognitionswissenschaft* (S. 153–156). Stuttgart: J.B. Metzler.
- Wilson, R. A. & Clark, A. (2009). How to Situate Cognition. Letting Nature Take Its Course. In P. Robbins & M. Aydede (Hrsg.), *The Cambridge Handbook of Situated Cognition* (S. 55–77). Cambridge: Cambridge University Press.

## **Zusammenfassung**

Mit der Theorie der erweiterten Kognition hat der Philosoph Andy Clark die These aufgestellt, dass kognitive Prozesse (manchmal) auch nicht-neuronale Ressourcen umfassen können. In diesen Fällen dehnt sich die kognitive Aktivität des Menschen über die eigenen biologischen Grenzen hinweg auf Teile der Umwelt aus und konstituiert dadurch, so Clark, einen erweiterten kognitiven Prozess. Mit der hier vorliegenden Arbeit soll nun der Versuch unternommen werden, möglichst genau darzustellen, was unter der These der erweiterten Kognition (TEK) zu verstehen ist und von welchen Vorstellungen sie sich dabei anleiten lässt. Es wird sich dabei zeigen, dass ihre Grundlagen tief im Gesamtwerk von Andy Clark verwurzelt sind, weshalb ein wesentlicher Teil der Arbeit auch darin bestehen wird, die der Erweiterungsthese zugrundeliegenden theoretischen Anschauungen von Clark klar zu umreißen.