

Antirepräsentationalismus – Kognition ohne Repräsentation?

Magisterarbeit
in der Philosophischen Fakultät I
(Philosophie, Sport und Kunstwissenschaften)

vorgelegt von

Rüdiger Heimgärtner

aus

Undorf

2002

Erstgutachter:

Prof. Dr. phil. Hans Rott

Zweitgutachter:

Dr. Wolfram Hinzen

Inhaltsverzeichnis

Abbildungsverzeichnis.....	IV
Tabellenverzeichnis.....	V
1. Einleitung.....	1
2. Antirepräsentationalismus.....	3
2.1. Was ist Antirepräsentationalismus?.....	3
2.2. Bemerkungen zum historischen Hintergrund des Antirepräsentationalismus.....	3
2.3. Varianten eines kognitionswissenschaftlichen Antirepräsentationalismus.....	5
2.3.1. Dynamische Systeme.....	5
2.3.2. Situated Action.....	7
2.3.3. Embodied Cognition.....	8
2.3.4. Perzeptuelle Symbolsysteme.....	10
2.3.5. Autopoietische Systeme.....	10
3. Gegenstand des Antirepräsentationalismus.....	11
3.1. Definition von Repräsentation.....	11
3.1.1. Der klassische Begriff von ‚Repräsentation‘.....	12
3.1.2. ‚Bedeutung B‘ und ‚Bedeutung $M \rightarrow B$ ‘.....	13
3.1.3. Definition von ‚Repräsentation B‘.....	16
3.1.4. Weitere begriffliche Festlegungen.....	22
3.2. Der Repräsentationsbegriff in der Kognitionswissenschaft.....	24
3.2.1. Symbolische Systeme.....	27
3.2.2. Konnektionistische Systeme.....	34
3.2.3. Dynamische Systeme.....	43
3.3. Weitere Aspekte des Repräsentationsbegriffs.....	52
3.3.1. Gradualität von Repräsentation.....	52
3.3.2. Implementation von Repräsentation.....	53
3.3.3. Computation und Repräsentation.....	54
3.3.4. Implizite und explizite Repräsentation.....	55
3.4. Zusammenfassung.....	59

4. Probleme des Antirepräsentationalismus.....	60
4.1. Methodologische Probleme.....	60
4.1.1. Grundsätzliche Probleme.....	60
4.1.2. Unterschiedliche Beschreibungsebenen.....	62
4.1.3. Low-High-Level-Cognition-Gap.....	66
4.1.4. Systemtheoretische Probleme.....	67
4.2. Inhaltliche Probleme.....	69
4.2.1. Kognition ohne Repräsentation.....	69
4.2.2. Repräsentation und Sprache.....	76
4.3. Ontologische Probleme.....	79
4.3.1. Repräsentation und Bedeutung.....	79
4.3.2. Repräsentation und Intentionalität.....	80
4.3.3. Emergenz oder nicht-reduktiver Physikalismus.....	82
4.3.4. Zusammenführung antirepräsentationalistischer und klassischer Ansätze.....	84
4.4. Zusammenfassung.....	86
5. Schlussfolgerungen.....	88
Literaturverzeichnis.....	94

Abbildungsverzeichnis

Abbildung 2.3.1:	Schematische Darstellung des Dampfdruckreglers von James Watt.....	6
Abbildung 2.3.3:	Mobot „Herbert“.....	9
Abbildung 3.1.1:	Dreiecksbeziehung der klassischen Vorstellung von Repräsentation.....	13
Abbildung 3.1.3:	Vertikale und horizontale Relation.....	17
Abbildung 3.2.1:	Die drei wichtigsten Paradigmen in der Kognitionswissenschaft.....	25
Abbildung 3.2.1.1:	Versuchsaufbau für die XOR-Funktion.....	31
Abbildung 3.2.1.2:	Symbolische Repräsentation (Implementierung der XOR-Funktion über eine Programmiersprache).....	31
Abbildung 3.2.2.1:	Versuchsaufbau mit einem neuronalen Netz (Multilayer-Perceptron: Feedforward-Netzwerk mit drei Neuronenschichten).....	37
Abbildung 3.2.2.2:	Subsymbolische Repräsentation (Die vier Zustände eines trainierten neuronalen Netzes für die XOR – Funktionalität).....	41
Abbildung 3.2.3.1:	Linearer gedämpfter fremderregter Schwinger (aus Argyris 1995).....	45
Abbildung 3.2.3.2:	Ein- und Ausgangssignale eines neurodynamischen Netzes (Elman-Netz), das die Semantik der XOR-Funktion repräsentiert.....	50
Abbildung 4.1.2.1:	Prinzip der Entscheidungsfindung.....	65
Abbildung 5.7.1:	Einordnung der kognitionswissenschaftlichen Ansätze unter Einbeziehung der Ergebnisse dieser Arbeit.....	93

Tabellenverzeichnis

Tabelle 2.3.2:	Unterschiede zwischen einem Watt-Regler und einem klassischen computational-symbolischen Regler (nach vanGelder 1995).....	6
Tabelle 3.1.3:	Gegenüberstellung der Eigenschaften von klassischer Repräsentation, Präsentation und Repräsentation B.....	21
Tabelle 3.2.1.1:	Semantik der XOR-Funktion.....	30
Tabelle 3.2.3.1:	„Differences among kinds of systems“ (aus vanGelder 1997).....	48
Tabelle 3.3.4.1:	Erklärungsebenen und Repräsentationsarten.....	57

1 EINLEITUNG

Diese Arbeit soll einen Eindruck von einem Antirepräsentationalismus vermitteln, wie er heute in der Kognitionswissenschaft vertreten wird. Dabei kann nur ein Überblick gegeben, aber nicht das gesamte Umfeld dargestellt werden. Vielmehr ist beabsichtigt, dass dem Leser durch diese Arbeit ermöglicht wird, ein „Gespür“ für den Antirepräsentationalismus, dessen konzeptuellen Inhalt sowie den daraus resultierenden Problemen und Implikationen zu entwickeln. Dabei werden einige wichtige Probleme des Antirepräsentationalismus diskutiert und Lösungsversuche dafür skizziert.

Ich werde auch versuchen zu zeigen, dass sich Kognitionswissenschaftler von einem Repräsentationalismus distanzieren, ein radikaler Antirepräsentationalismus (AR) aber nicht haltbar ist, da (menschliche) Kognition ohne einen gewissen „Aspekt“ von Repräsentation nicht möglich ist.

In Kapitel 2 wird zunächst näher erläutert, was Antirepräsentationalismus ist. Danach folgen Bemerkungen zum historischen Hintergrund des Antirepräsentationalismus und schliesslich werden die wichtigsten antirepräsentationalistischen Ansätze aus der Kognitionswissenschaft vorgestellt.

Ein antirepräsentationalistisch denkender Kognitionswissenschaftler¹ geht davon aus, dass kognitive Agenten ohne „innere, mentale Repräsentationen“ auskommen. Was sind aber innere, mentale Repräsentationen?

Um einer Antwort auf diese Frage näher zu kommen, wird auf Begriffe wie ‚Zeichen‘, ‚Symbol‘, ‚Referenz‘ und ‚Computation‘ eingegangen. Auch Begriffe wie ‚Intentionalität‘, ‚Bedeutung‘, ‚Emergenz‘ oder ‚Kontinuität‘ müssen berücksichtigt werden. Allerdings wird keine letztgültige Klärung dieser Begriffe angestrebt. Vielmehr geht es darum, diese Begriffe bezüglich eines Antirepräsentationalismus innerhalb der Kognitionswissenschaft in kohärenter Weise zueinander in Beziehung zu setzen.

Ob jemand Antirepräsentationalist ist oder nicht, hängt davon ab, welchen Begriff von Repräsentation man zugrunde legt. Es wird daher in Kapitel 3 versucht, einen Referenzbegriff von Repräsentation zu entwickeln, anhand dessen die Repräsentationalität der verschiedenen antirepräsentationalistischen Ansätze „gemessen“ werden kann. Dies geschieht sowohl aus philosophischer als auch aus

¹ Die feminine Wortform ist immer mitgemeint.

kognitionswissenschaftlicher Sicht anhand der wichtigsten systemtheoretischen Paradigmen in der Kognitionswissenschaft (Symbolische Systeme, Subsymbolische Systeme und Dynamische Systeme). Dabei wird auch demonstriert, wie der Repräsentationsbegriff von Kognitionswissenschaftlern verwendet wird.

Anschliessend muss näher bestimmt werden, was *Antirepräsentationalisten* unter Repräsentation verstehen. Denn ein von Antirepräsentationalisten anders definierter Begriff von Repräsentation oder eine unterschiedliche Verwendung dieses Begriffes könnte zu einer anderen Auffassung von Repräsentationalismus und damit von Antirepräsentationalismus führen und unterschiedliche Konsequenzen nach sich ziehen. Die Frage, ob kognitive Agenten ohne Repräsentationen auskommen oder nicht, würde u. U. anders entschieden. Anhand des eingeführten Referenzbegriffes kann geklärt werden, gegen welche Art von Repräsentation der Antirepräsentationalist eigentlich argumentiert.

Anschliessend kann in Kapitel 4 eine Analyse der antirepräsentationalistischen Argumente folgen und die Probleme des Antirepräsentationalismus können aufgezeigt und kritisch beleuchtet werden. Es wird untersucht, ob und wie die Probleme des Antirepräsentationalismus wie Memorierung, physikalische Unverbundenheit, High-Level-Cognition, Planung und Entscheidung einer Lösung durch den Ansatz dynamischer Systeme zugeführt werden können und welche Implikationen sich daraus ergeben. Dabei wird herausgearbeitet, inwiefern diese Ansätze im Hinblick auf den anfangs definierten Referenzbegriff von Repräsentation als antirepräsentationalistisch gelten können und somit Kognition erklären können, ohne dabei Repräsentation zu verwenden.

Am Ende der Kapitel 3 und 4 werden die Ergebnisse der jeweiligen Untersuchungen zusammengefasst und abschliessend in Kapitel 5 die Schlussfolgerungen daraus gezogen.

2 ANTIREPRÄSENTATIONALISMUS

2.1 Was ist Antirepräsentationalismus?

Antirepräsentationalismus ist das Gegenteil von Repräsentationalismus. Innerhalb des Repräsentationalismus wird unterstellt, dass die Erkenntnis materieller Dinge nur über sie vertretende mentale Objekte möglich ist: Keine Erkenntnis ohne Repräsentation. Ein Antirepräsentationalist hingegen vertritt die These, dass Erkenntnis über die Welt ohne interne Repräsentation von Dingen in der Welt möglich ist (vgl. Stufflebeam 1998a).

Kognitionswissenschaftler erforschen die kognitiven Prozesse, die Erkenntnis ermöglichen. Kognition umfasst alle Prozesse, die nötig sind, damit ein Lebewesen sich und andere Objekte in der Welt erkennen kann (z.B. Wahrnehmen, Erkennen, Vorstellen, Urteilen, Lernen, Erinnern, Denken, Vermuten, Erwarten, Planen und Problemlösen). Kognition ist auch die Fähigkeit des Menschen, sich gegenüber der Welt (d.h. seiner Umwelt) adaptiv zu verhalten und in ihr zurechtzukommen.

Alle klassischen Ansätze der Kognitionswissenschaft halten dabei an der Kognitivismusthese fest, die besagt, dass für Kognition interne Repräsentation notwendig ist. Antirepräsentationalistische Ansätze in der Kognitionswissenschaft behaupten hingegen, dass Kognition auch ohne interne Repräsentation möglich ist. Diese Haltung gegen Repräsentation kann auf zwei Arten eingenommen werden. Einerseits kann man leugnen, dass es „Repräsentation“ gibt und annehmen, dass kein kognitiver Agent über Repräsentationen verfügt, so dass Repräsentation für Kognition nicht wesentlich sein kann. Andererseits kann man behaupten, dass alle Erklärungsmodelle von Kognition auf den Begriff „Repräsentation“ verzichten können (vgl. Chemero 1999).

Bevor in Kapitel 3 untersucht wird, ob und wie kognitive Prozesse ohne Repräsentation möglich oder erklärbar sind, werden in diesem Kapitel kurz die antirepräsentationalistischen Ansätze der Kognitionswissenschaft vorgestellt.

2.2 Bemerkungen zum historischen Hintergrund des Antirepräsentationalismus

Ein Antirepräsentationalismus kann erst als Gegenbewegung zu einem Repräsentationalismus entstehen. Als Begründer des Repräsentationalismus gilt René Descartes: Mit dem Begriff des ‚cogito‘ (‚ich denke‘) geschieht eine Aufteilung in

Subjekt (Ich) und Objekt (Welt). Der Bezug zwischen Ich und Welt wird im Geist des Subjekts hergestellt durch Ideen, die Dinge der Welt repräsentieren. Denken geschieht über interne Repräsentationen (vgl. Descartes 1972). Denken im Sinne Descartes könnte man heute als „Kognition“ bezeichnen.

Antirepräsentationalistisch eingestellte Philosophen wie Heidegger (Existenzialismus), Rorty (Pragmatismus) und Quine (Naturalismus) sind gegen die Auffassung, dass Erkenntnis über die Welt mittels interner Repräsentation zustande kommt (vgl. Boros 1999).

Martin Heidegger versucht Descartes' Cogito zu invertieren. Er geht nicht vom Subjekt aus, in dem die Welt repräsentiert wird, sondern von den alltäglichen Fähigkeiten und Praktiken, die unser In-der-Welt-sein ausmachen und verneint damit eine Repräsentationsbeziehung zwischen Geist und Welt. Zum Wesen der Kognition gehört nicht Repräsentation als das Sich-auf-die-Welt-beziehen, sondern das In-der-Welt-sein (vgl. Heidegger 1927).

Richard Rorty argumentiert gegen Repräsentation, weil es für ihn eine ontologische Verwobenheit zwischen Form und Inhalt gibt und dadurch deren Separierung unmöglich ist. Die Welt spiegelt sich nicht mental im Kopf wieder. Vielmehr ist das Subjekt mit der Welt in kausaler Interaktion verwoben. In jeder Behauptung sind subjektive und objektive Anteile vermischt, so dass der Inhalt eines Gedankens nicht in die Anteile des Subjekts und jene Anteile, die von der Aussenwelt stammen, getrennt werden können (vgl. Rorty 1979).

W. O. Quine überträgt diesen Gedanken auf Kants Unterscheidung von analytischen und synthetischen Begriffen. Für Quine gibt es keinen Unterschied zwischen subjektiven und objektiven Wissenskonstituenten (vgl. Quine 1960).

Hubert L. Dreyfus postuliert Intelligenz ohne Repräsentation in Anlehnung an Merleau-Pontys Kritik an mentalen Repräsentationen (vgl. Dreyfus 1998). Dessen Theorien des „Intentional Arc“ und „Maximal Grip“ sollen erklären, wie man auch ohne mentale Repräsentationen mittels seiner Fähigkeiten in der Welt zurechtkommt (vgl. Merleau-Ponty ⁴1967). Durch die enge Verbindung zwischen dem Agenten und der Welt (= „intentionaler Bogen“) werden dessen erworbene Fähigkeiten nicht als Repräsentationen im Geist gespeichert, sondern als Dispositionen, auf die Herausforderungen der Situationen der Welt zu antworten. Dies wird unterstützt durch die Tendenz des Körpers, derart auf die Situationen der Welt zu reagieren, dass die gegenwärtige Situation der Vorstellung einer optimalen Gestalt im Agenten näher kommt (= „maximales Verständnis“).

Diese antirepräsentationalistischen Grundgedanken wirken auf die jüngsten Vertreter des Antirepräsentationalismus wie Maturana & Varela (1987), Freeman & Skarda (1990), Brooks (1991), Van Gelder (1995), Stufflebeam (1998b), Chemero (1999), Keijzer (1998) und Jäger (1996) fort (vgl. Clark 2001). Im Folgenden werden einige dieser Positionen vorgestellt.

2.3 Varianten eines kognitionswissenschaftlichen Antirepräsentationalismus

Die bekanntesten Varianten antirepräsentationalistischer Ansätze innerhalb der Kognitionswissenschaft sind dynamische Systeme, situative Agenten, embodied embedded cognition, perzeptuelle Symbolsysteme sowie neurobiologische Ansätze (vgl. Markman & Dietrich 2000).

2.3.1 Dynamische Systeme

Timothy van Gelder gibt drei Gründe dafür an, sich von einer repräsentationalen Theorie des Geistes abzuwenden (vgl. Van Gelder 1995, S. 380). Geist ist nicht nur etwas Inneres, sondern auch etwas Äußeres, mit der Welt Verbundenes. Ferner werden epistemologische Probleme umgangen, indem man menschliche Agenten als wesentlich in eine sich verändernde Welt eingebettet betrachtet. Und schliesslich kann menschliches Verhalten niemals kausal nur mittels kartesischer Repräsentation erklärt werden. Allerdings scheint der letzte Grund etwas schwach zu sein. Denn das menschliche Verhalten muss ja nicht kausal oder alleine über kartesische Repräsentation erklärt werden.

Van Gelder postuliert die dynamische Hypothese, welche besagt, dass kognitive Agenten dynamische Systeme sind und postuliert, dass dynamische Systeme nichts repräsentieren. Dies versucht er anhand des von James Watt konstruierten Dampfdruckreglers (Watt governor) zu demonstrieren (Abbildung 2.3.1).

Dieser rein mechanische Regler bewirkt, dass der Druck des ausströmenden Dampfes immer gleich gross bleibt. Er besteht aus einer Spindel und zwei mit Gewichten ausgestatteten Armen. Der ausströmende Dampf treibt die Spindel an. Je mehr Dampf ausströmt, desto schneller dreht sich die Spindel und desto weiter bewegen sich die Arme von der Spindel weg (durch die Fliehkraft der Gewichte), wodurch das Dampfauslassventil weiter geschlossen wird. Dadurch strömt weniger Dampf aus und die Spindel dreht sich langsamer, wodurch sich die Arme wieder näher zur Spindel

bewegen und damit das Dampfauslassventil weiter geöffnet wird und mehr Dampf ausströmt. Somit wird der Dampfdruck konstant gehalten.

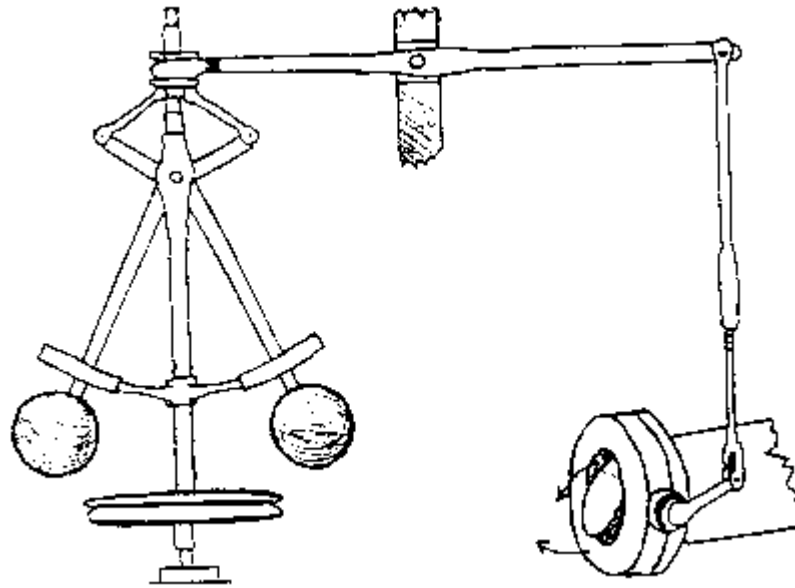


Abbildung 2.3.1: Schematische Darstellung des Dampfdruckreglers von James Watt

Man könnte diesen Regelvorgang algorithmisch beschreiben und als Programm in einem Computer implementieren. Aber genau das ist beim mechanischen Watt-Regler nicht notwendig. Dieser funktioniert auch ohne Repräsentationen und ist deshalb nicht repräsentational. Und wenn man wie van Gelder Repräsentationalität für Computationalität voraussetzt (also Computation als Manipulation von Symbolen versteht), dann ist der Watt-Regler auch nicht computational. Es gibt keine einzelnen diskreten „Schritte der Verarbeitung“ wie bei sequentiell-diskreten (symbolisch-digitalen) Systemen. Dynamische Systeme arbeiten kontinuierlich. Tabelle 2.3.2 stellt die Unterschiede zwischen einem Watt-Regler und einem klassischen computational-symbolischen Regler gegenüber (nach van Gelder 1995).

Kriterium	Computationaler Regler	Watt-Regler
Repräsentationalität	repräsentational	nicht-repräsentational
Computationalität	computational	nicht-computational
Prozessauflösung	sequentiell, diskret	nicht-sequentiell, kontinuierlich
Verarbeitungsart	zyklisch	nicht-zyklisch, simultan
Interaktionalität	homunkular, d.h. über Botschaften	nicht-homunkular, direkt
Zeitauflösung	natürliche Zahlen (integers)	reale Zahlen (real numbers)

Tabelle 2.3.2: Unterschiede zwischen einem Watt-Regler und einem klassischen computational-symbolischen Regler (nach van Gelder 1995)

Gemäss dem Neurodynamizisten Freeman brauchen Physiologen im Gegensatz zu Philosophen, Informatikern und Kognitionspsychologen keine „Repräsentationen“. Denn Denken in „Repräsentationen“ behindert die Erforschung der biologischen Algorithmen des Gehirns. „Repräsentationen“ sind für die Beschreibung und das Verstehen der Gehirndynamik nicht nötig. Wiederabruf (Retrieval) von z.B. Geruchserinnerungen ist keine Wiederherstellung eines gespeicherten Zustandes, sondern eine erneute Erzeugung eines (nicht mit dem Original identischen – aber sehr ähnlichen) Gerucheindrucks:

„Our physiological data show that episodic storage of odor trials does not happen, that „retrieval“ is not recovery but re-creation, always with differences, and that stimulus-bound patterns cannot coexist with re-created patterns to support matching procedures.“ (Freeman & Skarda 1990, S. 379)

Wenn man nicht fordert, dass der Inhalt einer Repräsentation über die Zeit hinweg identisch bleiben muss, dann könnte man auch Freemans „re-creation“ als Repräsentation bezeichnen. Dies wird in Kapitel 3 näher untersucht.

2.3.2 Situated action

Ansätze der situierten Agenten gehen davon aus, dass Kognition nicht ohne die Einflüsse der Umgebung auskommt (vgl. Clancey 1997, Vera & Simon 1993, Lai-Chong Law 1993). Kognition ist immer kontext- bzw. situationsabhängig. Wahrnehmen und Schlussfolgern ist ohne Handeln nicht möglich. Durch diese permanente Interaktion mit der Welt verändert sich das Wissen des Agenten ständig. Es ist kein statisches, gespeichertes Wissen, sondern ein dynamisches. Auf diese Weise wird Wissen auch in der Interaktion mit der Umwelt aus dieser „aufgenommen“. Es ist daher vieles in der Umgebung repräsentiert und braucht nicht mehr im Kopf repräsentiert zu sein. Die Welt muss daher nicht extensiv repräsentiert werden. Prognosen lassen sich leichter über direkte Information aus der Umwelt machen als über rein abstraktes Denken. Es müssen nur relevante, auf den Agenten bezogene Dinge der Welt berücksichtigt werden und nicht alle möglichen Fakten über die Welt, was eine Auflösung des Frameproblems bedeuten könnte (vgl. French & Anselme 1999).²

² Das Frameproblem bezeichnet das Problem eines kognitiven Systems, zu entscheiden, welche Information unverändert bleiben muss, nachdem das System eine bestimmte Handlung ausgeführt hat. Wenn sich ein Mann mit Hut von Punkt A nach Punkt B bewegt, so muss das System diese Bewegung erkennen, aber die meisten Informationen über den Mann gleich lassen, d.h. es muss auch den Hut mit dem Mann bewegen bzw. die Information, dass der Hut sich auf dem Kopf des Mannes befindet, unverändert lassen.

„Pengi“, ein am Computer simulierter autonomer Agent, muss Bienen und Holzkisten ausweichen (vgl. Agre & Chapman 1987). Pengi hat aber weder eine Vorstellung vom Spielfeld, auf dem er sich befindet, noch von der Anordnung der Bienen und Kisten auf diesem Spielfeld. Vielmehr verfügt Pengi nur über ein paar grundlegende Fähigkeiten, wie dem Ausweichen einer Biene oder dem Wegschieben einer Kiste. So muss der Agent nur das Problem lösen, das gerade unmittelbar durch die Umwelt (Situation) vorgegeben ist. Das Frameproblem verschwindet proportional zum Zuwachs dynamisch erzeugter kontextabhängiger Repräsentationen. Der Agent hat auf diese Weise immer weniger Probleme mit fehlenden oder irrelevanten Informationen, denn je mehr „Umweltbezug“ in die Generierung einer Repräsentation mit eingeht, desto domänenunabhängiger (oder weniger lückenhaft) ist diese.³

Allerdings kann Pengi nur Probleme lösen, die seine unmittelbare Umgebung betreffen. Innerhalb eines solchen Ansatzes (wie in allen Ansätzen, die mit sehr wenigen oder gar keinen Repräsentationen auszukommen versuchen) ist es schwer, höher kognitive Fähigkeiten wie kontrafaktisches Schliessen, abstraktes Denken, Memorierung oder das Erreichen eines (sowohl zeitlich als auch räumlich entfernten) Ziels zu erklären. Auf dieses Problem werde ich in Kapitel 4 ausführlicher eingehen.

2.3.3 Embodied Cognition

Dieser Ansatz ist verwandt mit dem der Situated action. Allerdings geht er über diesen hinaus, indem er behauptet, dass es notwendig ist, Agenten zu realisieren, die tatsächlich mit der realen Umgebung interagieren. Die Umgebung kann zur Lösung schwieriger Probleme herangezogen werden. Kognition benötigt die ständige Einbeziehung der Wahrnehmungen und Handlungen des Agenten. Durch Ausnutzen dieser sensomotorischen Koordination können interne Repräsentationen erheblich reduziert werden oder ganz entfallen. Vertreter von Embodied *Embedded* Cognition (EEC) behaupten darüber hinaus, dass für Kognition auch der Gebrauch externer Hilfsmittel (wie z.B. Notizen, Taschenrechner, Erinnern an schon Gewusstes durch Versetzen in eine ähnliche Situation usw.) notwendig ist (vgl. Brooks 1991). In der Robotik werden situative Agenten verwirklicht, die über „embodied cognition“ verfügen.

³ Eine Domäne ist ein Modell, das einen sehr kleinen Ausschnitt der Welt beschreibt. Je mehr „Welt“ eine Repräsentation enthält, desto unabhängiger ist sie von Domänen.



Abbildung 2.3.3: Mobot „Herbert“

Abbildung 2.3.3 zeigt den Mobot (mobilen Roboter) „Herbert“, der Cola-Dosen aufammelt. Kommt er wirklich ganz ohne Repräsentationen aus? Der Entwickler dieses Roboters, der Robotik-Forscher Rodney Brooks, lehnt jedenfalls jegliche Notwendigkeit von Repräsentation für die Kognition seiner „Kreaturen“ ab, weil sie zu stark von „Standardrepräsentationen“ abweichen:

„There need be no explicit representation on goals that some central (or distributed) process selects from to decide what is most appropriate for the Creature to do next. [...] [T]here need be no explicit representation of either the world or the intentions of the system to generate intelligent behaviours for a Creature. [...] Even at a local, level we do not have traditional AI representations. We never use tokens which have any semantics that can be attached to them. [...] An extremist might say that we really do have representations, but that they are just implicit. [...] However we are not happy with calling such things a representation. They differ from standard representations in too many ways.“ (Brooks 1997, S. 405f)

Mobots verfügen nicht über explizite Repräsentationen, da sie weder über Variablen noch über Regeln verfügen:

„There are no variables that need instantiation in reasoning processes. [...] There are no rules which need to be selected through pattern matching. There are no choices to be made. To a large extent the state of the world determines the action of the Creature.“ (Brooks 1997, S. 406)

Die Komplexität des Verhaltens ist nicht notwendigerweise der Komplexität der Kreaturen inhärent, sondern möglicherweise der Umwelt.

„We hypothesize (following Agre and Chapman) that much of even human level activity is similarly a reflection of the world through very simple mechanisms without detailed representations.“ (Brooks 1997, S.407)

Kognition könnte also eine Reflektion der Welt sein, ohne detaillierte Repräsentation. Brooks erläutert allerdings nicht, was detaillierte Repräsentation bedeutet.

2.3.4 Perzeptuelle Symbolsysteme

In klassischen symbolischen Systemen werden Repräsentationen auf unterschiedlich abstrakten Ebenen angenommen. Einige Repräsentationen korrespondieren direkt mit perzeptueller Erfahrung. Amodale Repräsentationen hingegen beziehen sich auf abstrakte Konzepte wie ‚Wahrheit‘ oder ‚Gerechtigkeit‘, und haben nichts mit perzeptueller Erfahrung zu tun. Allerdings sind letztere problematisch, weil sie aufgrund dessen die potentielle Vielfalt einfacher Ereignisse nicht einfangen können. In perzeptuellen Symbolsystemen werden Objekte und Ereignisse hingegen mittels des perzeptuellen Systems (Wahrnehmungssystems) simuliert, wodurch auf abstrakte Repräsentationen verzichtet werden kann (vgl. Barsalou 1999). Kategorisierung z.B. wird auf diese Weise durch Prozesse perzeptueller Merkmalsgenerierung ermöglicht. Es scheint also, dass Repräsentationen innerhalb des Ansatzes perzeptueller Symbolsysteme nicht abgeschafft werden, sondern anders gesehen werden – nämlich perzeptiv und weniger abstrakt.

2.3.5 Autopoietische Systeme

Die Neurobiologen Maturana & Varela (1987) behaupten, dass das Nervensystems nicht repräsentationalistisch ist. Es verfügt nicht über Ein- und Ausgänge, wie dies bei einem Computer der Fall ist, und denen einfach eine Bedeutung vom Entwickler zugeordnet werden kann. Vielmehr ist das Nervensystem über die Zeit hinweg dynamisch entstanden:

„Die Arbeitsweise des Nervensystems ist auch nicht repräsentationalistisch, da der strukturelle Zustand des Nervensystems bei jeder Interaktion spezifiziert, welche Perturbationen möglich sind und welche Veränderungen diese in seiner Dynamik von Zuständen auslösen.“ (Maturana & Varela 1987, S. 185).

Maturana & Varela bezeichnen Lebewesen als autopoietische Systeme, d.h. durch interne Relationen definierte Einheiten, die ihre Organisation und Struktur selbst aufrechterhalten. Durch die Interaktion mit der Umwelt wird die strukturelle Dynamik eines autopoietischen Systems gestört. Das autopoietische System versucht seinen Zustand aber aufrechtzuerhalten und kompensiert diese Perturbation (geringfügige Störung des Systems von aussen). Durch seine Struktur legt das System zu jedem Zeitpunkt fest, auf welche Perturbationen es reagiert und wie sich diese auf die Abfolge der Zustände des Systems auswirken. Ein autopoietisches System „lebt“, solange es sich selbst aufrechterhalten kann.

3 GEGENSTAND DES ANTIREPRÄSENTATIONALISMUS

Damit die Argumente des Antirepräsentationalismus gegen den Repräsentationalismus (Kognitivismus) greifen können, muss bekannt sein, wogegen sie sich genau richten. Es muss also geklärt werden, was Antirepräsentationalisten unter „(innerer/mentaler) Repräsentation“ verstehen. Um den Repräsentationsbegriff dabei besser fassen zu können und Unterschiede in dessen Verwendung bzw. Bedeutung herausarbeiten zu können, werden der klassische, konnektionistische und der dynamische Ansatz vorgestellt und auf diese Weise die wichtigsten historischen Veränderungen des Begriffes in der Kognitionswissenschaft erfasst.

Sowohl das theoretische Interesse an Repräsentation innerhalb der Philosophie als auch das praktische Interesse an Repräsentation innerhalb der Kognitionswissenschaft wird im Folgenden aufgezeigt. Allerdings wird das Hauptgewicht auf der Untersuchung des Repräsentationsbegriffs liegen, wie er in den Kognitionswissenschaften verwendet wird.

3.1 Definition von Repräsentation

Ein so abstrakter Begriff wie „Repräsentation“ ist sehr facettenreich und lässt sich schwer eindeutig definieren. Je nach Verwendungszweck und –bereich hat dieser Begriff unterschiedliche Bedeutungen. Verschiedene philosophische Schulen und Methoden sind ebenso dafür verantwortlich wie die sich durch die Technisierung („Computerzeitalter“) und der Spaltungen der Disziplinen (Wahrnehmungspsychologie, Kognitionspsychologie, Linguistik, Künstliche Intelligenz (KI), Philosophie des Geistes) im letzten Jahrhundert neu entwickelten Ansätze innerhalb von Psychologie und Kognitionswissenschaft (vgl. Scheerer 1992b).

Robert Cummins nennt vier Problembereiche mentaler Repräsentationen (vgl. Cummins 1996, S. 1). Zunächst muss definiert werden, was es für ein Ding heisst, ein anderes zu repräsentieren. Anschliessend kann untersucht werden, welche Inhalte mental repräsentiert werden, welche Form mentale Repräsentationen aufweisen und wie diese im Gehirn implementiert sind. Während die Definition von Repräsentation eine philosophische Frage darstellt, ist die Frage nach Inhalt, Form und Implementierung mentaler Repräsentationen eine empirische Frage. Das philosophische Interesse liegt in der Klärung der Relation zwischen Repräsentationen und dem, wovon sie

Repräsentationen sind (meist ist hier von „Repräsentation“ im Singular die Rede). Kognitionswissenschaftler versuchen herauszufinden, welche Zustände oder Objekte der Geist oder kognitive Systeme repräsentieren (meist ist hier die Rede von „Repräsentationen“ im Plural).

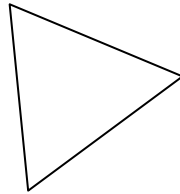
Im Folgenden werden sowohl kognitionswissenschaftliche Fragen angesprochen als auch auf die damit verbundenen philosophischen Probleme eingegangen. Es wird versucht zu klären, was „Repräsentation“ oder „Repräsentationen“ sind. Diese Begriffe werden von verschiedenen Seiten beleuchtet, um sich ihrer Verwendungsweise innerhalb des Antirepräsentationalismus anzunähern. Dazu werden einige gebräuchliche Kriterien des Begriffes hinterfragt. Es wird auch versucht, die Aspekte der Form, des Inhalts und der Implementierung von „Repräsentationen“ mit einzubeziehen.

3.1.1 Der klassische Begriff von ‚Repräsentation‘

In der Philosophie und Psychologie werden unter Repräsentationen meist Vorstellungen (mentale Zustände mit kognitivem Gehalt), Darstellungen (strukturenhaltende Abbildungen) oder Stellvertretungen verstanden (vgl. Scheerer 1992a, S. 790).

Die weiteste Lesart von Repräsentation in der Kognitionswissenschaft ist die der „Stellvertretung“ (vgl. Scheerer 1992b, S. 844). Allerdings muss diese Lesart präzisiert werden, indem spezifiziert wird, wie ein innerer Zustand als Repräsentation fungiert. Nach Cummins (1989) sind innere Zustände dann Repräsentationen, wenn sie von äusseren Zuständen kausal abhängig sind oder mit ihnen kovariieren. In der Kognitionswissenschaft werden solche inneren Zustände oft mit Symbolen gleichgesetzt. Die klassische Vorstellung von Repräsentation eines Symbols geht von einer Dreiecksbeziehung aus (vgl. Eco 1972, S. 69f). Abbildung 3.1.1 zeigt einige Begriffe verschiedener Autoren hierzu (vgl. Frege 1892 und Eco 1972). Nach Eco sind aber mehr als diese drei Grössen in einem Repräsentationsverhältnis verwickelt, wonach das Dreieck einem komplexen Polyeder weichen müsste.

Gedanke (Bedeutung, Repräsentationsinhalt, Referenz (Eco, Odgen), Sinn (Frege))



Gegenstand (Bezeichnetes, Referens (Eco),
Referent (Odgen), Bedeutung (Frege))



Gehirn (Zeichen, Signifikans (Eco), Symbol (Odgen, Frege))

Abbildung 3.1.1: Dreiecksbeziehung der klassischen Vorstellung von Repräsentation

Der Begriff ‚Repräsentation‘ ist analytisch relational. Gewöhnlich wird Repräsentation als eine zweistellige Relation aufgefasst: Etwas steht für etwas anderes (z.B. A steht für B). Es wird eine Beziehung zwischen Etwas (einer Entität A) und dem wofür dieses Etwas steht (einer von A verschiedene Entität B) angenommen (vgl. Keller 1995).

Für den späten Wittgenstein ergibt sich die Bedeutung (d.h. der Inhalt der Repräsentation) aus ihrem Gebrauch (besser: *ist* ihr Gebrauch (vgl. Wittgenstein¹² 1999, §117, §560)). Es wird dabei nichts *repräsentiert*. Man könnte demnach im Gegensatz zum frühen Wittgenstein den späten Wittgenstein auch als Antirepräsentationalisten bezeichnen.

Im Folgenden werde ich allerdings von einem klassischen Repräsentationsbegriff abweichen und für einen etwas anderen Repräsentationsbegriff argumentieren, der aufgrund seiner Definition als „eher antirepräsentationalistisch“ bezeichnet werden könnte und damit neueren, vor allem neurodynamischen Ansätzen in der Kognitionswissenschaft entgegen kommt.

3.1.2 ‚Bedeutung B‘ und ‚Bedeutung M→B‘

Durch das Sein eines intentionalen Wesens in der Welt und dessen Interaktion mit der Welt ergeben sich kausal Zustände im Gehirn, die sich situationsabhängig später wieder durch interne und externe kausale Triggerung einstellen können. Man könnte diese reproduzierten Zustände „Repräsentationen“ nennen, ohne dass es dazu ein

intentionales Wesen geben muss, das diesen Zuständen aktiv Bedeutung zuordnen müsste (implizite Repräsentation).

Um die Plausibilität der Verwendung dieses Repräsentationsbegriffes für die an dieser Stelle noch als „antirepräsentationalistisch“ bezeichneten Ansätze zu begründen, muss ich etwas ausholen.

Für meine Argumentation setze ich voraus, dass (i) es keine ontologische Identität gibt, (ii) alles Seiende miteinander interagiert und (iii) „Bedeutung“ bereits mit jeglichem strukturiertem Seienden durch dessen gegenseitige Interaktivität „vorhanden“ ist.⁴ Sie braucht nur noch (von einem intentionalen Wesen) abgelesen bzw. umgesetzt zu werden. Es können m. E. dann folgende zwei Arten von „Bedeutung“ unterschieden werden:

1. ‚Bedeutung B‘ (strukturinhärente Bedeutung): Ein System hat immer seinen Zweck (und wenn es nur der ist, sich selbst als Systemstruktur aufrecht zu erhalten), welcher für ein anderes System durchaus bedeutsam sein kann. Jedes System besteht aus Entitäten und Beziehungen zwischen diesen. Es ergibt sich so eine Systemstruktur. Alles kann als System mit einer Struktur betrachtet werden. Und auch wenn wir nichts als System betrachten, weil wir uns z. B. nicht einigen können, wo wir die Systemgrenzen ziehen, so bleibt dennoch in jedem Fall eine Struktur, eine Anordnung von Seiendem.

Jede Struktur hat eine potentielle Bedeutung, ob man ihr diese nun zuordnet oder nicht und ob man diese erkennt oder nicht. Denn durch die Interaktion der Struktur mit anderem Seienden hat diese zwingend eine ganz bestimmte Bedeutung für dieses andere Seiende: Trifft eine Struktur S1 auf eine andere Struktur S2 (oder kommt in deren Nähe), so gibt es folgende Möglichkeiten: Könnte S1 in S2 kausal eine Veränderung verursachen, dann ist S1 für S2 potentiell bedeutend (PB). PB wird für S2 aktual bedeutend (AB), wenn PB von S2 (bewusst-aktiv) erkannt oder (unbewusst-passiv) „erlitten“ wird, d.h. S2 kausal eine Veränderung erfährt. Legt man zwei Bücher aufeinander, so wird keine der PB der Strukturen der Bücher aktual. Bringt man allerdings ein brennendes Zündholz in die Nähe eines Buches, dann wird aus der PB des Zündholzes eine AB für das Buch, da dessen Struktur durch die Struktur des brennenden Zündholzes kausal verändert wird (verbrennt). AB kommt also nur zustande, wenn ein Prozess (Erkenntnisprozess oder „Erleidens“-prozess) zustande kommt, an dem beide Strukturen beteiligt sind.

⁴ Da es nichts Unstrukturiertes gibt, ist somit alles Seiende davon betroffen. Alles Seiende hat damit inhärent „Bedeutung“.

Ein Erkenntnisprozess erfolgt nur in einem bewussten intentionalen Wesen. Eine Strukturveränderung („Erleidensprozess“) kann aber auch erfolgen, ohne sich darüber bewusst zu werden. Z.B. verändert sich unser Körper innerlich ständig, ohne dass wir uns darüber ständig im Klaren sind. Dennoch ist es für uns sicherlich bedeutend, wenn plötzlich unser Herz aufhört zu schlagen. Es ist also ein intentionales Wesen nötig, das die aktuelle Bedeutung aus der potentiellen Bedeutung einer Struktur erkennt. Das intentionale Wesen erkennt allerdings nicht nur diese aktuelle Bedeutung einer Struktur, sondern vielmehr eine Gesamtbedeutung. Denn das intentionale Wesen selbst stellt auch eine Struktur mit einer strukturinhärenten potentiellen Bedeutung dar. Die erkannte Gesamtbedeutung ergibt sich somit aus der strukturinhärenten aktuellen Bedeutung der zu erkennenden Struktur und der strukturinhärenten aktuellen Bedeutung der Struktur des intentionalen Wesens. Diese Gesamtbedeutung entspricht der intentional zugeordneten Bedeutung $M \rightarrow B$ (s. Pkt. 2). Sie darf nicht mit der strukturinhärenten Bedeutung B verwechselt werden! Bedeutung B ist ungerichtet und relationslos und der Struktur inhärent. Sie könnte als eine Funktion einer Struktur aufgefasst werden.

Da diese Art der Bedeutung in der klassischen Bedeutungstheorie nicht vorkommt, nenne ich sie in dieser Arbeit Bedeutung im Sinne von ‚Bedeutung B‘ oder kurz: Bedeutung B. Bedeutung B ist zwar der Bedeutung im Sinne Wittgensteins sehr ähnlich: Die Bedeutung B einer Struktur ergibt sich aus dem Gebrauch dieser Struktur. Allerdings bezieht sich Wittgensteins Bedeutungstheorie nur auf die Sprache (vgl. Wittgenstein ¹² 1999). Bedeutung B hingegen ergibt sich aus dem Gebrauch irgendeiner Struktur.

2. ‚Bedeutung $M \rightarrow B$ ‘ (intentional zugeordnete Bedeutung): Die Bedeutung, die von einem intentionalen Wesen M einer Struktur mit der Bedeutung B zugeordnet wird. Diese Art der Bedeutung ist relational und gerichtet ($M \rightarrow B$).

Eine grosse Debatte in der KI-Forschung bezog sich darauf, dass aus Struktur bzw. aus Syntax (also einer linguistisch betrachteten Struktur) alleine niemals Semantik *entstehen* kann. Ich denke, dass diese These im Sinne der ‚Bedeutung $M \rightarrow B$ ‘ richtig ist. Allerdings ist ein *Entstehen* einer Bedeutung aus einer Struktur gar nicht notwendig. Aufgrund der kausalen Interaktion mit der Umwelt spielt immer die Situation eine entscheidende Rolle. Semantik im Geiste entstand mit der Evolution des Geistes selbst. Mit der Interaktion von Seiendem ist Bedeutung B bereits überall „vorhanden“. Bedeutung entsteht nicht nur erst durch Konvention oder Referenz, sondern ist den Dingen bereits potentiell „inhärent“ durch ihre Bestimmung in der Evolution. Syntax bedeutet für ein intentionales Wesen etwas, weil es die potentielle Bedeutung der

syntaktischen Struktur erkennt. Der Einwand, dass gleiche Struktur (Syntax) semantisch ambig sein kann, ist zu voreilig. Wenn eine Struktur (Syntax) ambig erscheint, dann nur deshalb, weil Teile dieser Struktur oder die Einbettung dieser Struktur in dessen Umfeld noch unerkannt oder unberücksichtigt sind. Wenn ich von einer Sitzbank gleichweit entfernt bin wie von einem Bankinstitut, dann erscheint der Satz „Ich gehe zur Bank.“ ambig zu sein. Aber aufgrund des Kontexts, in dem dieser Satz geäußert wird, kann diese potentielle Ambiguität aufgelöst werden. Trage ich eine Kleidung, die dem Besuch eines Bankinstitutes gerecht wird, bin ich in Eile, ist die Sitzbank schmutzig oder frisch gestrichen usw., wird man schnell wissen, was mit Bank gemeint ist. Gerade bei einem so komplexen System wie der Sprache kann dessen Struktur nicht in all ihren Facetten immer sofort durchschaut werden und bleibt daher oft semantisch ambig. Aber es spielen nicht nur alle strukturellen Aspekte der Sprache (wie die Einbettung in den Kontext, die Reihenfolge der Sätze etc.) eine Rolle, sondern auch alle nichtsprachlichen Aspekte (wie Stimmung, Umgebung usw.). Und durch die ununterbrochene Interaktion von Seiendem konstituieren sich die Bedeutungen (im Sinne der ‚Bedeutung B‘) als Inhalt mentaler Repräsentationen „in den Köpfen“ der Menschen (vgl. Gärdenfors 2000: Begriffsräume der kognitiven Semantik, vgl. auch Dynamische Systeme und Embodied Embedded Cognition).

Kurz: Aus jeglicher Struktur (bzw. Syntax) entsteht durch deren Gebrauch Semantik. Jede Struktur (bzw. Syntax) bedeutet potentiell bereits ihre Verwendungsweise. Daher bedeutet ein rein physikalisch arbeitendes System etwas, wie am menschlichen Gehirn eindrucksvoll zu sehen ist. Dies liegt an der Interaktivität des Seienden und an der Tatsache, dass alles strukturierte Seiende inhärent potentielle Bedeutung B trägt, die zur aktuellen Bedeutung wird, wenn ein Agent dazukommt, der diese Struktur benutzt. Je komplexer die Struktur eines Systems ist, desto wahrscheinlicher ist es, dass wir diesem System (z.B. Mensch) Intentionalität und Bewusstheit zuschreiben, da dieses aus sich selbst heraus mit Anderem (inter-) agiert und jenem Bedeutung im Sinne von ‚Bedeutung M → B‘ zuschreibt und so zu einem intentionalen System wird (M).

3.1.3 Definition von ‚Repräsentation B‘

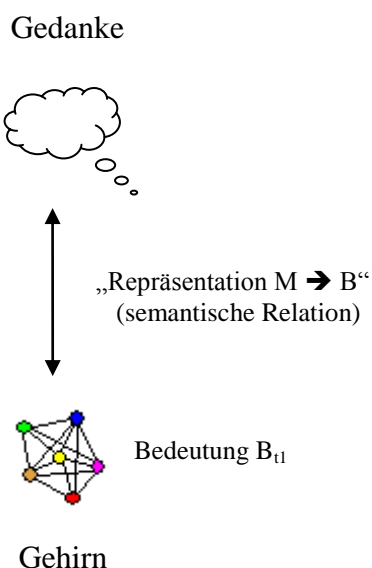
Die Unterscheidung dieser Arten von Bedeutung hat Auswirkungen auf die Definition des Repräsentationsbegriffs: Liegt Bedeutung B vor, dann tritt zunächst der Fall einer Präsentation ein. Ein physikalisch struktureller Zustand wird uns präsentiert

und damit auch seine potentielle Bedeutung. Diese B_{t1} könnte aber aufgrund bestimmter Umstände ein weiteres Mal eintreten, dann wird die bereits früher präsentierte Bedeutung dieses Zustandes wieder präsentiert B_{t2} , also „repräsentiert B_{t1} - B_{t2} “. Es wird also nicht eine Bedeutung, sondern ein Wiederauftreten einer potentiellen Bedeutung angezeigt. Dies ist aber ein ganz anderer Sinn von Repräsentation als der übliche, bei dem der physikalische Zustand für eine zugeordnete Bedeutung steht („repräsentiert $M \rightarrow B$ “). Es liegt also nicht mehr eine klassisch-semantische Repräsentation vor, welche ich als „vertikal“ bezeichne, sondern nur noch eine „zeitliche Repräsentation“ (Regeneration), welche ich „horizontal“ nenne. Abbildung 3.1.3 zeigt diese Unterschiede grafisch auf.

Der Begriff ‚Präsentation‘ ist dem Begriff ‚Repräsentation‘ sehr ähnlich. Das Prädikat ‚präsentieren‘ ist ebenfalls mindestens zweistellig. Man kann nicht sagen, „X repräsentiert“ oder „X präsentiert“, sondern nur „X repräsentiert Y“ oder „X präsentiert Y“. Allerdings sind die Arten der Relationen unterschiedlich. Bei der „Präsentation“ liegt nur eine Bedeutungsrelation vor. Etwas wird präsentiert. Im Falle „Repräsentation“ liegt neben der Bedeutungsrelation auch eine Zeitrelation vor. Etwas wird wieder präsentiert.

VERTIKALE RELATION

(semantisch)



HORIZONTALE RELATION

(zeitlich)

REPRÄSENTATION B

„Repräsentation B_{t1} - B_{t2} “
(zeitliche Relation)

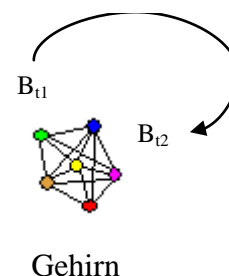


Abbildung 3.1.3: Vertikale und horizontale Relation

Damit dieses „Etwas“ aber als Bedeutung zum Vorschein kommt, muss es interpretiert und explizit gemacht werden. Dazu ist wiederum Etwas nötig, dass diese Interpretation vornimmt (d.h. auf einer höheren Ebene – meist ein intentionales Wesen). Dann liegt explizit eine Repräsentationsrelation vor. Liegt aber auch dann eine Präsentation oder Repräsentation vor, wenn niemand deren Bedeutung interpretiert?

Sowohl der Inhalt einer Präsentation als auch der Inhalt einer Repräsentation sind Bedeutungen. Es ist aber für die Art der Repräsentation entscheidend, von welcher Art diese Bedeutung ist:

1. Bedeutung im Sinne von ‚Bedeutung $M \rightarrow B$ ‘:

Die Bedeutung wurde (von einem intentionalen Wesen) zugeschrieben. Es liegt eine *explizite Repräsentation* vor.

2. Bedeutung im Sinne von ‘Bedeutung B’:

Die Bedeutung wird nicht abgelesen oder gebraucht, liegt aber mit der Struktur bereits fest. Bedeutung fällt mit der Struktur zusammen in dem Sinne, dass sie sich aus der Verwendung dieser Struktur ergibt, d.h. B wird bedeutsam bzw. ergibt sich erst in Verbindung mit einem Prozess, der mit dieser Struktur etwas anfangen kann. Die Bedeutung wird auch von niemandem zugeschrieben. Es liegt also keine explizite Repräsentation vor, sondern eine *implizite Repräsentation*.

Gewöhnlich wird man im ersten Fall von echter Präsentation oder Repräsentation sprechen. Im 2. Fall hingegen kann man m. E. auf dreifache Weise verfahren:

- (i) Man betrachtet Repräsentationen als nicht-relationale Repräsentationen, was aber der analytischen Relationalität des Repräsentationsbegriffes widerspricht.
- (ii) Man betrachtet Repräsentationen nicht als Repräsentationen. In diesem Fall wäre man methodischer Antirepräsentationalist. Allerdings ist man dann den Problemen des Antirepräsentationalismus ausgesetzt (s. Kapitel 4), was mit der dritten Möglichkeit vermieden wird.
- (iii) Man betrachtet Repräsentation als eine von der klassischen Repräsentation verschiedenen Art. Im Folgenden wird versucht, diese dritte Art des Repräsentationsbegriffs zu definieren.

Wenn man Bedeutung als Inhalt einer Repräsentation versteht und den Repräsentationsbegriff analytisch als relational ansieht, dann darf man bei strukturinherenten, ungerichteten und („semantisch“) nichtrelationalen Bedeutungen im Sinne der ‚Bedeutung B‘ also nicht von Repräsentation im klassischen Sinne (wie in Abbildung 3.1.1 dargestellt) sprechen. Bei den gerichteten, intentionalen Bedeutungen

im Sinne ‚Bedeutung $M \rightarrow B$ ‘, welche die Voraussetzung der („semantischen“) Relationalität einer Repräsentation erfüllen, kann hingegen sehr wohl von Repräsentation gesprochen werden.

Eine klassische Repräsentation R besteht aus dem Repräsentierten, dem Repräsentierenden und einer Repräsentationsrelation RR : Das Repräsentierte ist der Gegenstand G der Repräsentation R (z.B. ein Stein). Das Repräsentierende besteht aus zwei Teilen: dem mentalen Inhalt I der Repräsentation R (z.B. einer Vorstellung von diesem Stein) und der physikalischen Implementierung M des Inhalts I der Repräsentation R (z.B. durch die neuronale Implementierung dieser Vorstellung des Steins im Gehirn). Die klassische Repräsentationsrelation RR wird als eine Beziehung zwischen I und G verstanden. Es sind aber noch weitere Beziehungen zu berücksichtigen. Die Beziehung zwischen I und M und die Beziehung zwischen M und G . Alle diese Beziehungen sind problematisch.

Allerdings denke ich, dass sich die Probleme mit Hilfe des Ansatzes dynamischer Systeme lösen lassen, ohne dabei auf diese Beziehungen und somit auf den Repräsentationsbegriff verzichten zu müssen. Innerhalb dynamischer Systeme könnte man dann von einer neuen Art von Repräsentation sprechen. Um dies zu demonstrieren, soll ein Beispiel helfen, das die Beziehung zwischen G und I klarer macht.

Als Gegenstand G nehme ich einen Stein. I soll die Vorstellung von dem Stein sein. Ich kann von dem Stein keine Vorstellung haben, ohne ihn zumindest einmal wahrgenommen zu haben, denn Begriffe ohne Anschauung sind leer (vgl. Kant 1787). Ich muss die „Bekannschaft“ mit diesem Stein machen, ihn kennen lernen, d.h. eine Repräsentation von ihm erzeugen, ihn erkennen (vgl. Russell 1918/19). Dies geschieht, indem durch den Stein eine Wahrnehmung des Steines als Muster im Gehirn kausal hervorgerufen wird. Aufgrund des dynamischen Charakters des Gehirns wird dieses Muster mit bereits vorhandenen Mustern zu etwas verrechnet (kombiniert), das man als die Vorstellung von dem Stein bezeichnen könnte. Immer wenn ich etwas Ähnliches erblicke wie diesen Stein, dann wird wieder eine Vorstellung von einem Stein dynamisch aufgebaut. Natürlich sind die Vorstellungen aufgrund der Dynamik des Systems nicht identisch. Dadurch ist es auch möglich, sich zu täuschen, d.h. auch Missrepräsentationen sind möglich: In der Dämmerung kann man schon mal eine Kuh für ein Pferd halten oder umgekehrt.

Ich habe nun eine Repräsentation von dem Stein als Vorstellung von dem Stein im Gehirn in Form eines stabilen physikalischen Systemzustandes. Die Verknüpfung zwischen G und I wurde aufgebaut. I repräsentiert G . Immer wenn G wahrgenommen

wird, stellt sich I ein. Aufgrund des physikalisch kausal operierenden dynamischen Systems (Gehirn), ist es auch möglich, dass sich I aufgrund eines anderen I' einstellt. D.h. eine Vorstellung kann durch eine andere hervorgerufen werden. Anders ausgedrückt: ein stabiler Systemzustand kann aufgrund physikalischer Vorgänge im System einen anderen stabilen Systemzustand kausal verursachen. Die Vorstellung von dem Stein könnte z.B. die Vorstellung von einem Haus hervorrufen, wenn man z.B. vorher gelernt hat, dass Häuser aus Steinen gebaut werden können. Die Beziehung zwischen I und G ist (im Falle dass G ein Gegenstand der Welt ist) zusammengesetzt aus der Beziehung zwischen I und M und der Beziehung zwischen M und G. Es gibt also keine direkte Verbindung zwischen G und I, sondern nur eine indirekte über G und M sowie M und I.

Alle diese Relationen sind auf physikalischer Ebene angesiedelt: Die Relation zwischen I und G kommt auf rein physikalischem Wege zustande. RR ist nicht kausal, sondern korrelativ, d.h. es gilt meist folgendes: immer wenn G vorliegt, liegt auch I vor. Diese Relation ist deshalb nicht kausal, weil es ab und zu sein kann, dass das System aufgrund von Ausnahmefällen (d.h. durch interne und/oder externe Einflüsse) eine andere innere Dynamik (= zeitlich bedingtes Verhalten) aufweist. Je öfter G vorliegt, desto wahrscheinlicher ist es, dass M vorliegt. Je öfter M mit I verbunden wird, desto wahrscheinlicher ist es, dass eine Repräsentationsrelation RR zwischen I und G besteht und somit eine Repräsentation R vorliegt. Dynamische Systeme verfügen über stabile Zustände, die sich kausal aus ihrer Systemdynamik ergeben (z.B. sogenannte Attraktoren)⁵. Diesen können Bedeutungen zugeordnet werden und somit als explizite Repräsentationen betrachtet werden. Aufgrund ihrer Entstehungsgeschichte innerhalb des Gesamtgefüges Mensch-Umwelt können diese Zustände aber auch aus sich (d.h. implizite) Bedeutung haben und daher als implizite Repräsentationen angesehen werden. Die als antirepräsentationalistisch bezeichneten dynamischen Systeme können somit sowohl über implizite als auch über explizite Repräsentationen verfügen.

Die Beziehung zwischen I und G ergibt sich also aus der Interaktion des Systems mit eigenen Teilsystemen und mit der Umwelt über die Beziehungen zwischen I und M sowie M und G. Innerhalb des Systems gibt es aber auch Beziehungen zwischen den verschiedenen I's untereinander.

⁵ Für die weitere Diskussion reicht es vorerst aus, einen Attraktor als stabilen Systemzustand bzw. als stabiles Systemverhalten eines dynamischen Systems zu betrachten. Für genauere Hinweise zu Attraktoren und weiteren Fachbegriffen aus der dynamischen Systemtheorie siehe S. 44f in Abschnitt 3.2.3.

Innerhalb dynamischer Systeme sind I und M identisch (d.h. die vertikale Relation $(M \rightarrow B)$ wird eliminiert), sodass die Beziehung zwischen I und G nur noch aus der Verbindung zwischen M und G besteht (sofern G ein Ding in der Welt ist).

Ist G hingegen etwas Abstraktes (Mentales), dann fällt G mit irgendeinem I zusammen. Und wegen $I=M$ gilt: G als Abstraktes (Mentales) fällt mit M zusammen. Die Beziehung zwischen M und G (egal, ob G abstrakt oder konkret ist) ist wie oben bereits erwähnt nicht kausal, sondern korrelativ.

Zusammenfassend: Es ist angebracht, nicht nur die Unterschiede zwischen Repräsentation und Präsentation aufzuzeigen, sondern auch eine neue Art von Repräsentation einzuführen: Repräsentation im Sinne einer „zeitlichen Repräsentation“ (Repräsentation B) als Wiederholung eines Zustandes innerhalb eines dynamischen Systems. Die Repräsentation B verfügt über keine vertikale Relation, d.h. über keine semantische Relation im Sinne der expliziten Bedeutung ‚Bedeutung $M \rightarrow B'$ ‘. Sie ist eine implizite Repräsentation. Die Begriffe ‚Präsentation‘ und ‚Repräsentation‘ müssen analytisch als relational verstanden werden. Deren Bedeutungs- und Zeitrelationalität ist zu unterscheiden. Tabelle 3.1.3 fasst die Ergebnisse noch einmal kurz zusammen.

	klassische Repräsentation	Repräsentation B	Präsentation
Repräsentationsinhalt / Bedeutung	Referenz (Denotat)	Wiederholung bzw. wiederholte Verwendung eines stabilen Zustandes/Verhaltens innerhalb eines dynamischen Systems	Gebrauch
Repräsentationsart	explizit	Implizit	-
Semantisch relational im Sinne der impliziten ‚Bedeutung B‘	Nein	Ja	Nein
Semantisch relational im Sinne der expliziten ‚Bedeutung $M \rightarrow B'$ ‘ (vertikal)	Ja	Nein	Ja
Zeitlich relational (horizontal)	Ja	Ja	Nein
Vertreter	Frege	Pasemann, Jäger, Freeman, Maturana, Beer	Wittgenstein

Tabelle 3.1.3: Gegenüberstellung der Eigenschaften von klassischer Repräsentation, Präsentation und Repräsentation B

Vor diesem Hintergrund lege ich nun folgende Definition des Begriffes ‚Repräsentation B‘ fest, um einen begrifflichen Bezugspunkt für die weitere Untersuchung zu haben, in der die Repräsentationalität der in dieser Arbeit aufgezeigten Ansätze festgestellt werden soll.

Eine ‚Repräsentation B‘ liegt dann vor, wenn gilt:

- (i) Das Repräsentierte ist der Gegenstand G (z.B. ein Stein – kann auch etwas Abstraktes sein (wie z.B. die Zahl „5“ oder „Pegasus“), d.h. G kann sich innerhalb oder ausserhalb eines (kognitiven) Systems befinden).
- (ii) Das Repräsentierende I besteht aus dem mentalen Inhalt I der Repräsentation B (z.B. einer Vorstellung von dem Stein) als Eigenschaft eines dynamischen Systems aufgrund seiner Systemdynamik (z.B. die Vorstellung des Steins entspricht einem stabilen Systemzustand bzw. – Verhalten eines neuronalen Netzes).
- (iii) Die Repräsentationsrelation RR einer Repräsentation B ist eine Beziehung zwischen dem Repräsentierenden I und dem Repräsentierten G.

Zur Relation zwischen I und G ist noch folgendes zu sagen: I ist das, was man sich vorstellt, was man anderen zuschreibt, was man denkt. Alles dies passiert auf einer konzeptuellen Ebene. Da aber alles auf physischer Ebene implementiert ist, fällt die konzeptuelle Ebene mit der physischen zusammen. I kann in folgendem Sinne als emergente Eigenschaft aus I bezeichnet werden: I ist ein stabiler Systemzustand innerhalb eines dynamischen Systems, der sich aufgrund der Systemdynamik ergibt. I korreliert daher mit G (ob G nun etwas Konkretes ausserhalb des Systems oder etwas Abstraktes (als I) innerhalb des Systems ist). Dies sind explizite Repräsentationen, d.h. stabile Zustände, welche erkannt wurden und denen Ausdrücke bzw. Laute zugeordnet wurden. Die Bedeutung der Laute ergab sich aus dem Gebrauch der Laute in der entsprechenden Situation, d.h. durch eine implizite Konvention durch den bedeutsamen Gebrauch der Laute in der Gesellschaft. I ist ein stabiler Systemzustand, der sich aufgrund anderer, niedrigerer Systemzustände ergibt. Dazu können auch neuronale Implementationen solcher stabilen Systemzustände gehören oder gespeicherte wahrgenommene physikalische Muster (subsymbolische Repräsentationen vgl. Abschnitt 3.2.2, bes. S. 37ff). Es gibt durch die Dynamik des Systems eine ständige Korrelation zwischen solchen I-Zuständen und G-Zuständen im oder ausserhalb des Systems, wobei G-Zustände innerhalb des Systems auch I-Zustände sein können.

3.1.4 Weitere begriffliche Festlegungen

Implizite Repräsentationen werden im Gegensatz zu expliziten Repräsentationen nicht benannt bzw. wurden noch nicht als Repräsentationen erkannt. Sie repräsentieren

Dinge, denen wir uns meist gar nicht bewusst sind, wie z.B. interne Systemzustände, internes Systemverhalten, Einflüsse der Umwelt auf uns usw. Oft werden syntaktisch strukturierte Repräsentationen (Symbole) mit expliziten Repräsentationen bzw. konnektionistisch verteilte Repräsentationen (Subsymbole) mit impliziten Repräsentationen gleichgesetzt, weil explizite Repräsentationen im Vergleich zu impliziten leicht erkannt und benannt werden können. Explizite Repräsentation ist an ein Subjekt gebunden, implizite Repräsentation nicht. In diesem Sinne kann also von nicht erkennbaren und erkennbaren (impliziten) Repräsentationen und erkannten bzw. bereits benannten (expliziten) Repräsentationen gesprochen werden. Implizite Repräsentationen können vorhanden sein, auch wenn man sie niemals entdecken würde oder niemals entdecken könnte.

Symbole sind explizite Repräsentationen. Der Unterschied zwischen Symbolen und Variablen liegt darin, dass die Bedeutung von Variablen häufig undefiniert wird. Innerhalb symbolischer Systeme wird allerdings meist nur von Symbolen gesprochen, obwohl damit sowohl Symbole als auch Variablen gemeint werden. Variablen werden durch Interpretation (d.h. durch Zuordnung einer Bedeutung) qualitativ.

Interne Repräsentation findet „im Kopf“ eines kognitiven Agenten statt. Diese wird als „mentale Repräsentation“ bezeichnet. Externe Repräsentation findet ausserhalb des Kopfes statt. Ein Foto ist eine typische externe Repräsentation, eine Vorstellung eine typisch interne/mentale Repräsentation.

Der allgemeine Repräsentationalismus erfordert keine besonderen Entitäten wie Variablen, Symbole oder dergleichen. Vielmehr ist, um als Repräsentation gelten zu können, nur ein stabiler Zustand nötig, der einen Systemzustand oder ein Systemverhalten repräsentiert („einfriert“), der in einer späteren Operation noch einmal benötigt wird (entspricht der Repräsentation B).

Daher ist jegliches System, das über einen stabilen Zustand verfügt (der den Systemzustand oder das Systemverhalten repräsentiert, in dem sich das System beim Einstellen dieses stabilen Zustandes befand), der später von diesem System noch einmal gebraucht wird, ein repräsentationales System. Innerhalb eines Systems liegt dann eine Repräsentation B vor, wenn ein stabiler Zustand bzw. ein stabiles Systemverhalten innerhalb eines dynamischen Systems von diesem wiederholt verwendet wird. Repräsentationen B haben aufgrund der Dynamik dynamischer Systeme und der permanenten Interaktivität solcher Systeme mit der Umwelt sowohl interne als auch externe Bedeutung, d.h. Bedeutung, die aus dem inneren des Systems herrührt, als auch Bedeutung, die aus der Umwelt stammt.

Man kann innerhalb eines dynamischen Systems erst dann nicht mehr von internen Repräsentationen sprechen, wenn es keine Attraktoren bzw. kein stabiles Systemverhalten mehr gibt. Und es ist der Normalfall innerhalb eines dynamischen Systems, dass es über solche quasi-stabile Zustände verfügt und sich nicht ständig chaotisch verändert.

In der Kognitionswissenschaft werden noch weitere Arten von Repräsentationen unterschieden, auf die ich teilweise Bezug nehmen werde (vgl. Scheerer 1992b, S. 843):

- strukturiert vs. holistisch: Teile können semantisch interpretiert werden (kompositionelle Semantik) vs. Teile können nicht semantisch interpretiert werden.
- lokal vs. verteilt: Jede Repräsentation hat ihren eigenen „Ort“ vs. ein „Ort“ ist Teil verschiedener Repräsentationen.
- analog (numerisch) vs. digital (symbolisch): Zwischen zwei Zuständen ist immer noch ein dritter Zustand, der semantisch interpretiert werden könnte vs. diskret-semantisch-interpretierbare Zustände.
- propositional vs. nicht-propositional (z.B. räumlich): Dem Inhalt propositionaler Repräsentationen kann ein Wahrheitswert zugesprochen werden, dem Inhalt räumlicher Repräsentationen nicht.
- intrinsisch vs. extrinsisch: Die durch die Repräsentation abgebildeten Relationen sind als Relationen in der Repräsentation selbst enthalten vs. Relationen werden explizit benannt.
- primär (physisch) vs. sekundär (funktional) isomorph: Relationen oder Eigenschaften des Inhalts der Repräsentation sind in ihr selbst enthalten vs. Ähnlichkeiten zwischen den Objekten der Domäne sind durch Relationen zwischen den Elementen eines Repräsentationssystems abgebildet.

3.2 Der Repräsentationsbegriff in der Kognitionswissenschaft

Den bisher erfolgreichsten theoretischen und experimentellen Ansatz zur Erforschung des Geistes stellt nach Paul Thagard die computational-repräsentationale Vorstellung des Geistes (CRUM – computational representational understanding of mind) dar (vgl. Thagard 1999, S. 22f). CRUM basiert auf der Kognitivismusthese („Keine Kognition ohne Repräsentation“) und der Computationalitätshypothese („Keine Kognition ohne Computation“) und erklärt, wie Menschen ein bestimmtes intelligentes

Verhalten aufweisen können, also wie Kognition möglich ist, wenn man Kognition für intelligentes Verhalten voraussetzt.

CRUMBS (computational representational understanding of mind – biological-sociological) ist eine Erweiterung von CRUM durch neuere biologische und soziologische Ansätze, um die signifikante Rolle der physischen Umwelt und die Bedeutung von Emotion und Bewusstsein für das menschliche Denken zu berücksichtigen (vgl. Damasio 1994).

Antirepräsentationalisten verneinen zumindest das „R“ in CRUMBS (z.B. Stufflebeam 1998a), wenn nicht sogar zusätzlich das „C“ (z.B. van Gelder 1995), wie wir in diesem Kapitel noch sehen werden. Die Abbildung 3.2.1 soll einen Eindruck davon vermitteln, auf welche Weise die verschiedenen Ansätze in der kognitionswissenschaftlichen Literatur m. E. als „antirepräsentationalistisch“ eingeordnet werden (vgl. auch Metzinger 1998).

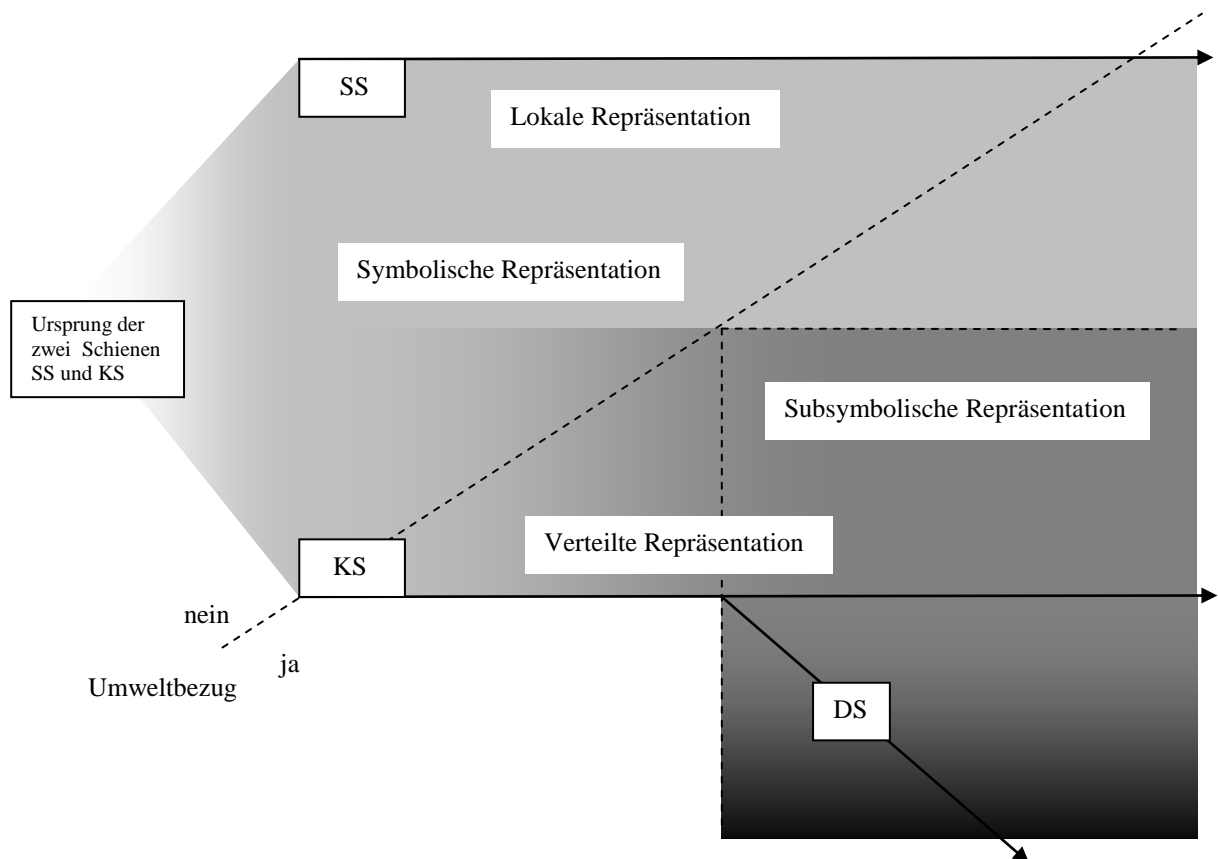


Abbildung 3.2.1: Die drei wichtigsten Paradigmen in der Kognitionswissenschaft (SS: symbolische Systeme, KS: konnektionistische Systeme, DS: dynamische Systeme; je dunkler, desto antirepräsentationalistischer)

In diesem Abschnitt werden die drei wichtigsten Paradigmen innerhalb der Kognitionswissenschaft zur Erklärung der Kognition näher vorgestellt: symbolische Systeme (SS), konnektionistische Systeme (KS) und dynamische Systeme (DS). Dabei

wird insbesondere auf jene Aspekte eingegangen, welche die Repräsentationalität der Ansätze betreffen, da gezeigt werden soll, dass und wie sich die Sicht auf Repräsentationen innerhalb der Kognitionswissenschaft in Bezug auf deren Inhalt, Form und Implementierung verändert hat. Zuvor werden kurz jene Begriffe aus der Systemtheorie angesprochen, die für die Darstellung der Ansätze nötig sind.

Ein System besteht aus einer bestimmten Anzahl strukturell angeordneter Elemente, die sich gegenseitig beeinflussen (Rückkopplungen). Jedes System erfüllt innerhalb seiner Systemgrenzen, die es von seiner Umwelt trennt, einen bestimmten Zweck. Über Systemein- und -ausgänge kann dabei ein Austausch von Stoff, Energie und Information erfolgen (offenes System) – im Gegensatz zu geschlossenen Systemen ohne jegliche Schnittstellen zur Umwelt. Die dabei in der Zeit entstehenden sequentiellen Zustände als Summe der gegenseitigen Abhängigkeiten der Elemente ergeben über einen bestimmten Beobachtungszeitraum den sogenannten Zustandsraum.

„Ein System besteht aus einem oder mehreren strukturell verbundenen Elementen, deren Zustände von anderen Elementen (oder sich selbst) abhängen und die die Zustände anderer Elemente (oder sich selbst) beeinflussen.“ (Bossel 1989, S. 10).

Jedes System verändert sich mit der Zeit. In konkreten Systemen stehen konkrete Ereignisse in einer zeitlichen Beziehung (Reihenfolge) zueinander. Abstrakte Systeme unterliegen nicht der (Echt-) Zeit und verwenden daher eine (geordnete) Zahlenmenge als Zeitmenge (Reihenfolge). Reziproke, direkte Abhängigkeit wird als Koppelung bezeichnet. Systemvariablen und verkoppelte Parameter bilden miteinander ein größeres System. Was ein System ist, wird willkürlich festgelegt, indem man bestimmt, welche Menge dieser Variablen und Parameter man als zu *einem* System gehörend untersuchen will. Ob diese Menge dann tatsächlich ein System bildet und wie es sich verhält, zeigt sich in der Realität. Ein System wird durch ein Modell beschrieben, welches durch seinen Modellzweck bestimmt wird. Aufgrund der zunehmenden Komplexität der Modelle werden oft Computersimulationsmodelle herangezogen, um formalisierte, d.h. mathematische oder logische Modelle mit Hilfe von schnellen Rechnern zu simulieren. Da die Systemtheorie innerhalb sehr vieler Forschungsrichtungen eingesetzt wird, gibt es entsprechend viele Ausprägungen von „Systemtheorien“ (vgl. Jäger 1996, S. 152ff).⁶ Gemeinsam ist ihnen aber allen, dass „ein System etwas ist, das sich in der Zeit

⁶ U. a. wurden folgende Systemtheorien entwickelt: Kybernetik (Wiener), Mess- und Regelungstechnik, Automatentheorie, Autopoietische Systeme (Maturana & Varela), Synergetik (Haken), Allgemeine Systemtheorie (Bertalanffy), Evolutionsdynamik, Selbstorganisation (Haken), Informationstheorie (Shannon), Thermodynamik, zelluläre Automaten (von Neumann), mathematisch-dynamische Systeme usw. Eine übersichtliche Gegenüberstellung der Paradigmen von Systemtheorie und klassischer Naturwissenschaft gibt Krieger (vgl. Krieger 1998, S. 19/20 Fussnote).

entwickelt und aus Teilsystemen oder Komponenten besteht, so dass das Ganze mehr als die Summe der Teile ist.“ (Jäger 1996, S. 152) (vgl. Emergenztheorie in Abschnitt 4.3.3, S. 82ff).

3.2.1 Symbolische Systeme

Symbolische Systeme entstammen der Mathematik. Alan Turing bewies, dass jedes formal erfassbare Problem automatisch mit einer Maschine in endlichen Schritten (Algorithmus) gelöst werden kann (Turing-Maschine). Alonzo Church erweiterte diesen Algorithmus zu einer universellen Turing-Maschine. John von Neumann konstruierte eine Maschine, die diesen universellen Algorithmus implementierte. Es war ein Computer, der über eine Zentraleinheit und einen Speicher, sowie eine Ein- und eine Ausgabeeinheit verfügte.⁷

Die Hypothese physikalischer Symbolsysteme von Newell (1980) besagt, dass physikalische symbolische Systeme über die notwendigen und hinreichenden Mittel für allgemeine (d.h. auch menschlich) intelligente Handlungen verfügen. Kognition kommt der physikalischen Manipulation von Symbolen gleich und wird daher auch Computationalitätshypothese genannt (vgl. Newell & Simon 1981 und van Gelder 1998):

“Physical Symbol System Hypothesis: The necessary and sufficient condition for a physical system to exhibit general intelligent action is that it be a physical symbol system. Necessary means that any physical system that exhibits general intelligence will be an instance of a physical symbol system. Sufficient means that any physical symbol system can be organized further to exhibit general intelligent action. General intelligent action means the same scope of intelligence seen in human action: that in real situations behaviour appropriate to the ends of the system and adaptive to the demands of the environment can occur, within some physical limits. The hypothesis takes as given the identity of symbol systems and universal systems, and asserts their connection to rationality, a concept which did not enter into their formulation. “ (Newell 1980, S. 170)

Symbole zeichnen sich dadurch aus, dass sie für etwas stehen. Sie verfügen über eine Repräsentationsrelation und können daher als Repräsentationen angesehen werden. Ein Symbol kann als eine physikalische oder mentale Struktur betrachtet werden, welche als eine Repräsentation von etwas interpretiert werden kann (vgl. Lai-Chong Law 1993). D.h. Symbole sind explizite Repräsentationen (vgl. Abschnitt 3.1.4, S. 22f).

⁷ Vgl. Haugeland 1985 und Haugeland 1997.

Philosophisch wurde die Computationalisthypothese von Jerry Fodor durch seine Repräsentationale Theorie des Geistes (RTG) gestützt (vgl. Fodor 1975). Der Geist soll so funktionieren wie ein physikalisches Symbolsystem (z.B. wie ein Digitalcomputer). Die Forschung im Bereich der künstlichen Intelligenz förderte die Attraktivität dieser Theorie. Es zeigte sich anhand der raschen Verbreitung von Computern, dass sich die symbolische Informationsverarbeitung formaler Art in vieler Hinsicht sehr bewährte.

Fodor geht davon aus, dass ein Wesen genau dann einen intentionalen Zustand mit einem bestimmten Inhalt hat, wenn sich dieses Wesen in einer funktionalen oder computationalen Relation zu einer mentalen Repräsentation befindet, deren Bedeutung diesem Inhalt entspricht. Er schreibt „Wesen“, weil er damit auch „Computer“ mit einschließen möchte. Weiterhin geht Fodor davon aus, dass der Geist eine eigene Sprache – „Mentalesisch“ – spricht (Language of Thought). Dies bedeutet, dass die mentalen Repräsentationen quasi-syntaktisch strukturiert sind. Deshalb sind auch typidentische Teile innerhalb dieser mentalen Repräsentationen untereinander austauschbar. Die Bedeutung komplexer mentaler Repräsentationen setzen sich regelhaft aus den Bedeutungen von mentalen Repräsentationen geringerer Komplexität und zusammen. Intentionale Zustände stehen in einem kausalen Verhältnis zueinander. Diese ergeben sich durch die symbolische Verarbeitung der strukturierten mentalen Repräsentationen. Bedeutungen ergeben sich kontextunabhängig nur aus der Kombination der Bedeutung einzelner Symbole und ihren kausalen Beziehungen zur Welt.⁸ Es werden also die Symbole nur syntaktisch erkannt, ihnen anschließend eine Bedeutung zugeordnet und daraus die Gesamtbedeutung errechnet.

Ein großes Problem stellt aber dabei der Übergang von Syntax zur Semantik dar.⁹ Fodor entgeht diesem Problem scheinbar, indem er mentale Symbole postuliert, die bereits über Semantik verfügen. Aber Fodor kann sein Postulat „mentaler Symbole“ empirisch nicht untermauern. Antirepräsentationalisten hätten mit der Zuordnung von Semantik zur Syntax keine Probleme, wenn sie auf Repräsentationen wirklich verzichten würden. Denn wenn der Inhalt einer Repräsentation als Bedeutung

⁸ Diese Beziehung könnte sich ergeben, wie ich es in Abschnitt 3.1.3 auf S. 20 geschildert habe.

⁹ Stanley P. Franklin nennt diesen Streit um den Übergang von Syntax zur Semantik „First AI Debate“. Insgesamt vermutet er in der Geschichte der Künstlichen Intelligenz bisher drei solcher Streits (vgl. Franklin 1995). Die „Second AI Debate“ behandelt die Kontroverse zwischen symbolischen und subsymbolischen Systemen. Der jüngste Streit in der „Third AI Debate“ konzentriert sich um die Plausibilität des Antirepräsentationalismus: Sind Repräsentationen essentiell für menschliche Kognition oder nicht? Diese Debatte befindet sich somit im Hauptinteresse dieser Arbeit. Um die Veränderung der Verwendung des Repräsentationsbegriffs zu demonstrieren, wird aber auch kurz auf den Gehalt der zweiten Debatte eingegangen. Und auch die erste Debatte um das Symbol Grounding Problem von John Searle (1980) wird in der dritten Debatte wieder virulent (vgl. Abschnitte 3.1.2, S. 15 und 3.2.1, S. 33)

verstanden wird, dann entfällt mit der Repräsentation auch der Inhalt dieser und somit auch die Bedeutung, welche einer Syntax zugeordnet werden könnte. Eine Lösung dafür könnte sein, solche mentalen Symbole als Repräsentationen im Sinne der Repräsentation B zu verstehen, wie ich es in Abschnitt 3.1.2 geschildert habe (S. 15f).

Fodors repräsentationale (und computationale) Theorie des Geistes enthält folgende Prinzipien:¹⁰

- (i) Kompositionalität (die Bedeutung komplexer strukturierter mentaler Repräsentationen ergibt sich regelhaft aus der struktursensitiven Verarbeitung der Bedeutungen von strukturierten mentalen Repräsentationen geringerer Komplexität).
- (ii) Produktivität (eine endliche Anzahl von Regeln erlaubt eine unendliche Anzahl den Regeln entsprechender Strukturen (Produktionen)). Z.B. Die Regel, rekursiv Adjektive verwenden zu dürfen, erlaubt die Bildung unendlich langer Sätze: Die attraktive grosse schlanke nette gepflegte gutaussehende gutverdienende ... Frau steht dort.
- (iii) Systematizität (kategorial gleiche Komponenten sind systematisch vertauschbar). Z.B. Nominalphrasen (NP) der Satzstruktur NP-V(erb)-NP: Mary loves John. John loves Mary.
- (iv) Inferentielle Kohärenz (das Ziehen von Schlüssen folgt logischen Gesetzen). Z.B. entspricht folgendes Argument der logischen Schlussform „Modus Ponens“: Wenn es regnet dann nehme ich einen Regenschirm mit. Es regnet, also nehme ich einen Regenschirm mit.

Symbolische Systeme sind repräsentationale Systeme, weil sie über physikalische Symbole verfügen, die bestimmte Ausschnitte der Aussenwelt (durch die Zuordnung einer Bedeutung an Symbole und an Verarbeitungsprinzipien) repräsentieren. Innerhalb eines solchen Systems wird ein Teil der Welt mittels Repräsentationen modelliert.

Anhand des Beispiels der logischen „exklusiven ODER-Funktion“ (XOR-Funktion) wird demonstriert, ob und wie die verschiedenen Ansätze (SS, KS und DS) diese Funktion repräsentieren. Der Ausgang eines XOR-Gatters (funktionaler Teil eines elektronischen Bausteins) ist nur dann aktiv, wenn von dessen Eingängen nur genau einer aktiv ist. Ein praktisches Beispiel für das „XOR-Prinzip“ ist folgendes: Man kann entweder atmen oder schlucken, nicht aber beides gleichzeitig und auf beides zu verzichten wäre auch fatal. Tabelle 3.2.1.1 fasst die Semantik der XOR-Funktion

¹⁰ Die RTG ist repräsentational aufgrund der Annahme von strukturierten mentalen Repräsentationen und computational aufgrund der struktursensitiven Manipulation dieser strukturierten mentalen Repräsentationen.

zusammen (wahr/aktiv/möglich = 1, falsch/inaktiv/unmöglich = 0):

Input A	Input B	Output X
0	0	0
0	1	1
1	0	1
1	1	0

Tabelle 3.2.2.1: Semantik der XOR-Funktion

Logische Schlussfolgerungen sind grundlegende Operationen des kognitiven Systems. Alle Systeme, die die XOR-Funktion beherrschen, können auch all die anderen booleschen Funktionen wie z.B. OR, AND und NOT. Die NOT-Funktion, also die Negation, ist eine ausgesprochen typische Fähigkeit der menschlichen Kognition (wie z.B. in „Heute regnet es nicht.“). Sicherlich könnte man intuitivere Beispiele für Kognition anführen. Die XOR-Funktion wird aber deshalb als Beispiel herangezogen, weil sich bereits einmal herausstellte, dass diese nicht von jedem System repräsentiert werden kann. Minsky & Papert (1969) zeigten nämlich, dass ein Ein-Ebenen-Perceptron (so, wie es Rosenblatt 1962 entwarf, d.h. ein einfaches neuronales Netz mit einer Ebene von Neuronen) die XOR-Funktion nicht repräsentieren kann. Fortan wurde angenommen, dass nicht alle Funktionen, die auch auf symbolischen Maschinen durchführbar waren, auch von neuronalen Netzen durchführbar sind. Dies war ein Schlag gegen den Konnektionismus. Die Fachwelt favorisierte daraufhin symbolische Systeme. Erst als sich durch die Arbeiten von Anderson (1983) sowie Rumelhart & McClelland (1986) und „ihrer PDP-Research-Group“ herausstellte, dass Multilayer-Perceptrons, also Netze mit mehreren Neuronenebenen, doch in der Lage sind, alle Funktion symbolischer Systeme nachzubilden, wurde die intensive konnektionistische Forschung wieder aufgenommen (vgl. Rumelhart 1986). Und wenn durch diesen Schlag der Konnektionismus kurzzeitig in Misskredit geraten war, so könnte diese kleine Funktion auch anderen Ansätzen das Fürchten lehren (z.B. dem Ansatz dynamischer Systeme).

Die Umsetzung der XOR-Funktion innerhalb eines symbolischen Systems erfolgt meist über ein Programm. Zur Demonstration der Implementierung der XOR-Funktion in solchen Systemen habe ich ein kleines Programm geschrieben, das den folgenden Versuchsaufbau realisiert (Abbildung 3.2.1.1):

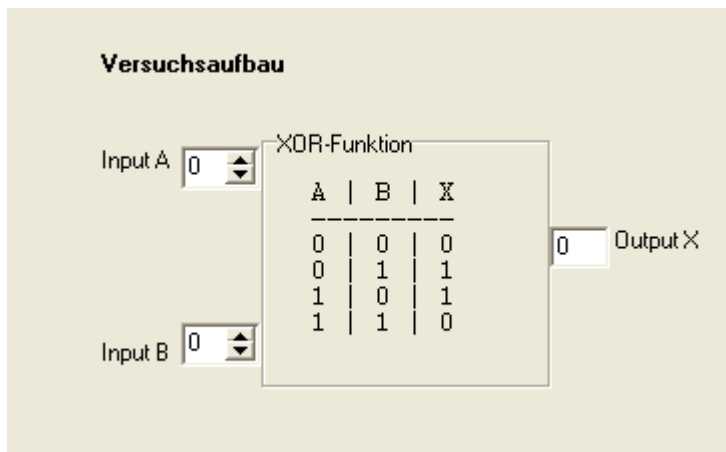


Abbildung 3.2.1.1: Versuchsaufbau für die XOR-Funktion

Abbildung 3.2.1.2 zeigt, wie die XOR-Funktion in der Programmiersprache Pascal implementiert wurde:

Symbolische Repräsentation

```

XOR-Funktion
function output_x (input_a,input_b:integer):integer;
begin
  if (input_a=0) and (input_b=0) then output_x:=0;
  if (input_a=0) and (input_b=1) then output_x:=1;
  if (input_a=1) and (input_b=0) then output_x:=1;
  if (input_a=1) and (input_b=1) then output_x:=0;
end;

```

Abbildung 3.2.1.2: Symbolische Repräsentation (Implementierung der XOR-Funktion über eine Programmiersprache)

Die Semantik der XOR-Funktion ist in dieser Funktionsroutine symbolisch definiert. Das Ergebnis der Funktion ergibt sich aus der Verknüpfung der Eingangszustände. Die Verknüpfung stellt nichts anderes als eine Manipulation von Symbolen (Variablen) dar und wird durch das Programm vorgegeben. In einem Computer mit von Neumann-Architektur gelangen die angelegten Eingangssignale (1 = Strom vorhanden, 0 = kein Strom vorhanden) zum Prozessor, dieser verarbeitet sie und das Ergebnis liegt in Form des Ausgangssignals vor. Die dabei ablaufenden elektronischen Prozesse werden kausal von elektronischen Bauteilen bestimmt. Die Verschaltung und softwaremässige Aktivierung dieser Bauteile implementiert die gewünschte Funktionalität. Diese wird vom Programmierer über eine Software

vorgegeben (hier über die Funktion `output_x`), welche die Semantik der XOR-Funktion innerhalb der Programmiersprache Pascal repräsentiert.¹¹ Die Befehle (Symbole) dieser Programmiersprache werden vom Prozessor sequentiell abgearbeitet und führen schliesslich zum gewünschten Ergebnis am Ausgang (`output_x`). Insgesamt wird dadurch die XOR-Funktion von dem Gesamtsystem repräsentiert und zwar durch die Manipulation von Symbolen (bzw. Variablen, s. Abschn. 3.1.4, S. 23) (in der function `output_x`), in Form von Produktionsregeln (IF ... THEN ... -Anweisungen).

Die Ausgangssignale sind ebenso exakt und diskret wie die Eingangssignale. Es erfordert viel Programmieraufwand, diese Programme oder Funktionen „unscharf“ (fuzzy) zu machen oder analoge Ausgangssignale auszugeben oder eine kontinuierliche Verarbeitungsweise anstatt einer schrittweisen (diskreten) auszuführen.¹² Dies ist eine inhärente Schwäche klassischer symbolischer Systeme.

Die verwendeten Symbole stehen für bestimmte, ihnen zugewiesene Bedeutungen. Das Programm besteht aus Variablen wie `input_a`, `input_b` und `input_x`. Variablen können – sobald sie einen Wert zugewiesen bekommen haben – als Symbole betrachtet werden. Sie repräsentieren in unserem Beispiel die Werte 0 oder 1 der XOR-Funktion. Das Programm besteht auch aus Symbolen wie „if“, „then“, „and“, „integer“ und „function“. Die Bedeutungen dieser Symbole sind unveränderbar und bedeuten immer das gleiche, nämlich: „wenn“, „dann“, „und“, „natürliche Zahl“ und „Funktion“. Aus allen diesen Elementen wird ein Algorithmus konstruiert, der im Programm in Abbildung 3.2.1.2 zum Ausdruck kommt. Alle diese Symbole sind Repräsentationen im klassischen Sinne. Symbolische Systeme sind also repräsentationale Systeme.

Man kann die gesamte XOR-Funktion aber auch als Repräsentation B auffassen: Der Gegenstand der Repräsentation ist die XOR-Funktion. Der Inhalt der Repräsentation ist die Bedeutung der XOR-Funktion, welcher durch die elektronischen Bauteile des PCs physikalisch implementiert wird. Der PC repräsentiert die XOR-Funktion.

Aber symbolische Ansätze haben einige Probleme zu lösen, wenn sie Kognition erklären wollen. Eine der Hauptfragen in der ersten Debatte der KI¹³ war, wie ein Symbol zu seiner Bedeutung gelangt.

¹¹ Zum Verhältnis zwischen ‚implementieren‘ und ‚repräsentieren‘ siehe Abschnitt 3.3.2, S. 53f.

¹² Wie dies bei Analogcomputern, neuronalen Netzen oder dynamischen Systemen der Fall ist. In diesen Systemen ist es auch wesentlich leichter „unscharfe Logiken“ (Fuzzy-Logiken) zu implementieren.

¹³ Künstliche Intelligenz (- Forschung)

Eine Möglichkeit ist, dem Symbol durch explizite Zuweisung (oder Konvention bzw. Verwendungsweise) eine Bedeutung zu geben. Es wird eine Zuordnung „von aussen“ getroffen (,Bedeutung $M \rightarrow B$ ’) (vgl. Abschnitt 3.1.3, S.15). Für diese explizite Bedeutungszuweisung an ein Symbol ist m. E. immer ein intentionales Wesen nötig, das diesen Zusammenhang durch Assoziation herstellt. Innerhalb der Modelltheorie (oder Typentheorie) geschieht dies, indem Ontologien entworfen und den Symbolen Dinge der externen Welt über eine Denotationszuweisungsfunktion zugeordnet werden (vgl. Cann 1993). Symbole erhalten so eine Referenz auf Dinge in der (auch möglichen) Welt und damit eine Bedeutung. Das Symbol repräsentiert dieses Ding, es steht für dieses Ding. Diese Repräsentationsbeziehung ist somit explizit relational. Die Bedeutung (oder der Inhalt einer Repräsentation) ergibt sich aus der Referenz auf das Bezeichnete.

Wenn man Symbole als physische Strukturen betrachtet, können Symbole aber auch an sich etwas im Sinne der ‘Bedeutung B’ bedeuten. In diesem Falle können Symbole als implizite Repräsentationen aufgefasst werden. Damit wird das Symbol Grounding Problem gelöst: Wie gelangt ein symbolisches Token, das nur von bedeutungslosen anderen symbolischen Tokens umgeben ist, zu einer Bedeutung? (vgl. Searles Argument vom Chinesischen Zimmer, Searle 1980).

Dieses Problem tritt vor allem in einem rein symbolischen Modell auf, weil darin die Verbindung zwischen den Symbolen und dem, worauf sie sich beziehen fehlt (meist die Aussenwelt). Wird ein solches symbolisches System aber mit einem konnektionistischen System verbunden, so kann eine ununterbrochene Verbindung durch die ständige Interaktion mit der Umwelt entstehen. Die Schnittstelle eines symbolischen Systems ist dafür nicht geeignet, da sie über zu wenig *gleichzeitig aktive* Ein- und Ausgänge verfügt. Ausserdem werden diese im Gegensatz zu neuronalen Netzen (Gehirnen) nacheinander (sequentiell) und nicht parallel verarbeitet. Symbole werden mit nichtsymbolischen Repräsentationen (z.B. wahrgenommene Bilder) verbunden, welche dann jene Objekte auswählen können, auf die sich die Symbole beziehen (vgl. Harnad 1990). Die Bedeutungskonstitution der Bedeutung B funktioniert ähnlich (vgl. Abschnitt 3.1.2, S. 14).

Vertreter des Konnektionismus behaupten, dass der symbolische Repräsentationsbegriff auf subsymbolische Systeme nicht übertragbar ist, da Symbole Konzepte zum Inhalt haben, Subsymbole aber nur Merkmale oder Teile von Merkmalen von Konzepten repräsentieren können (vgl. Smolensky 1988). Es können daher auf der subsymbolischen Ebene keine symbolischen Konzepte verwendet werden. Auf dieser Ebene ist Repräsentation nur mittels Subsymbolen möglich. Man nennt solche

subsymbolische Repräsentation „verteilt“ im Gegensatz zur „lokalen“ symbolischen Repräsentation. Man könnte zwar lokal den Ein- und Ausgängen eines neuronalen Netzes (wenn so interpretiert) Bedeutung zuordnen, so dass diese symbolhaften Charakter annehmen. Dies ist aber *innerhalb* des Netzes nicht möglich, da hier nur subsymbolische, verteilte Repräsentation vorliegt, die von aussen nicht interpretiert werden kann: weder einzelne Merkmale von Konzepten noch die Konzepte selbst (aufgrund der ständigen Änderung der Signale im Netz) können erkannt werden.

Die klassische, symbolische KI ist mit weiteren Problemen konfrontiert, die auf dem Festhalten an Repräsentationen für Kognition basieren. Das „Binding-Problem“ besteht in der Frage, wie Daten aus den verschiedenen Sinnesmodalitäten im Gehirn zusammengeführt werden. Die adäquate Verfolgung von Ereignissen und Objektveränderungen wird als „Frame-Problem“ bezeichnet (vgl. Abschnitt 2.3.2, S. 7, Fussnote 2 und Pylyshyn 1987 sowie French & Anselm 1999).

Unter anderem aus diesen Gründen wurden Alternativen zu symbolischen Ansätzen gesucht. Für die Erklärung menschlicher Kognition lag es nahe, solche Ansätze hervorzubringen, die sich stärker mit dem menschlichen „Prozessor“ für Kognition – dem Gehirn – auseinandersetzen.

3.2.2 Konnektionistische Systeme

Anfangs wurde der Konnektionismus eher ablehnend behandelt, nicht nur weil auch er mit einigen Problemen (z.B. Parametrisierungsproblem: welchen optimalen initialen Zustand soll das neuronale Netz (NN) haben, Labeling-Problem: wie kann man Neuronen Bedeutung zuordnen) zu kämpfen hatte (vgl. Rumelhart 1986 sowie Marcus 2001). Aber nach Abklingen der „Second AI Debate“ insbesondere zwischen den Symbolisten Fodor/Pylyshyn und dem Subsymbolisten Smolensky zeigte sich, dass beide Ansätze fruchtbar innerhalb der Kognitionswissenschaft weiter bestehen können (vgl. Fodor/Pylyshyn 1988, Smolensky 1988 und 1990). Dies führte auch zu hybriden Ansätzen, in denen symbolische und subsymbolische Ansätze vereint wurden (vgl. z.B. Harnad 1990).

D. O. Hebb (1949) schrieb dem Gehirn eine bestimmte Methode des Lernens zu. Er postulierte die Lernregel für die Verbindung zweier Neuronen: Werden zwei Neuronen gleichzeitig erregt, dann erhöhe deren Verbindungsstärke (Hebbsche Lernregel vereinfacht). In Anlehnung an dieses Erkenntnis wurden künstliche neuronale Netze als mathematische Modelle geschaffen.

Smolensky untermauerte mit seiner Subsymbolismus-Hypothese die Grundlage des konnektionistischen Paradigmas:

„The subsymbolic hypothesis: The intuitive processor is a subconceptual connectionist dynamical system that does not admit a complete, formal, and precise conceptual-level description. This hypothesis is the cornerstone of the subsymbolic paradigm.” (Smolensky 1988, S. 782)

Die Subsymbolismus-Hypothese besteht aus folgenden drei Unterthesen:

“[a] The connectionist dynamical system hypothesis [SYNTAX]: The state of the intuitive processor at any moment is precisely defined by a vector of numerical values (one for each unit). The dynamics of the intuitive processor are governed by a differential equation. The numerical parameters in this equation constitute the processor’s program or knowledge. In learning systems, these parameters change according to another differential equation. [...]

[b] The subconceptual unit hypothesis [VERTEILTE REPRÄSENTATION (SEMANTIK)]: The entities in the intuitive processor with the semantics of conscious concepts of the task domain are complex patterns of activity over many units. Each unit participates in many such patterns. [..]

[c] The subconceptual level hypothesis [SUBSYMBOLISCHE REPRÄSENTATION (SEMANTIK)]: Complete, formal, and precise descriptions of the intuitive processor are generally tractable not at the conceptual level, but only at the subconceptual level.” (Smolensky 1988, S. 780f, Einfügungen in eckigen Klammern von mir)

Verhalten kann auch subsymbolisch und ohne einsichtige kausale „Grundketten“ von Überzeugungen und Wünschen dargestellt und erklärt werden. Dies ist zwar nur innerhalb kleiner Bereiche möglich. Aber über eingegrenzte Problemräume (Domänen) sind die klassischen symbolischen Systeme auch noch nicht hinausgekommen.

Es wurde auch nachgewiesen, dass neuronale Netze – zwar in einer ganz anderen Art und Weise wie symbolische Systeme¹⁴ – systematisch im Sinne Fodors sind (vgl. Abschnitt 3.2.1, S. 29).

Die biologische Plausibilität von Kognition kann nicht auf kognitive Prozesse allein beschränkt werden, wie dies Ansätze klassischer Symbolsysteme in der Regel tun, um erfolgreich zu sein, sondern es muss der gesamte Kontext der menschlichen Perzeption und Handlung mit einbezogen werden (vgl. EEC in Kapitel 2.3.3, S. 8).

Gerade die Stärke des subsymbolischen Ansatzes, mit Unschärfe umgehen zu können, prädestiniert ihn, die menschliche Kognition adäquater zu simulieren als dies der symbolische Ansatz vermag. Denn ein grosser Teil der kognitiven Leistung des Menschen ist die perzeptuelle Wahrnehmung und das Generieren von Vorstellungen, wobei diese nicht immer ganz exakt oder identisch sind. Vielmehr hat es der Mensch

¹⁴ Vgl. Clark 2001, S. 77. Nach Clark zeigen Smolensky 1990 und van Gelder 1990, dass Systematizität über Vektormultiplikationen von Basisvektoren erzeugt werden kann.

oft mit sehr unterschiedlichen und meist nur ähnlichen, aber nur sehr selten mit identischen Gegebenheiten in der Welt zu tun. Der Mensch erkennt aus unvollständigen oder verrauschten Mustern oder Bildern dennoch deren Inhalt. Auch Einflussfaktoren wie Körperbewegungen, Werkzeuge, die lokale Umgebung sowie externe symbolische Strukturen werden dabei berücksichtigt.

Dies alles ist vor allem durch Prozesse der Mustererkennung und der Mustervervollständigung möglich, wie es neuronale Netze (insbesondere Gehirne) leisten. Die Daten werden direkt in den neuronalen Prozessen mit verarbeitet und durch diese Verflechtung von inneren und äußeren Prozessen eine kontextsensitive Kognition geschaffen, die sich gänzlich vom symbolischen Ansatz unterscheidet.

Das „Wissen“ eines neuronalen Netzes wird vorwiegend in dessen Struktur und Gewichtungen gespeichert. Dieses Wissen ist aber für einen Menschen in dieser Form nicht erkennbar oder einsichtig. Mittels statistischer Verfahren wie der Clusteranalyse¹⁵ können jedoch Extraktionen vorgenommen werden, so dass das gespeicherte Wissen symbolisch sichtbar wird (vgl. Clark 2001, S. 69f). Für zeitsensitive sogenannte „rekurrente Netze“ kann die sogenannte prinzipielle Komponentanalyse oder PCA (principal component analysis) für denselben Zweck angewandt werden (vgl. Clark 1993, S. 60-67).¹⁶

Ein anderer Vorteil konnektionistischer Systeme gegenüber symbolischen Systemen ist deren parallele Arbeitsweise.¹⁷ Es gibt keinen zentralen Prozessor mehr, sondern sehr viele davon – Neuronen. Dadurch wird eine dynamische, nicht-lineare Informationsverarbeitung bereits in der Grundstruktur eines solchen Systems möglich.¹⁸ Durch die Erweiterung einfacher Netzwerke mit zusätzlichen Neuronen, Verbindungen, Parametern und Berechnungsvorschriften werden heute biologisch adäquatere Lösungen erreicht, die zeitverzögerte Prozesszyklen sowie zeitkontinuierliche und analoge Verarbeitung¹⁹ mit Rauschen (d.h. geringen Störungen) zulassen.

¹⁵ Clauß³1999, S. 303ff: Mit Hilfe der Clusteranalyse kann eine Menge gegebener Objekte gemäss ihrer Merkmale gruppiert bzw. klassifiziert werden. Dazu wird ein Mindestmass an Ähnlichkeit definiert (Ähnlichkeits- und Distanzmaße). Es können hierarchische Gruppen („Ober- und Untergruppen“) und disjunkte Gruppen („Lage der Gruppen zueinander“) unterschieden werden.

¹⁶ PCA dient zur linearen Merkmalsextraktion, d.h. es können – im Gegensatz zur nicht-linearen Merkmalsextraktion nur maximal zwei Merkmalsklassen erzeugt werden. Das Verhalten von NNs kann statisch erfasst werden. Eliasmith sieht Gehirne als statistische Abhängigkeiten an (vgl. Eliasmith 1996).

¹⁷ Obwohl es auch symbolische Systeme gibt, deren Prozessoren parallel arbeiten. Allerdings verfügen diese Systeme im Gegensatz zu den konnektionistischen Systemen über keine flexible Systemstruktur.

¹⁸ Im Gegensatz zur linearen Verarbeitung ohne Rückkopplung und einer gerichteten Verbindungsstruktur spielen in nicht-linearen Systemen Rückkopplungen zwischen allen beteiligten Systemelementen eine grosse Rolle.

¹⁹ Im Gegensatz zur digitalen, diskreten Verarbeitung. Zur Unterscheidung zwischen analog und digital siehe S. 24 in Abschnitt 3.1.4.

Diese konnektionistischen Modelle der dritten Generation²⁰ entfernen sich immer weiter weg von der bisher vorherrschenden Vorstellung, dass der Kognition Symbole innewohnen – wie dies die Kognitivisten (wie z.B. Fodor) vertreten – hin zu einer eher anitrepräsentationalistischen Vorstellung.

Um einen Eindruck von der Repräsentationsfähigkeit neuronaler Netze und der Art, wie sie repräsentieren, zu bekommen, wird nun gezeigt, wie die XOR-Funktion in einem neuronalen Netz implementiert wird (s. dazu auch Marcus 2001, S. 7-16).

Abbildung 3.2.2.1 zeigt ein Multilayer-Perceptron, das aus drei Neuronenschichten besteht: einer Eingabeschicht mit zwei Neuronen (links) für Input A und Input B, einer versteckten Schicht mit vier Neuronen²¹ (mitte) und einer Ausgabeschicht mit einem Neuron (rechts) als Output X.

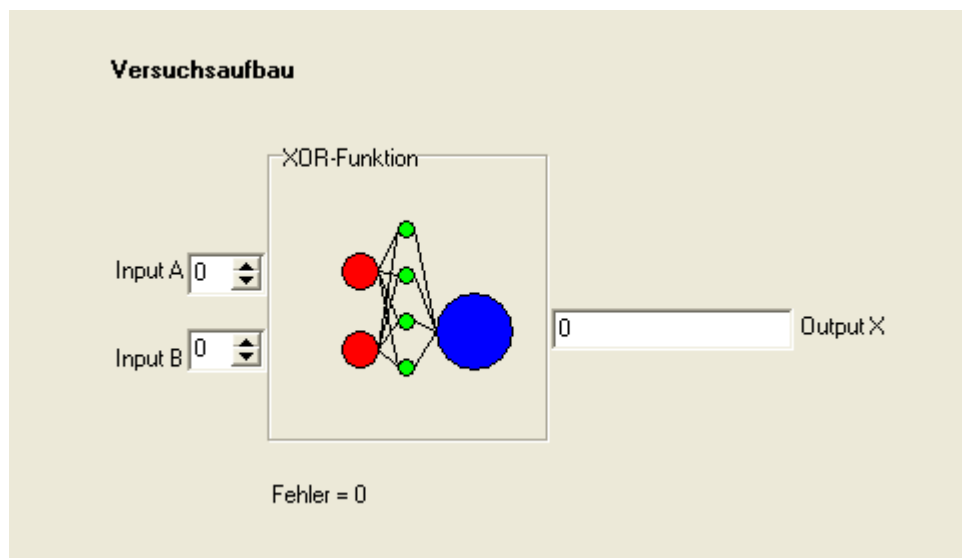


Abbildung 3.2.2.1: Versuchsaufbau mit einem neuronalen Netz (Multilayer-Perceptron: Feedforward-Netzwerk mit drei Neuronenschichten)

Wie gelangt nun dieses neuronale Netz zur Bedeutung der XOR-Funktion? Im Gegensatz zu symbolischen Systemen muss ein neuronales Netz (NN) trainiert werden. Dazu werden an die Eingangsneuronen des Netzes die Eingangswerte der XOR-Funktion angelegt (z.B. Input A=0 und Input B=0). In unserem Netz wirkt die Aktivierung der Eingangsneuronen über die versteckten Neuronen auf die

²⁰ Konnektionistische Modelle der ersten Generation sind Perceptrons (keine versteckte Neuronenschicht) im Gegensatz zu Multilayer-Perceptrons der zweiten Generation, die über versteckte Neuronenschichten verfügen. Ein zusätzliches Merkmal der Modelle der dritten Generation ist Kopplung. Jedes Neuron eines Netzes kann auf alle anderen Neuronen rekurren, da es mit allen anderen Neuronen verbunden ist. Daher werden solche Netze auch rekurrente Netze oder Elman-Netze genannt (nach ihrem Erfinder).

²¹ Zwei Neuronen als versteckte Schicht sind mindestens notwendig, um die XOR-Funktion repräsentieren zu können. In diesem Beispiel werden aber vier Neuronen in der versteckten Schicht verwendet, damit der unterschiedliche numerische Aktivierungsgrad dieser Neuronen besser erkennbar wird.

Ausgangsneuronen, d.h. in Vorwärtsrichtung. Deshalb wird diese Art von Netz auch „Feedforward-Netz“ genannt. Dies geschieht über die jeweilige Berechnung der Aktivierung der Neuronen, die mit einem vorhergehenden Neuron verbunden sind. Jedes Neuron kann deshalb als eigener Prozessor betrachtet werden, weil es aus den Ausgangssignalen anderer mit ihm verbundener Neuronen, der Gewichtung der Stärke der Verbindungen und der eigenen aktuellen Aktivationshöhe seinen eigenen neuen neuronalen Aktivationsgrad „berechnet“. Dieses Ergebnis wird an alle folgenden mit ihm verbundenen Neuronen weitergegeben, bis das Ausgangsneuron erreicht ist und dort das Ergebnis abgelesen werden kann (z.B. $X = 0.7$). Aufgrund der Differenz oder des „Fehlers“ zwischen diesem abgelesenen Wert und dem gewünschten Ausgangswert der XOR-Funktion (wenn $A = 0$ und $B = 0$, dann soll $X_{\text{soll}} = 0$ sein, $X_{\text{ist}} = 0.7$, d.h. die Differenz ist $X_{\text{diff}} = 0.7$), wird nun innerhalb des Netzes die Gewichtung der Verbindungen zwischen den Neuronen nach einem vorgegebenen Algorithmus, einer sogenannten „Lernregel“, neu berechnet. In unserem Netz wird die sogenannte „Backpropagation“-Lernregel verwendet. Dabei werden die Gewichtungen der Verbindungen zwischen den Neuronen nacheinander – angefangen bei den Ausgangsneuronen – zurück bis zu den Eingangsneuronen neu berechnet. Anschliessend werden wieder Wertepaare der XOR-Funktion angelegt und der Fehler erneut berechnet. Dieser Vorgang wird für alle „Zeilen“ der XOR-Funktion in Tabelle 3.2.1.1 (S. 30) solange durchgeführt, bis der Fehler Null wird. Aus Zeitgründen wird das Training aber meist vorher abgebrochen. Denn in Abhängigkeit der Struktur des Netzes und der Güte der Trainingsdaten ist die notwendige Anzahl der Trainingsdurchläufe, um das gleiche Ergebnis zu erzielen, unterschiedlich oder sogar unendlich gross.

Durch diesen völlig anderen Aufbau des Verarbeitungsprozesses von Signalen in einem neuronalen Netz können keine Symbole oder Regeln (wie die Produktionsregeln im symbolischen Beispiel in Abbildung 3.2.1.2, S. 31) direkt verarbeitet werden, obwohl einzelnen Neuronen symbolische Bedeutung zugeordnet werden kann (insbesondere den Eingängen – wie es im Beispiel der XOR-Funktion der Fall ist). Dieses Neuron trägt dann lokal eine Bedeutung und es liegt somit lokale Repräsentation vor. Allerdings ist es für neuronale Netze typisch, dass eine Repräsentation über alle Neuronen des Netzes verteilt ist, so dass meist nur verteilte Repräsentation vorliegt. Es erstreckt sich somit eine subsymbolische Bedeutungsrepräsentation über das gesamte Netz. Jedes Subsymbol (an einem Neuron) repräsentiert einen Bedeutungsanteil der gesamten Bedeutung der verteilten Repräsentation (vgl. Rumelhart 1986). Die

Bedeutung der XOR-Funktion wird also nicht von einem bestimmten Neuron, d.h. lokal repräsentiert, sondern sie wird verteilt repräsentiert, d.h. alle vier Neuronen des Netzes sind an der Konstitution der Bedeutung der XOR-Funktion beteiligt.

Es können sogar gleiche Bedeutungszustände aufgrund unterschiedlicher neuronaler Verbindungen und Aktivierungen bzw. Aktivationszustände der Neuronen von unterschiedlichen neuronalen Gesamtzuständen repräsentiert werden, d.h. die Semantik der XOR-Funktion könnte auch von einem Netz mit 10 versteckten Neuronen oder mit einem Netz, das zwar auch vier versteckte Neuronen hat, aber andere Aktivationszustände und Gewichtungen der Verbindungen aufweist, repräsentiert werden.

Aber auch umgekehrt können unterschiedliche Bedeutungszustände von denselben Neuronenverbänden gleicher Verbindungsstruktur, aber verschiedenen Aktivierungszuständen repräsentiert werden. D.h. anstatt der XOR-Funktion könnte auch eine OR-Funktion innerhalb desselben Netzes repräsentiert werden. Natürlich müsste es mit der OR-Funktion anstatt der XOR-Funktion trainiert worden sein.

Durch zusätzliche „Steuereingänge“ können sogar mit demselben Netz Multifunktionalität über Superposition (d.h. einer Überlagerung verschiedener Aktivationszustände) erreicht werden, indem man die gewünschte Funktion durch entsprechendes Aktivieren der Steuereingänge „auswählt“ (vgl. Pasemann 1996).

Die Aktivierung oder Hemmung der Feuerung (d.h. der Weitergabe des Aktivationszustandes) von Neuronen ist abhängig von den Gewichtungen der exzitatorischen (verstärkenden) oder inhibitorischen (hemmenden) Verbindungen zwischen den Neuronen, die sich aus der dynamischen Interaktion aller Neuronen ergeben.

Im XOR-Beispiel repräsentieren die Eingangsneuronen die beiden Eingangssignale der XOR-Funktion und das Ausgangsneuron das Ausgangssignal der XOR-Funktion. Entsprechend können nur noch die versteckten Neuronen in ihrer Gesamtheit die Semantik der XOR-Funktion repräsentieren. In der Tat wird durch das Aktivierungsmuster dieser Neuronen (bzw. durch die Veränderung der Gewichtungsstärken ihrer Verbindungen während des Trainings) dieses Wissen über die (bzw. Bedeutung der) XOR-Funktion in diesem Netz gespeichert. Im Gegensatz zum symbolverarbeitenden Computer, in dem das Wissen innerhalb elektronischer Bauteile wie z.B. in einer begrenzten Anzahl von Speicherbausteinen gespeichert wird, ist die Speicherfähigkeit neuronaler Netze während der Laufzeit variabel. Neuronale Netze sind plastisch, d.h. sie können während des Betriebes ihre Struktur und damit auch ihre

Speicherstruktur verändern.²² Allerdings muss diese „Speicherung“ in einem mit Sensoren und Effektoren verbundenem Gehirn dynamisch, nicht statisch gesehen werden, da sich zu jedem Zeitpunkt ein anderer Systemzustand (Aktivationsmuster) einstellt – in Abhängigkeit der sich ständig ändernden Eingangssignale (vgl. Riegler 1999, S. 15).

Jedes Neuron ist nur Teil einer Repräsentation. Viele Neuronen (hier vier) ergeben zusammen die Repräsentation der XOR-Funktion. Da alle vier Neuronen der versteckten Schicht gemeinsam an jeweils einer der vier Regeln der XOR-Funktion beteiligt sind, gibt es eine Überlagerung (= Superposition) von vier Repräsentationen verteilt über alle vier Neuronen (vgl. Jäger 1996, Pasemann 1996, Smolensky 1988). Es gibt jeweils eine Repräsentation für jede Regel, an denen aber jeweils alle 4 Neuronen gleichzeitig beteiligt sind.

Abbildung 3.2.2.2 zeigt die vier Zustände eines trainierten neuronalen Netzes für die XOR-Funktionalität: Die unterschiedlichen Aktivierungsmuster aller Neuronen (grau abgestuft) entsprechen der Implementierung der XOR-Funktion in diesem Netz. Für jedes der vier unterschiedlichen Muster an den Eingängen haben die Neuronen der versteckten Schicht und das Ausgangsneuron unterschiedliche Aktivationsstärke. Der Initialisierungszustand des neuronalen Netzes (d.h. die Anfangsaktivierung der Neuronen der versteckten Schicht) beeinflusst die Aktivationsmuster, die sich nach dem Lernen der XOR-Funktionalität in der versteckten Schicht einstellen. Nach jedem erneuten Training mit jeweils anderem Initialisierungszustand sind die Aktivationsmuster anders verteilt.

Es ist zu erkennen, dass die Neuronen der versteckten Schicht (mitte) für jede Wahrheitsbedingung der XOR-Funktion unterschiedliche Aktivierungsgrade aufweisen (unterschiedliche Grauschattierungen in der Abbildung). Damit liegt eine auf vier Neuronen verteilte Repräsentation dieser Funktion vor. Der Unterschied zum klassischen Paradigma, wo eine auf vier Programmzeilen verteilte Repräsentation oder auf vier elektronische Bauteile verteilte Repräsentation vorliegt, liegt in der fehlenden Plastizität fest verdrahteter Computer.

Am Eingang und am Ausgang liegt eine lokale, keine verteilte Repräsentation vor, weil jedem der beiden Inputs und dem Output jeweils genau ein Neuron mit einer bestimmten Bedeutung zugeordnet wird.

²² Die Überlagerung verschiedener Speicherzustände („superpositionale Repräsentationen“) auf der gleichen Speicherstruktur ist zwar bisher nur in neuronalen Netzen verwirklicht. Es wird aber an Quantencomputern gearbeitet, die über Speicher verfügen könnten, bei denen ebenfalls mehrfache Superposition möglich ist (vgl. Borchert 2000).

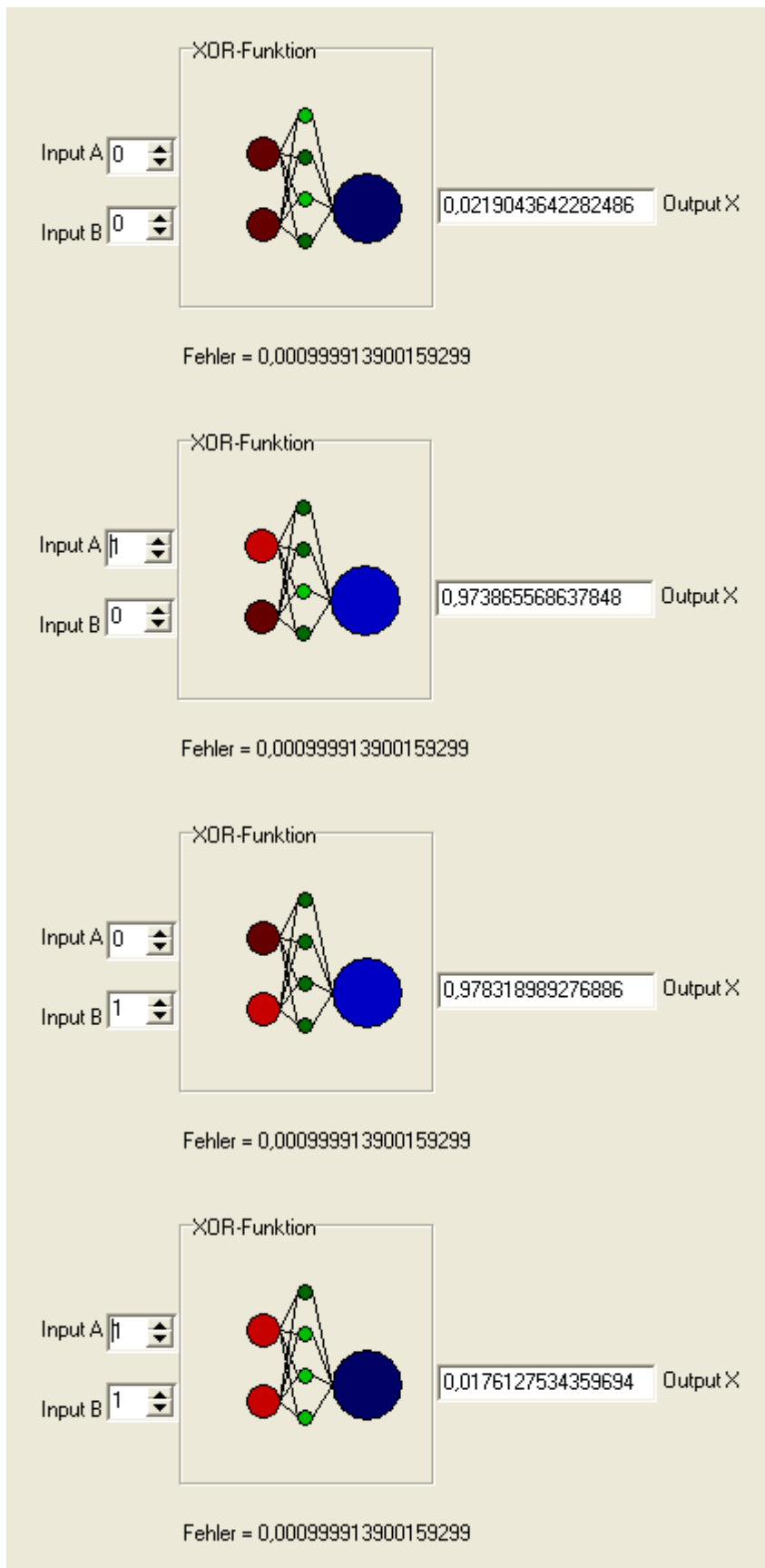


Abbildung 3.2.2.2: Subsymbolische Repräsentation (Die vier Zustände eines trainierten neuronalen Netzes für die XOR-Funktionalität)

Ferner ist im Gegensatz zur symbolischen Repräsentation das Ausgangssignal nicht notwendig exakt 0 oder 1, obwohl der Ausgangswert umso näher an 0 oder 1 liegt, je besser das Netzwerk entworfen und trainiert wurde.

Nun kann man einwenden, dass die XOR-Funktion damit nicht ordnungsgemäss repräsentiert ist, weil für diese nur exakte, diskrete Ein- und Ausgangswerte, nämlich 0 oder 1, erlaubt sind. Dieses Problem könnte aber durch ein weiteres neuronales Netzwerk, das dem XOR-Netz nachgeschaltet ist, beseitigt werden. Dieses Netzwerk müsste bei Eingangswerten zwischen 0.5 und 1 den Wert 1 am Ausgang liefern und bei Eingangswerten zwischen 0 und 0.5 einen Ausgangswert von 0. Dadurch wäre die XOR-Funktion exakt implementiert.

Aber dadurch würde man gerade die Stärken des neuronalen Netzes untergraben. Denn ohne dieses nachgeschaltete Netzwerk würde eine „unscharfe“ XOR-Funktionalität simuliert. Es ist nämlich auch möglich, „unscharfe“ Eingangswerte anzulegen. Zwar würde der Ausgangswert dann „schlechter“ werden. Aber das grosse Plus dieses neuronalen Netzwerkes ist es, dass immer noch annähernd die Semantik der XOR-Funktion (die ja normalerweise nur für 0-1-Werte definiert ist) repräsentiert und erkennbar ist. Erst wenn das „Rauschen“ am Eingang zu gross wird und dadurch die Aktivierungsgrade „zufällig“ werden, stellt sich am Ausgang ein „zufälliger“ Wert zwischen 0 und 1 ein.²³ Der Vorteil dieses Ansatzes liegt gerade darin, dass ihm menschliche mentale Verarbeitungsprozesse wie Lernen, Vergessen, Fehlermachen (z.B. bei zu grossem Rauschen) während des Schließens etc. aufgrund seiner „Unschärfe“ bereits inhärent sind. Ausserdem gibt es aufgrund dieser inhärenten Ungenauigkeit bzw. unmöglicher scharfer, exakter Werte – insbesondere durch das Fehlen von Extremwerten wie z.B. der Null – niemals einen „Geteilt durch Null“-Fehler, der in symbolischen Systemen zwangsläufig zu einem Systemabsturz führt (da es nicht definiert ist, durch Null zu teilen), wenn er nicht explizit durch das Programm abgefangen wird. Um ein NN zum Absturz zu bringen, müssen mehr Strukturen des Netzes zerstört werden als von diesem selbständig regeneriert werden können (z.B. bei Alzheimer). Funktionen von Neuronen können bis zu einem bestimmten Grad auch von anderen Neuronen übernommen werden, indem die Struktur des Systems von diesem selbst den Erfordernissen entsprechend angepasst wird.²⁴

²³ Dieses Signal folgt aber nicht dem Zufall, sondern wird durch ein determiniertes Chaos hervorgerufen (vgl. Haken 1982 und Krieger ²1998. Vgl. Abschnitt 3.2.3, S. 46. NN können als dynamische Systeme aufgefasst werden (vgl. Harmonietheorie v. Smolensky 1986 und Smolensky 1988).

²⁴ Dies wird durch Selbstorganisationsprozesse und die Dynamik des Systems sowie der Plastizität der Systemstruktur erreicht. Darauf wird ausführlicher im nächsten Abschnitt über „Dynamische Systeme“ eingegangen.

Es steht auch hier fest: Konnektionistische Systeme sind repräsentationale Systeme. Die verwendeten Subsymbole (auf physischer Ebene) sind Repräsentationen von Merkmalen oder Teilmerkmalen von Konzepten (bzw. der Bedeutung von Symbolen) auf globaler (konzeptueller) Ebene. Jedes Neuron kann als Subsymbol aufgefasst werden. Es repräsentiert einen Teil der Bedeutung eines Gesamtkonzeptes (z.B. der XOR-Funktion). Die Bedeutung des einzelnen Neuron ergibt sich erst während der Verarbeitung dynamisch aus den aktuellen Signalen im Netz und den gespeicherten Gewichtungen der Verbindungen, die sich durch das Trainieren des Netzes ergeben haben.

In Bezug auf unseren Referenzbegriff von Repräsentation aus Abschnitt 3.1.3 (S. 22) lässt sich folgendes sagen: Der Gegenstand der Repräsentation ist die XOR-Funktion. Der Inhalt der Repräsentation ist die Bedeutung der XOR-Funktion, welche durch die Stärke der Gewichtungen der Verbindungen zwischen den Neuronen des Netzes physikalisch implementiert wird. Das neuronale Netz repräsentiert die XOR-Funktion.

Eine Mischform aus konnektionistischen und dynamischen Systemen sind neurodynamische Systeme, welche zwar über eine neuronale Struktur verfügen, die Veränderungen in dieser Struktur und diese Struktur selbst kann man aber als dynamisches System betrachten und mit Hilfe der Theorie dynamischer Systeme beschreiben (vgl. Pasemann 1996). Pasemann geht davon aus, dass auch in neurodynamischen Systemen noch etwas repräsentiert wird, wenn auch „nur“ sowohl räumlich als auch zeitlich verteilt. Es gibt „nur“ noch ein kurzes „Aufblitzen“ von „Repräsentationsbruchstücken“, die zusammengenommen höhere Kognition und somit klassische Repräsentationen erzeugen (vgl. Pasemann 1996 und Metzinger 1998).²⁵ Pasemann scheint also durchaus Repräsentationen anzuerkennen, zwar nicht in einem klassisch-symbolischen Sinne, aber im Sinne der Repräsentation B (vgl. Tabelle 3.1.3, S. 21).

3.2.3 Dynamische Systeme

Die Theorie dynamischer Systeme wird seit langem in der Physik angewandt. Ebenso werden in der Psychologie und Soziologie sowie der Ökonomie dynamische Systeme eingesetzt (vgl. Jäger 1996 und Krieger ²1998). Wettervorhersagen werden

²⁵ Ein ähnliches Konzept (Dynamische Symbolsysteme) legt Jäger (1995) vor (vgl. dazu Abschnitt 3.2.3, S. 50ff).

aufgrund solcher Modelle gewonnen und wurden im Laufe der Zeit optimiert und dadurch immer sicherer in der Prognose.

Ähnliches soll nun mit der menschlichen Kognition geschehen. Das Verhalten und die Überlegungen des Menschen sollen berechnet und vorhergesagt werden. Denn je besser dies gelingt, desto besser sind die Erklärungen bezüglich der Kognition des Menschen. Daher beschäftigen sich immer mehr Kognitionswissenschaftler mit dynamischen Systemen (vgl. Jäger 1996).

Um die Ausführungen so verständlich wie möglich zu gestalten, wird im Folgenden nur die Geometrie dynamischer Systeme, nicht aber deren Algebra dargestellt, obwohl dynamische Systeme innerhalb der mathematischen Theorie dynamischer Systeme auch nur mittels Differentialgleichungen beschrieben werden können. Jäger beschreibt ein dynamisches System informell folgendermassen:

„Ein dynamisches System besteht aus einem kontinuierlichen *Zustandsraum* (synonym: *Phasenraum*), dem eine Dynamik aufgeprägt ist, d.h. eine Gesetzmässigkeit, die festlegt, wie im Laufe der Zeit die Zustände aufeinander folgen. Ein solcher Zustandsraum ist durch reell-wertige Dimensionen x_1, \dots, x_n bestimmt. Ein Zustand ist also ein n -dimensionaler Vektor (x_1, \dots, x_n) . Betrachtet man ein DS als Modell für ein reales System, so repräsentiert jede Dimension eine „Messgrösse“, eine Observable. Man nennt sie auch *Systemvariable*, *Systemgrösse*, *Systemparameter* oder noch anders. Ein Zustandsvektor (x_1, \dots, x_n) gibt also an, welche numerischen Werte zu einem bestimmten Zeitpunkt die beobachteten Systemvariablen einnehmen.“ (Jäger 1996, S. 154).

Die Systemvariablen können sich zeitlich kontinuierlich oder diskret verändern. Man erhält so eine Spur in der Zeit, eine Trajektorie, welche, in einem Diagramm dargestellt, das Phasenportrait oder Phasenbild ergibt. Ein Attraktor als ein bestimmtes Gebiet eines Phasenraumes zieht benachbarte Trajektorien an. Der Anziehungsbereich wird Attraktionsbecken genannt. Punktförmige Attraktoren werden Fixpunktattraktoren bezeichnet und entsprechen „Ruhezuständen“, in denen das System trotz kleinerer Störungen von innen oder aussen stabil bleibt. Es liegt damit ein quasi-stabiler Systemzustand bzw. ein stabiles Systemverhalten vor. Ein zyklischer Attraktor ist eine geschlossene Trajektorie und entspricht „stabilen Schwingungen“ (z.B. der eines Pendels). Auch hier liegt ein stabiles Systemverhalten vor.

Abbildung 3.2.3.1 zeigt den Phasenraum eines dynamischen Systems (eines linear gedämpften fremderregten Schwingers) mit einem zyklischen Attraktor (die geschlossene Trajektorie). Der Phasenraum wird hier aufgrund der zwei Parameter des Systems, Ort x und Geschwindigkeit v , als zweidimensionales Koordinatensystem aufgetragen. Wäre der Schwinger ideal, würde sich seine Trajektorie immer auf der des zyklischen Attraktors befinden.

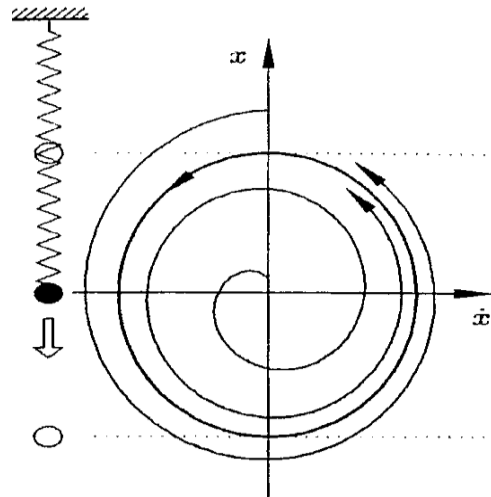


Abbildung 3.2.3.1: Linearer gedämpfter fremderregter Schwinger (aus Argyris 1995)

Da dieser Schwinger aber ein reales System darstellt und somit gedämpft wird, werden Auslenkung und Geschwindigkeit immer geringer, bis der Schwinger schliesslich zur Ruhe kommt. Man könnte aber auch immer wieder Energie zuführen, indem man den Schwinger im jeweils richtigen Moment anstösst, sodass die Trajektorie des Schwingers immer vom zyklischen Attraktor „angezogen“ bleibt (s. die geschlossene Trajektorie in Abb. 3.2.3.1). In komplexen selbstorganisatorischen dynamischen Systemen ist genau dies der Fall (z.B. in autopoietischen Systemen).²⁶

Ist eine Trajektorie nicht zyklisch geschlossen, so liegt ein quasi-periodischer Attraktor vor. Schliesslich spricht man von chaotischen Attraktoren, wenn sie aus unendlich vielen Trajektorien bestehen, die fraktal angeordnet sind, d.h. selbstähnlich sind. In fraktalen Strukturen sind alle Teilstrukturen der gesamten Struktur ähnlich. Chaotische Attraktoren sind nicht vorhersagbar. Das Verhalten chaotischer Systeme ist zwar determiniert, aber dennoch nicht prognostizierbar.

Mit einem Bifurkationspunkt wird der Zeitpunkt in einem dynamischen System bezeichnet, an dem plötzlich ein Wechsel von einem stabilen Systemverhalten in ein anderes erfolgt. Es gibt einen qualitativen Wechsel. Dieser kann nur sprunghaft, nicht aber kontinuierlich erfolgen (Stabilitätsbereich → Bifurkationspunkt → qualitativ neuartiger Stabilitätsbereich). Bifurkation bedeutet einen qualitativen Umschlag und somit etwas Neues (vgl. Jäger 1996, S. 157). Synergieeffekte (wie Bifurkationen) innerhalb von dynamischen Systemen ergeben sich durch ein deterministisches Chaos mit anschliessender Selbstorganisation (vgl. Haken 1982 und Kelso 1995).

²⁶ Vgl. Abschnitt 2.3.5, S. 10.

Deterministisches Chaos wird durch Berechnungsvorschriften (Differentialgleichungen) bestimmt. Geringe Veränderungen der Parameter (semantisch klassifizierte Variablen) in einer Berechnungsvorschrift können sehr grosse Schwankungen im Ergebnis der Berechnung hervorrufen. Sind mehrere Differentialgleichungen an dem Zustandekommen des Ergebnisses beteiligt, so können die Schwankungen im Ergebnis, aufgrund der gegenseitigen dynamischen Beeinflussung der Parameter in den Gleichungen, nicht mehr vorhergesagt werden. Das System befindet sich in einem chaotischen Zustand – allerdings in einem durch Formeln determinierten: im deterministischen Chaos. Innerhalb dieses Chaos können sich die Parameter sehr stark verändern, so dass sich ein ganz anderes Systemverhalten einstellen kann, und damit das Chaos plötzlich wieder verschwindet. Das System organisiert sich innerhalb dieses Prozesses selbst um. Daher wird dieser Vorgang auch Selbstorganisationsprozess genannt. Zwar ist diese Darstellung der Terminologie dynamischer Systeme nicht vollständig. Für die weitere Argumentation in dieser Arbeit sollte sie aber ausreichend sein.²⁷

Viele der neueren Ansätze in der Kognitionswissenschaft versuchen die Interaktion mit der Umwelt zu berücksichtigen, um damit die Erklärungsfähigkeit kognitiver Theorien zu verbessern. Dynamische Systeme sollen computationale, klassisch-symbolische Systeme ablösen und die Kognition besser erklären können (vgl. Port & van Gelder 1995, S. 2). Sie verwenden keine diskreten oder permanenten Symbole. Vielmehr sind Repräsentationen zeitlich an die Information in der repräsentierten Welt gebunden: Wenn der Zustand der repräsentierten Welt sich ändert, dann auch die Repräsentation (vgl. Kelso 1995, S. 45ff („Finger motion“) sowie Clark & Toribio 1994).

Port & van Gelder (1995) sowie Thelen & Smith (1994) argumentieren, dass das Verhalten kognitiver Agenten mit dem Ansatz der Dynamischen System Theorie beschrieben werden kann. Durch die zeitliche Kontinuität der Beschreibung kann die Plastizität der Kognition dargestellt und durch das Aufzeigen von Details in den kognitiven Prozessen auch individuellen Unterschieden (innerhalb der Prozesse in verschiedenen kognitiven Agenten) Rechnung getragen werden. Van Gelder wendet die allgemeine dynamische Systemtheorie auf die menschliche Kognition an und postuliert als Entgegnung auf die Computationalitätshypothese von Newell & Simon und die konnektionistische Subsymbolismus-Hypothese von Smolensky die

²⁷ Für eine weitergehende Einführung in die Terminologie dynamischer Systeme siehe z.B. Jäger 1996, Port & van Gelder 1995 (Kapitel 1 und 2), Guastello 1995 und Haken 1982.

Dynamizitätshypothese, welche besagt, dass kognitive Agenten dynamische Systeme sind:

„The *dynamical hypothesis* (DH) is the unifying essence of dynamical approaches to cognition. It is encapsulated in the simple slogan, *cognitive agents are dynamical systems*.“ (Van Gelder 1998, S. 615)

Der Einwand, dass es trivial ist, die menschliche Kognition als dynamisches System zu bezeichnen, weil jedes System als dynamisches System betrachtet werden kann, ist falsch. Denn erstens geht es darum, dass die *dynamische Systemtheorie* die menschliche Kognition am besten erklären kann, weil es ein *dynamisches System* ist. Und zweitens können nur solche Systeme in vollem Umfange als dynamisch bezeichnet werden, deren Strukturen sich aufgrund von Selbstorganisationsprozessen verändern können. Ausserdem sollten solche Systeme wirklich zeitkritisch arbeiten und die charakteristischen Eigenschaften eines dynamischen Systems aufweisen, d.h. determiniert, komplex, zeitsensitiv, mehrdimensional, gekoppelt und situiert sein (vgl. NNs der 3. Generation in Abschnitt 3.2.2, S. 37ff). Zweifellos sind Gehirne (wie alle Systeme ausreichender Plastizität ihrer Systemstruktur) dynamische Systeme. Diese tragen dem zeitlichen Charakter von Kognition Rechnung, weil sie komplexes zeitliches Verhalten modellieren können.

Die Dynamizitätshypothese besteht aus zwei Teilen. Die Wissenshypothese (Knowledge Hypothesis) besagt, dass kognitive Agenten dynamisch verstanden werden können und sollten, und die Wesenshypothese (Nature Hypothesis) besagt, dass kognitive Agenten dynamische Systeme sind (vgl. Van Gelder 1998, S. 9). Kognitive Agenten sind nicht selbst Systeme, sondern Objekte, deren Eigenschaften Systeme bilden können, d.h. sie instantiieren (während der Kognition) zahlreiche dynamische Systeme. Ein kognitiver Agent ist also nicht ein spezielles dynamisches System, sondern das Konglomerat aller für Kognition notwendigen dynamischen Systeme. Die Variablen sind nicht low-level-Quantitäten, sondern makroskopische Quantitäten – etwa auf dem Niveau der kognitiven Leistung selbst. Kognitive dynamische Systeme verfügen auch über Variablen außerhalb des Agenten (z.B. Parameter der Umwelt, die das System beeinflussen). Es gibt keine qualitativen Zustände mehr auf unterster Ebene, sondern nur noch quantitativ-interagierende Zahlen (vgl. Van Gelder & Port 1995).

Tabelle 3.2.3.1 gibt einen zusammenfassenden Überblick über die wichtigsten Eigenschaften dynamischer Systeme im Vergleich zu symbolischen und subsymbolischen Systemen (übernommen aus Van Gelder 1997 (Table 16.1, S. 434)).

	Computational systems	Connectionist systems	Dynamical systems
Informal description	Symbol manipulators	Networks of neural units	Set of coupled magnitudes
Classic description	Turing machine; LISP machine	Perceptron; Hopfield net	Solar system; Watt governor
Kinds of variable	Digital - often syntactical	Quantitative - activation levels	Quantitative - states and rates
Changes in states	Discrete steps (sequential)	Propagated interaction	Interdependent in "real" time
Tools for description	Transition rules ("programs")	Weighted-sum equations	Differential equations
General character	Interpretable as representations	Homogenous & high-dimensional	Coupled - with environment too

Tabelle 3.2.3.1: „Differences among kinds of systems“ (aus Van Gelder 1997)

Für van Gelder (1995) fällt nur digitale Symbolverarbeitung unter Computation. Neuronale Netze und dynamische Systeme sind für ihn nicht zwingend computational. Van Gelder geht zusätzlich davon aus, dass dynamische Systeme nichts repräsentieren.

Nicht nur anhand des Watt-Reglers wird nach van Gelder die Antirepräsentationalität eines dynamischen Systems demonstriert (vgl. Abschnitt 2.3.1, S. 5f). Auch konkrete, auf Kognition bezogene Modelle sollen dies nach van Gelder zeigen. In Busemeyer & Townsends Entscheidungsfeldtheorie (Decision Field Theory (DFT))-Modell wird die Entscheidungssituation nicht durch Symbole, sondern über kontinuierliche Quantitäten repräsentiert (vgl. Van Gelder 1998, S. 3). Varianten des Haken-Kelso-Bunz (HKB)-Modells für die Koordination des Zeigefingers wurden in verschiedenen kognitiven Domänen angewandt (vgl. Kelso 1995). Koordination wurde dabei nicht durch das Versenden von symbolischen Befehlen (d.h. homunkular) erreicht, sondern entstand als emergente²⁸ Eigenschaft eines dynamischen Systems, das sich selbst um Instabilitäten organisierte (vgl. Van Gelder 1998, S. 4).

Auch an dieser Stelle soll versucht werden, die XOR-Funktion zu implementieren, um die Repräsentationsfähigkeit dieses Ansatzes zu demonstrieren. Im Gegensatz zu vorherigen Implementierungen, werden in einem dynamischen System die Eingangssignale nicht simultan, sondern sequentiell angelegt, um das entsprechende Ausgangssignal zu erhalten. Als Beispiel eines dynamischen Systems wird ein rekurrentes neuronales Elman-Netz verwendet.²⁹

Im Gegensatz zu unserem Feedforward-Netz in Abschnitt 3.2.2 kann sich die Aktivierung innerhalb des Elman-Netzes nicht nur vorwärts, d.h. von den Eingangsneuronen über die versteckten Neuronen zu den Ausgangsneuronen ausbreiten,

²⁸ Vgl. dazu Abschnitt 4.3.3, S. 82f.

²⁹ Vgl. zur Art dieses Netzes auch Fussnote 20 auf S. 37.

sondern in alle Richtungen innerhalb des Netzes, weil jedes Neuron mit jedem anderen Neuron in diesem Netz verbunden ist.

Daraus ergeben sich interessante Eigenschaften. Auch wenn kein Eingang des Netzes aktiv ist, kann das System aufgrund seiner Eigendynamik eine eigene Aktivität entfalten. Externer Input kann diese intrinsische Dynamik des Systems stören. Allerdings kann das System diesen Störungen bis zu einem bestimmten Grad widerstehen oder auch von ihm geführt werden. Das Verhalten des Netzes kann über den Ausgang seine Umgebung beeinflussen, welche über die Eingänge des Netzes wiederum die Dynamik des Netzes beeinflusst usw. Dadurch, dass die Eingangswerte nicht gleichzeitig, sondern sequentiell, angelegt werden, gibt es nur noch einen Eingang und einen Ausgang.

Da in der Realität immer die Zeit eine Rolle spielt, liegen reale Probleme für die Kognition meist zeitversetzt, also sequentiell und nicht gleichzeitig vor. Wir wären vollkommen überfordert, wenn alle Probleme gleichzeitig auftreten würden. Daher arbeiten wir eines nach dem anderen ab.

Dieses dynamische Netz repräsentiert die XOR-Funktion folgendermassen: Das Eingangssignal ist eine Sequenz von Nullen und Einsen. Für die XOR-Funktion haben wir die Sequenzen 00, 01, 10 und 11. Am Ausgang muss dann entsprechend der gewünschten Semantik der XOR-Funktion die Sequenz 0110 erscheinen. Realisiert wird dies mit nur zwei Neuronen. Neuron A wird von jeder Sequenz, die Einsen enthält, aktiviert und verbleibt einige Zeit in diesem Zustand aufgrund einer exhibitorischen Verbindung mit sich selbst. Neuron B wird aktiviert, wenn eine Sequenz mehr als eine Eins hintereinander enthält. Wenn Neuron B aktiviert ist, deaktiviert es über eine inhibitorische Verbindung das Neuron A. Durch diese Verschaltung der Neuronen A und B ergibt sich am Ausgang die gewünschte Sequenz von Einsen und Nullen. Abbildung 3.2.3.2 stellt die Ein- und Ausgangssignale graphisch dar.

Durch die reflexive Verbindung des Neurons A wird dessen Zustand für einige Zeit aufrechterhalten. Es ist also für einige Zeit ein interner stabiler Zustand vorhanden. Das System repräsentiert für diese Zeit das Eingangssignal, das kurz vorher noch angelegen hat. Dies ist ein Beleg dafür, dass innerhalb eines dynamischen Systems mittels eines stabilen Zustandes Repräsentation B vorhanden ist. Dynamische Systeme verfügen also über Repräsentation.

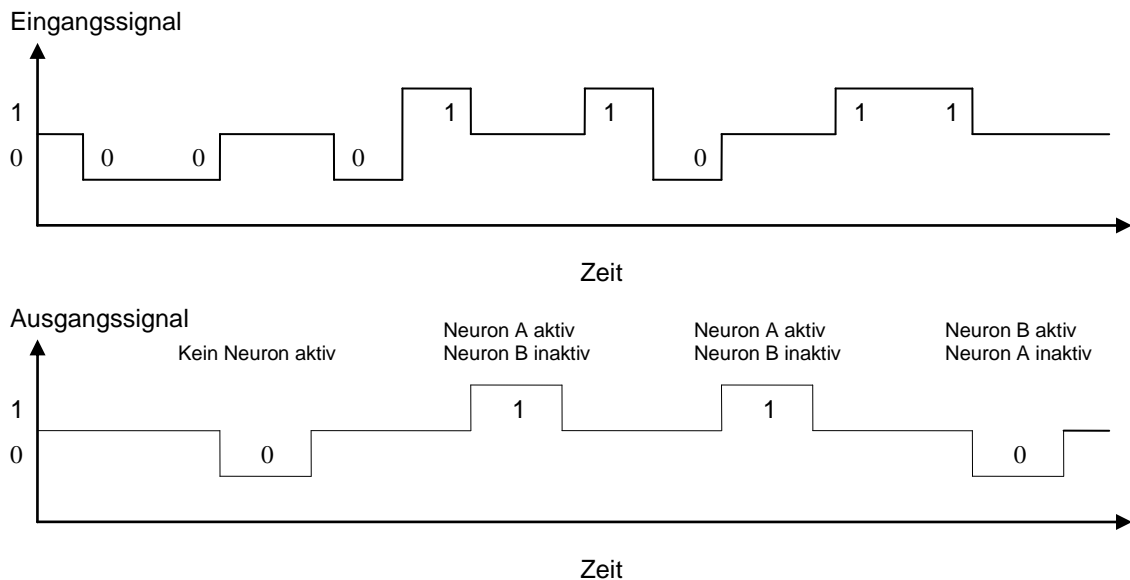


Abbildung 3.2.3.2: Ein- und Ausgangssignale eines neurodynamischen Netzes (Elman-Netz), das die Semantik der XOR-Funktion repräsentiert

Man kann sich denken, dass innerhalb eines Gehirns sehr viele solche Zustände möglich sind und über selbsterhaltende Prozesse auch über sehr lange Zeit aufrechterhalten werden können.

Jäger baut dieses Prinzip innerhalb dynamischer Systeme weiter aus zu einem dynamischen Symbolsystem (DSS), wonach dynamische Systeme dynamische Symbole enthalten können (vgl. Jäger 1996). Ein dynamisches Symbol ist eine beobachtbare Regularität innerhalb eines Informationsverarbeitungssystems:

“Dynamic symbols *are* physically observable, repeatedly identifiable, dynamic regularities in a physical information processing system. They are physical events of some duration, like any other events researched in the natural sciences, e.g., the collision of two elementary particles. A dynamic symbol derives its identity from some empirical observation procedure that reliably indicates the presence of the regularity. A dynamic symbol is *not* a “platonian” entity of any sort. Typical examples are attractor states in neural networks, but transient regularities other than classical attractors make perfect dynamic symbols, too. (Jäger 1995, S. 111)

Allerdings haben dynamische Symbole innerhalb eines DSS nichts mit klassischer Repräsentation zu tun: Einem dynamischen Symbol wird von aussen ein formales Symbol (Label) *ohne* extensionale Bedeutung zugeordnet. Dieses Label dient nur zur Kennzeichnung des dynamischen Symbols, um es wieder zu erkennen, sobald es im System wieder auftritt und um darüber sprechen zu können bzw. um es innerhalb der DSS-Theorie verwenden zu können. Jäger lehnt klassische Repräsentation innerhalb der Theorie dynamischer Symbolsysteme deshalb ab, weil die Bedeutung eines dynamischen Symbols sich nicht über eine Referenz auf die Welt, sondern aus der

internen Dynamik des Systems ergibt, in dem es auftritt. Jäger bezieht sich dabei epistemologisch auf den radikalen Konstruktivismus von Maturana & Varela (1987):

“Dynamical symbol systems combine a symbolic format of „representation“ (without an extensional semantics being intended) with a self-organizing dynamics.” (Jäger 1995, S. 119)

Aber gerade durch die selbstorganisierende Dynamik und die permanente Interaktion dynamischer Systeme mit der Umwelt erlangen m. E. dynamische Symbole im Inneren des Systems externe Semantik, d.h. auf die externe Umwelt bezogene Informationen. Sie können somit als implizite Repräsentation (Repräsentation B) verstanden werden, womit auch dynamische Systeme als repräsentational zu betrachten sind. Ausserdem könnte die permanente Veränderung der Systemstruktur des DSS erklären, wie Begriffswandel möglich ist und als Indiz dafür gelten, dass es keine identischen Konzepte, Begriffe und Vorstellungen geben kann.

Es scheint also möglich zu sein, innerhalb dynamischer Systeme dynamische Symbole oder „Symbolketten“ zu manipulieren. Für Jäger fallen sogar beide Systeme, DSS und die Copycat-Architektur Hofstadters, in die gleiche Klasse der selbstorganisierenden Dynamiken mit symbolischer Repräsentation (vgl. Jäger 1995). Ich denke, dass über solche Prozesse stabile Systemzustände erreicht werden können, die (z.B. als Attraktoren) über sehr lange Zeit aufrechterhalten werden und somit auch als Langzeitgedächtnis (Speicher) fungieren könnten. Die sich aus der Dynamik des Systems ergebenden permanenten Zustände werden zwar ständig aktualisiert und möglicherweise sogar erneuert.³⁰ Aber gerade deshalb kann diese Art der Speicherung als Repräsentation B verstanden werden als Wiederholung stabiler Zustände (Bedeutungen B) (vgl. Tabelle 3.1.3, S. 21).

Jäger schränkt die Möglichkeiten von Dynamischen Symbolsystemen (DSS) ein: Aufgrund der dynamischen Informationsverarbeitung können keine numerischen, logischen oder sonstigen „exakten“ Schlussoperationen durchgeführt werden. Dennoch weisen DSS Zustände auf, denen man Individuen, Klassen und Relationen zuordnen kann. Sie sind sogar in der Lage, einige Probleme der logik-orientierten KI (d.h. einige Probleme des Repräsentationalismus) zu lösen, wozu z.B. terminologische Zirkel, analoge Repräsentationen und nichtmonotone Vererbung³¹ zählen (vgl. Jäger 1995).

³⁰ Ähnlich wie die Informationen (Bits, d.h. 0- oder 1-Zustände) in elektronischen Speichern DRAMs (dynamic random access memory), die im Vergleich zu SRAMs (static random access memory) ständig neu „aufgefrischt“ werden müssen (d.h. alle 10ms von neuem mit Strom versorgt werden), da sie ihre Informationen sonst unwiderruflich verlieren würden.

³¹ Z.B. wird ein Pinguin als Vogel klassifiziert, obwohl er nicht fliegen kann, d.h. ein übliches Merkmal von Vögeln gar nicht hat.

Daher sollten nach Jäger DSS als Möglichkeit betrachtet werden, eine Verbindung von dynamischer Systemtheorie mit klassischer AI anzusteuern:

“[...] DSS contributes to bridging the gap between dynamics-oriented, bottom-up techniques, and representation-oriented, top-down perspectives in AI.” (Jäger 1995, S. 119)

Es lässt sich also auch hier folgendes konstatieren: Dynamische Systeme sind repräsentationale Systeme im Sinne von ‚Repräsentation B‘. Stabile Systemzustände repräsentieren vergangene interne Systemzustände oder externe Zustände der Umwelt. Mit der Zeit wechseln sich diese stabilen Zustände ab und repräsentieren auf diese dynamische Weise insgesamt gesehen bestimmte Semantiken auf Dauer.

Dieses wurde am Beispiel der XOR-Funktion demonstriert. Unser dynamisches System verfügt über Repräsentationen B: Der Gegenstand der Repräsentation ist die XOR-Funktion. Der Inhalt der Repräsentation ist die Bedeutung der XOR-Funktion, welcher durch die dynamische Interaktion von Eingangswerten und die rückgekoppelte Verschaltung von Neuronen eines Elman-Netzes physikalisch implementiert wird. Dieses als dynamisches System betrachtete neuronale Netz repräsentiert die XOR-Funktion.

3.3 Weitere Aspekte des Repräsentationsbegriffs

3.3.1 Gradualität von Repräsentation

Man könnte Repräsentation auch graduell auffassen. Clark & Toribio (1994) behaupten, dass Repräsentation nicht als Dichotomie, sondern als Kontinuum verstanden werden muss. Dieses erstreckt sich von numerischer Repräsentation sensorischer Muster direkt gekoppelter Systeme über die Aktivationsmuster neuronaler Netze als verteilte Repräsentationen bis hin zu explizit symbolischen Repräsentationen in klassischen KI-Systemen.

Ich denke allerdings, dass entweder Repräsentation vorliegt oder eben nicht. Es liegt an der Definition, was man als Repräsentation auffasst. Allerdings zeigt die Forderung nach der Gradualität von Repräsentation auch, dass es verschiedene Kontexte gibt, innerhalb derer man sich des Repräsentationsbegriffes anders bedient. Daher kommen auch die vielen Arten von Repräsentation zustande.

3.3.2 Implementation von Repräsentation

Wenn ein System A auf einer niedrigeren Ebene (z.B. ein neuronales System) dasselbe Verhalten zeigt wie ein System B auf einer höheren Ebene (z.B. ein Symbolsystem) und die Systeme A und B nicht identisch sind, dann implementiert das System A das System B:

„A lower-level system will be said to *implement* a higher-level system when the variables of the latter are somehow constructed out of variables of the former. Note that implementation licenses us to identify the behavior of the one system with the behavior of the other, despite failure of strict identity between the systems themselves.“ (Van Gelder 1998, S. 5)

Implementieren ist der Vorgang, einem bestimmten System eine bestimmte Funktionalität „einzupflanzen“. Man arrangiert die Teile des Systems so, dass sich das gewünschte Systemverhalten ergibt.

Eine Implementierung ist keine Repräsentation, aber sie führt zu einer Repräsentation: Erst die implementierte Funktion eines Systems, die über eine bestimmte Bedeutung verfügt, kann als Repräsentation angesehen werden. Wenn man die XOR-Funktion in einem System implementiert, dann wird die Semantik der XOR-Funktion von diesem System repräsentiert. Es kann also gesagt werden: sobald eine Bedeutung implementiert (d.h. physisch realisiert) wurde, wird diese Bedeutung von dem physischen Träger auch repräsentiert. In Computern erfolgt die Implementation sämtlicher Funktionen auf physischer Ebene mittels der sinnvollen (d.h. bedeutungsvollen) Verschaltung elektronischer Bauteile. In Gehirnen sind die Gehirnfunktionen auf physischer Ebene in einem sinnvoll strukturierten Netz von Neuronen implementiert.

In der zweiten AI-Debatte wurde diskutiert, ob subsymbolische Systeme eine Implementation symbolischer Systeme sind oder nicht (vgl. Clark 2001, S. 97ff und bes. Fodor & Pylyshyn 1988 und Smolensky 1988). Smolensky (1988) vertritt die These, dass konnektionistische Systeme, d.h. subsymbolische Systeme ohne explizite Repräsentationen eine computational-repräsentationale Theorie des Geistes nicht implementieren. Subsymbolische Systeme verfügen über „Subsymbole“, welche mit Symbolen aufgrund ihrer Systemstruktur nicht kompatibel sind (vgl. Kap. 3.2.2, S. 35f).

Ich bin allerdings der Meinung, dass konnektionistische und dynamische Systeme (also subsymbolisch-repräsentationale Systeme) symbolische Systeme nicht nur implementieren, sondern auch über adäquatere Schnittstellen zur Umwelt verfügen: Sie sind umfassender und direkter sowie permanent und in jedem ihrer Prozesse

physikalisch kausal mit der Welt verbunden (vgl. Abschnitt 3.2.1, S. 33). Dabei spielen grundsätzliche Prinzipien eine Rolle (z.B. Kausalität, Selbstorganisation).

3.3.3 Computation und Repräsentation

Nach van Gelder (1998) ist Computation ein effektiver Prozess, in welchem durch einen Algorithmus spezifizierte Operationen systematisch in endlichen und zeitlich diskreten Schritten Startzustände in Endzustände überführt werden. Digitale Computer sind Systeme, die aufgrund eines Algorithmus effektive Computation über Repräsentationen (d.h. über qualitativen Variablen aufgrund Interpretation) ausführen, wodurch ihr Verhalten bestimmt wird. Dynamische Systeme verfügen hingegen über numerische Variablen. Da Zahlen quantitative Eigenschaften haben, sind dynamische kognitive Systeme quantitative Systeme, da es darin zeitliche und zuständige Abstände gibt, welche zu deren Verhalten beitragen. Da dynamischen Systemen die qualitativen Variablen fehlen, können sie weder repräsentational noch computational genannt werden.

Allerdings ist diese Schlussfolgerung zu voreilig. Denn dynamische Systeme verfügen über stabile Systemzustände, die ein bestimmtes Systemverhalten repräsentieren. Wenn ein Mensch steht, dann wird dies, wenn man die Beinbewegungen dieses Menschen in Begriffen der dynamischen Systemtheorie beschreibt, durch einen stabilen Zustand innerhalb des Systems „Mensch“ angezeigt (Fixpunktattraktor). Geht dieser Mensch hingegen spazieren, dann werden seine Beinbewegungen durch eine andere Art von stabilem Zustand innerhalb des Systems „Mensch“ beschrieben (Zyklischer Attraktor). Dynamische Systeme repräsentieren aufgrund stabiler Systemzustände (z.B. Attraktoren) etwas, ob nun äussere, in der Umwelt vorfindbare Zustände oder innere im System befindliche Zustände (vgl. auch Kelso 1995). Dynamische Systeme verfügen somit über Repräsentationen und können daher als repräsentational bezeichnet werden.

Damit kann auch das folgende Argument von Stufflebeam (1998b) entkräftet werden. Er argumentiert, dass Computation auch ohne Repräsentationen möglich ist, weil es nicht nur digitale (qualitative), sondern auch analoge (quantitative) Computation gibt und man in einer Kontinuität nicht weiss, an welcher Stelle man einen Zustand beschreiben kann oder soll. Es können somit keine (diskreten) Repräsentationen ausgemacht werden. Dem kann entgegnet werden, dass sich trotz analoger, kontinuierlicher Computation solcher Systeme „diskrete“ Zustände (stabile

Systemzustände oder stabiles Systemverhalten) feststellen lassen, die als Repräsentation B zu verstehen sind. Stufflebeam gibt eine gewisse Repräsentationalität des Gehirns zu, indem er schliesslich feststellt, dass analoge Verarbeitungsprozesse im Gehirn gerade Repräsentationen wie sprachliche Ausdrücke oder Vorstellungen produzieren und somit nicht ganz repräsentationsfrei sind.

3.3.4 Implizite und explizite Repräsentation

Computer und Gehirne kann man als dynamische Systeme auffassen, da sie auf physischer Ebene zeitkritisch arbeiten. Beschreibt man nun Computer als auch Gehirne in der Terminologie dynamischer Systeme, dann ergibt sich ein identisches Bild für beide Systemarten.

Sowohl Computer als auch Gehirne verfügen über stabile Zustände (Attraktoren), die ein bestimmtes Systemverhalten repräsentieren, und die später wieder gebraucht werden. Dazu gehören auch Vorstellungen, Gedanken, Begriffe usw. Über dynamische Symbolsysteme kann daher auf konzeptueller Ebene im Gehirn wie im Computer auch klassische Symbolverarbeitung (allerdings mit dynamischen Symbolen) durchgeführt werden (vgl. Abschnitt 3.2.3, S. 50ff und Jäger 1995). Auf dieser (konzeptuellen) Ebene muss, durch dessen Entkopplung von der Echtzeitverarbeitung, mit diskreten Einheiten (Symbolen und Variablen) gearbeitet werden. Auf physikalischer Ebene sind aufgrund zeitkritischer Echtzeitverarbeitung nur kontinuierliche Signale möglich.

Man darf daher konzeptuelle und physische Systemebenen eines Systems nicht verwechseln. Computation (analoge, nichtsymbolische Prozessierung) ohne klassische Repräsentationen (Symbole) findet auf der physischen Ebene eines neuronalen Netzes statt. Aber Computation (diskrete, symbolische Prozessierung) findet auch auf der konzeptuellen Ebene eines neuronalen Netzes statt. Und dass unser Gehirn als neuronales Netz über eine konzeptuelle Ebene verfügt, ist offensichtlich, sonst würde ich meine Gedanken hier nicht aufschreiben und der Leser sie nicht verstehen können. Die konzeptuelle Ebene neuronaler Netze kann über die Theorie dynamischer Systeme erklärt werden, wie sich im Folgenden herausstellen wird (vgl. dazu auch die Abschnitte 3.2.3 (S. 44f), 4.3.2 (S. 83f), 3.1.3 (S. 20) und 3.3.4 (S. 58)).

Das Gehirn ist auf physischer Ebene eher vergleichbar mit einem Analogrechner, der nicht mit diskreten Variablen, sondern mit Kontinuitäten umgeht.

Auf konzeptueller Ebene ist ein Gehirn eher mit einem Digitalrechner vergleichbar, der über permanente, diskrete Repräsentationen verfügt. Aber auch auf physischer Ebene können aufgrund dynamischer Prozesse permanente, diskrete Repräsentationen (z.B. Attraktoren) ausgemacht werden.

Beer & Gallagher (1992) entwickelten einen autonomen Agenten, indem sie Repräsentation systematisch durch die Theorie dynamischer Systeme ersetzen. Danach ist die computationale Theorie unpassend, um das Verhalten autonomer Agenten im Allgemeinen zu charakterisieren. Oft braucht man nämlich keine internen Repräsentationen, um zwischen Input und Output zu vermitteln. M. E. beziehen sich Beer & Gallagher dabei auf die physische Ebene.

Aber gerade auf physischer Ebene gibt es innerhalb der Dynamik eines dynamischen Systems ontologisch Entitäten (Attraktoren) die sich von ihrer Umgebung abheben. Über sie kann aber nur auf konzeptueller Ebene gesprochen werden. Auf dieser Ebene ordnen wir physikalischen Entitäten auch deren Bedeutung zu. Interpretation erfolgt auf konzeptueller Ebene. Da aber Gedanken (mentale Inhalte wie z.B. Konzepte und Begriffe) nichts anderes sind als Ansammlungen von Attraktoren, kann man davon ausgehen, dass konzeptuelle und physikalische Ebene eins sind, d.h. Physisches ist nicht das, was keine Intentionalität besitzt, sondern Intentionalität fällt mit Physischem zusammen (vgl. Abschnitt 3.1.3, S. 20: Steinbeispiel und I=M). Möglicherweise könnte man sogar sagen, dass Inhalt und Form ebenfalls zusammenfallen, d.h. Eins sind.

Wir haben also in jedem Fall, d.h. sowohl auf physischer als auch auf konzeptueller Systemebene, stabile Zustände und damit Repräsentation B, über die diskrete Computation möglich ist. Da Vorstellungen, Gedanken, Intentionen usw. diskret sind und als interne Repräsentationen betrachtet werden können, müssen Prozesse, die über diese Entitäten operieren sollen, diskrete Prozesse sein. Also kann Computation, wenn es um Kognition geht, nur diskret sein. Und wie soll mit diskreten Tokens sinnvoll operiert werden, wenn nicht computational? Es gibt also keine Alternativen zu CRUM. Kognition kann weder nicht-repräsentational noch nicht-computational sein. Dies gilt sowohl für die konzeptuelle als auch für die physische Ebene.

Die dritte Debatte in der KI scheint also auf der unterschiedlichen Verwendung des Begriffes „Repräsentation“ und darauf zu beruhen, dass man auf physischer Ebene nur noch nicht lange genug nach Repräsentationen gesucht hat. Auf dieser gibt es nämlich Repräsentationen B. Daher beziehen sich viele Antirepräsentationalisten in

ihren Argumenten nur auf explizite, symbolische Repräsentation, die sie einerseits ablehnen, weil sie den Repräsentationsbegriff anders verwenden und andererseits implizite Repräsentationen dabei völlig ausser Acht lassen. In der ersten Spalte der Tabelle 3.3.4.1 sind die Ebenen aufgeführt, innerhalb der von Antirepräsentationalisten die jeweiligen Erklärungen bezüglich Repräsentationen gemacht werden. Das symbolische und das subsymbolische Paradigma werden von den Antirepräsentationalisten richtig interpretiert. Hinsichtlich dynamischer Systeme begehen sie einen induktiven Fehler: weil sie dort auf der physischen Ebene keine explizite Repräsentation finden können, leugnen sie einfach die Existenz von Repräsentation völlig. Hätten sie aber etwas weiter gesucht, wären sie auf der physischen Ebene desselben Systems fündig geworden (sie hätten implizite Repräsentationen B gefunden) und hätten somit auch erkennen können, dass daher eine antirepräsentationalistische Position nicht haltbar ist.

Erklärungs- ebene	Computations- Paradigma	Interne Repräsentation	Beispiel
konzeptuell	symbolisch-digital	Lokale, diskrete, explizite, symbolische, qualitative Tokens	Klassische AI
physisch	nichtsymbolisch- analog	Verteilte, kontinuierliche Mengen direkter quantitativer Input-Output-Relationen	PDP
physisch	nichtsymbolisch- analog	Stabile(s) Systemzustände und – verhalten, Attraktoren	Klassische AI, PDP, DS
konzeptuell	symbolisch-digital	Dynamische Symbole: diskrete, explizite, symbolische, qualitative Tokens	DS

Tabelle 3.3.4.1: Erklärungsebenen und Repräsentationsarten

Die gestrichelte Linie in der Tabelle 3.3.4.1 soll verdeutlichen, dass die physikalische und die konzeptuelle Ebene in dynamischen Systemen zusammenfallen. Diese Vereinigung dieser Ebenen findet sich auch im Referenzbegriff von Repräsentation B (der auch über Bedeutung B verfügen kann) aus Abschnitt 3.1.3 (S. 21f) wieder. Physikalische und konzeptuelle Ebenen auf ontologischer Ebene vereint zu betrachten, bedeutet möglicherweise einen grossen Schritt, die Lücke zwischen

mentalen und physikalischen Phänomenen ohne Rekurs auf Emergenz und andere problematische Entitäten zu schliessen.

Ich vertrete dabei keinen Dualismus, sondern einen Physikalismus: bestimmte intentionale Termini entsprechen physikalischen Termini, d.h. korrespondieren mit bestimmten physikalischen Phänomenen. Eine Repräsentation korrespondiert mit einem Attraktor. Alle intentionalen Begriffe, die nicht physischen Begriffen entsprechen, können als instrumentalistisch betrachtet werden.

Da sensomotorische Systeme auf Gedanken (als neurodynamische stabile Zustände) zugreifen und weiterverarbeitet werden, können diese neurodynamisch stabilen Zustände als (implizite) Repräsentationen B betrachtet werden.

Nach Eliasmith (1996) könnte man innerhalb dynamischer Systeme auf bestimmte (intentionale) Begriffe zur Erklärung von Kognition verzichten, weil diese Begriffe nicht die (physikalische) Welt beschreiben. Es hätte nur instrumentellen Nutzen, intentionale Rede zu betreiben, nämlich den, über eine konzeptuelle Ebene die Möglichkeit zu haben, über Theorien nachzudenken.

Ich denke jedoch, dass hier ein Missverständnis vorliegt, da physische Termini immer schon intentional sind. Intentionale Begriffe sind reale physikalische Strukturen. Die konzeptuelle Ebene ist keine normative Ebene. Es liegt aber auch kein radikaler Physikalismus vor, da es keine radikale Typidentität gibt, weil kein struktureller Isomorphismus zwischen Gedanken und Attraktoren vorliegt. Vielmehr setzen sich Gedanken (mentale Inhalte, Bedeutung) aus vielen solcher Attraktoren zusammen. Semantik liegt bereits in der ihr korrespondierenden physikalischen Struktur vor bzw. ist bereits durch die ihr korrespondierende physikalische Struktur festgelegt (vgl. Abschnitt 3.1.2, S. 14: ‚Bedeutung B‘, die dann durch ihren Gebrauch aktual wird).

Die einzige Übereinstimmung zwischen Gedanken und Gehirn scheint darin zu liegen, dass beide strukturiert sind. Möglicherweise können systematische (strukturelle) Synergieeffekte dynamischer Systeme aufgrund ihrer Struktursensitivität diese Kluft zwischen Geist und Gehirn überwinden (vgl. Haken 1982 sowie Jäger 1996):

„Bei näherem Hinsehen findet man, dass systemtheoretische Modelle auf zwei Arten verwendet werden können. Die eine Art ist schlicht und einfach die gewohnte reduktionistische. Die andere Art ist zwar in gewissem Sinne nicht-reduktionistisch, sie beruht aber auf mathematisch klaren Universalitätseigenschaften dynamischer Systeme, denen nichts Geheimnisvolles anhaftet. Ausserdem sind nicht-reduktionistische Erklärungen dieser Art nur „oberflächlich“ und einer reduktionistischen Fundierung fähig.“ (Jäger 1996, S. 168)

3.4 Zusammenfassung

Wie bereits deutlich geworden ist, verändert sich die Art der Repräsentation in kognitionswissenschaftlichen Systemen sehr stark hinsichtlich Form und Implementierung. Es scheint, dass Vertreter konnektionistischer und dynamischer Modelle immer mehr von der bisher vorherrschenden Vorstellung abkommen, dass der Kognition explizite Symbole innewohnen – wie dies Kognitivisten (wie z.B. Fodor) vertreten – und stattdessen antirepräsentationalistische Ansätze bevorzugen.

Die Darstellung und Untersuchung der wichtigsten Ansätze der Kognitionswissenschaft zeigte aber, dass es nicht gelingt, ganz von Repräsentationen Abschied zu nehmen. In allen Ansätzen sind Zustände bzw. Entitäten auszumachen, die als Repräsentation bezeichnet werden können, d.h. sie verwenden implizit Repräsentationen B. Sie können daher nicht als antirepräsentationalistisch angesehen werden.

Der allgemeine Repräsentationalismus erfordert nur einen stabilen Zustand (d.h. Repräsentation B), der einen Systemzustand oder ein Systemverhalten repräsentiert, der in einer späteren Operation noch einmal benötigt wird. Daher ist jegliches System, das über Repräsentation B verfügt, ein repräsentationales System.

Man kann innerhalb eines dynamischen Systems erst dann nicht mehr von internen Repräsentationen sprechen, wenn es keine Attraktoren bzw. kein stabiles Systemverhalten mehr gibt. Diese stabilen Zustände sind aber das einzige Ziel dynamischer Systeme (vor allem autopoietischer Systeme wie dem Gehirn).

Antirepräsentationalisten vergleichen die physische Ebene dynamischer Systeme mit der konzeptuellen Ebene des Computers. Da sie innerhalb dynamischer Systeme auf konzeptueller Ebene keine expliziten Repräsentationen finden, vertreten sie einen Antirepräsentationalismus. Dabei übersehen sie aber die Existenz impliziter Repräsentationen auf der physischen Ebene dynamischer Systeme (Attraktoren in dynamischen Systemen).

Daher ist die Position des Antirepräsentationalismus nicht haltbar: dynamische Systeme gehören zu CRUMBS. Es gibt auch keine Alternativen zu CRUM, da Kognition auf konzeptueller Ebene diskret und damit computational sein muss. Dasselbe gilt auch für die physikalische Ebene, da beide Ebenen (die konzeptuelle wie auch die physikalische) auf der physikalischen Ebene (ontologisch) zusammenfallen. Kognition kann also weder nicht-repräsentational noch nicht-computational sein.

4 PROBLEME DES ANTIREPRÄSENTATIONALISMUS

Grundsätzlich weisen alle wissenschaftlichen Fragestellungen zumindest methodologische und ontologische Aspekte auf. Das Hauptziel einer Frage besteht in ihrer inhaltlichen Beantwortung. Dazu muss geklärt werden, wie man dabei vorgehen will, also welche Methode man verwendet, vor welchem Hintergrund man die Untersuchung durchführt und welche Motivation dahinter steckt. Dabei wird man, je tiefer man den Dingen auf den Grund geht, feststellen, dass sowohl die Frage, als auch die Antworten darauf nicht eindeutig zu formulieren sind, weil immer mehr und tiefliegendere Aspekte hinzukommen und eine Rolle spielen, bis man schliesslich bei metaphysischen Annahmen ankommt.

4.1 Methodologische Probleme

4.1.1 Grundsätzliche Probleme

Wenn keine theorienübergreifenden Begriffsdefinitionen vorhanden sind, führt dies im wissenschaftlichen Diskurs zu Problemen. Nach Stufflebeam kommt die Ablehnung interner Repräsentationen oft daher, weil die Definitionen von Repräsentation von Theorie zu Theorie variieren und es nicht offensichtlich ist, dass alles, was als Repräsentation bezeichnet wird, auch diesen Namen verdient (vgl. Stufflebeam 1998a).

Meines Erachtens ist dies eine zu starke Behauptung, weil sie die Gefahr der wissenschaftlichen Unfruchtbarkeit in sich birgt. Denn so könnte man mit jedem ungenauen Begriff verfahren. Und da zumindest interdisziplinär Begriffe aufgrund unterschiedlicher Beschreibungssprachen oft unterschiedlich verwendet werden können, könnte man mit so einer Begründung jegliche interdisziplinäre wissenschaftliche Forschung einstellen.

Elektrochemische Prozesse im Gehirn können mit psychologischem Vokabular nur grob beschrieben werden, was mit physikalischem Vokabular hingegen sehr detailliert möglich ist. Es würde wesentlich länger dauern, wenn man innerhalb von Erklärungen bezüglich Kognition anstelle der Verwendung des psychologischen Begriffs „Repräsentation“ viele physikalische Begriffe heranziehen würde wie z.B.: „Ein stabiler Zustand innerhalb eines dynamischen Systems, der sich aufgrund von

Selbstorganisationsprozessen und der Interaktion des Systems mit der Umwelt aufgrund bestimmter Zustände ausserhalb des Systems (d.h. in der Umwelt) einstellte, entspricht nach wie vor diesen bestimmten Zuständen ausserhalb des Systems“. Man könnte diesen Zustand in der Terminologie dynamischer Systeme kurz „Attraktor“ nennen. „Attraktor“ kann aufgrund des Referenzbegriffs von Repräsentation aus Abschnitt 3.1.3 auch „Repräsentation B“ genannt werden. Dass selbst Wissenschaftler, die sich als Antirepräsentationalisten bezeichnen würden, dennoch so etwas wie Repräsentationen in ihren Theorien zulassen, bestätigt nur diese Auffassung (vgl. Abschnitt 3.3.3, S. 54ff und z.B. Stufflebeam 1998a und Van Gelder 1995).

Ein weiteres grundsätzliches Problem aller theoretischen Modelle ist es, dass sie die Wirklichkeit niemals vollständig abbilden, sondern diese lediglich abstrahiert, idealisiert und sehr vereinfacht (und daher möglicherweise verzerrt oder sogar inadäquat) darstellen.

Da dieses Problem aber alle Theorien trifft, kann man es dem Antirepräsentationalismus nicht anlasten. Das Problem ist auch aus einem anderen Grund nicht so bedrückend. Denn solange pragmatische Schlüsse, die für das Überleben notwendig sind, in der Regel richtig sind, reicht auch ein unvollkommenes, abstrahiertes Abbild der Welt völlig aus. Man findet sich in der Welt auch so zurecht. Dennoch könnte man innerhalb der Modelltheorie davon ausgehen, dass eine „vollständige Repräsentation der Welt“ gar nicht möglich ist.

Ein anderes Problem ergibt sich daraus, dass viele Vertreter antirepräsentationalistischer Ansätze von künstlichen Modellen ausgehen und Ergebnisse aus deren Untersuchungen dann auf die menschliche Kognition übertragen. Diese Vorgehensweise ist aber höchst problematisch, denn dies bedeutet noch lange nicht, dass die menschliche Kognition genauso funktioniert wie die simulierte Kognition im Computer.

Nach Newell & Simon (1981) ist die klassische AI-Forschung als empirische Wissenschaft zu verstehen. Aber alle empirischen Fragen müssen induktiv durch wissenschaftliche Experimente bestätigt werden und können nicht a priori entschieden werden. Empirische Fragen können keine Allgemeingültigkeit nach sich ziehen. Apriori-Argumente haben dafür wesentlich bessere Karten. Aber welche apriori-Gründe für oder gegen einen Antirepräsentationalismus gibt es?

Ein solcher Grund könnte sein, dass es unmöglich ist, alle Erklärungslücken zu schliessen. Da sich Vertreter symbolischer Systeme explanatorisch „top-down“ von abstrakten Ebenen zur konkreten, materiellen Ebene vorarbeiten und Vertreter

subsymbolischer Systeme explanatorisch „bottom-up“ vom Gehirn zur Kognition vordringen, wird die Erklärungslücke zwischen Gedanken und Gehirn scheinbar immer kleiner. Aber andererseits bleiben immer noch und immer wieder neue Lücken. Die grössten Probleme liegen dabei im Schliessen von Erklärungslücken während des Integrationsvorganges von verschiedenen Ansätzen, Analyseebenen (Makro-Mikroebene des Geistes) und Disziplinen aufgrund unterschiedlicher Beschreibungsebenen und Terminologie (vgl. Clark 2001). In der Physik werden z.B. immer wieder kleinere Teilchen entdeckt und dadurch immer wieder neue Fragen aufgeworfen, die wiederum auf Antworten warten. Ähnlich scheint es in allen anderen wissenschaftlichen Bereichen zu sein – auch in der Kognitionsforschung. Es könnte also sein, dass die sogenannte „Low-High-Level-Cognition-Gap“ nie geschlossen werden kann. Bevor dies noch genauer untersucht wird, wird noch kurz auf das Problem der Unterschiedlichkeit von Beschreibungsebenen eingegangen.

4.1.2 Unterschiedliche Beschreibungsebenen

Die antirepräsentationalistische Position ist laut Clark & Toribio (1994) aus drei Gründen schwach. Sie besage erstens, dass Systeme mit expliziten Repräsentationen mit der Verkopplung der Umwelt in Echtzeit nicht zurechtkommen, beachte implizite Repräsentationen aber dabei gar nicht. Die Ergebnisse aus Kapitel 3 dieser Arbeit bestätigen dies: Es lässt sich innerhalb dynamischer Systeme immer etwas finden, das als Repräsentation bezeichnet werden könnte (vgl. besonders die Abschnitte 3.1.3, S. 16ff und 3.3.4, S. 55ff).

Zweitens mache sie keinen Unterschied zwischen symbolischer Repräsentation und allgemeiner Repräsentation. Diese Unterschiede etwas zu erhellen wird meines Erachtens aber zu wenig forciert. Weder Clark & Toribio (1994) noch ihre Gegner schaffen es, auf eine eindeutige Art die Verwendungsweise der Begriffe „Repräsentation“ und „Repräsentationen“ darzulegen. Dadurch entstehen Argumente, deren Argumentationskraft sich nur schwer ausmachen lässt. In Kapitel 3 wurden diese begrifflichen Probleme näher beleuchtet und versucht, sie gegeneinander abzugrenzen. Dabei kristallisierte sich der Repräsentationsbegriff ‚Repräsentation B‘ heraus. Repräsentationen dieser Art sind selbst noch in dynamischen Systemen zu finden. Aufgrund der verschiedenen Disziplinen, die an der Kognitionsforschung teilnehmen, wird sich aber eine einheitliche Terminologie schwer und wenn, dann nur langfristig durchsetzen lassen.

Drittens würden Antirepräsentationalisten zu Unrecht behaupten, dass ihre Systeme nicht repräsentational wären. Der von den Dynamizisten als antirepräsentationalistisch bezeichnete Ansatz dynamischer Systeme ist nach Clark & Toribio (1994) nicht antirepräsentational, sondern repräsentational, weil deren Tools nicht intrinsisch nicht-repräsentational sind.

Denn eine Erklärung über dynamische Systemtheorie kann durchaus repräsentational beladen sein, insbesondere mit zunehmender Dimensionalität des Systems, d.h. je mehr Parameter eine Rolle spielen und je komplexer die Dynamik des Systems wird. Denn dann ergeben sich immer mehr stabile Zustände, die als Repräsentationen bezeichnet werden können (Clark & Toribio 1994, S. 423). Clark & Toribio (1994) lehnen daher die radikale dynamische Systeme–These kognitiver Systeme ab, welche besagt, dass Kognition Evolution von Zustandsräumen in Watt-Regler-artigen Systemen sei.

Ich stimme dem in dem folgenden Sinne zu: Watt-Regler-artigen dynamischen Systemen fehlt jegliche Plastizität. Dadurch können sie nicht über genügend interne Repräsentationen verfügen, um hohe kognitive Fähigkeiten zu erlangen. Aber die Gehirne der Menschen haben sich im Laufe der Evolution zu optimalen plastischen dynamischen Systemen entwickelt, die die höchste Kognition hervorbringen, von der wir momentan empirisch wissen. Aufgrund ihrer hohen Plastizität können sie viele interne Repräsentationen haben und so menschliche Kognition verrichten. Und dynamische Systeme dieser Komplexität und Plastizität verfügen über viele Attraktoren und somit massiv über Repräsentationen B.

Clark und Toribio (1994) meinen, dass die Art des Antirepräsentationalismus, wie ihn Theoretiker dynamischer Systeme vertreten, übertrieben ist und aufgrund einer Vermengung des allgemeinen Repräsentationsbegriffes mit vielfältigen restriktiven expliziten Repräsentationsbegriffen und solchen, die intuitive bekannte Inhalte tragen, entstanden ist. Ausserdem beziehen sich Antirepräsentationalisten nur auf explizite interne Repräsentationen, nicht auf implizite interne Repräsentationen (was in Abschnitt 3.3.4 dieser Arbeit bestätigt wurde). Nach Clark & Toribio (1994) geht Brooks von der Ablehnung expliziter Repräsentationen dazu über, Repräsentationen im Allgemeinen zu leugnen. Heideggers Angriffe zielen auch hauptsächlich auf explizite Repräsentationen ab. Auch Beer konzentriert sich hauptsächlich auf explizite Repräsentation, indem er zeigt, dass keine solchen für die Funktion seines Systems nötig sind. Daher bleiben die Fragen bezüglich impliziter Repräsentationen offen, womit ein radikaler Antirepräsentationalismus nicht haltbar ist.

Dieses Argument von Clark & Toribio ist m. E. aber nicht sehr stichhaltig. Denn es könnte doch Antirepräsentationalisten geben, die auch erfolgreich gegen implizite Repräsentationen argumentieren.

Clark & Toribio (1994) favorisieren „modest representations“. Jeder Prozess, in dem ein physikalisch definierter Eingangsraum so transformiert wird, dass einige Merkmale unterdrückt und andere hervorgehoben werden, ist eine Instanz von „modest representations“. Je grösser der computationale Aufwand ist, solche Transformationen durchzuführen, desto repräsentationalistischer ist die Lösung der Transformation. Der Repräsentationsbegriff ist dabei kontinuierlich zu verstehen: Es gibt ein Kontinuum von Repräsentationen. Dieses reicht von klassischer symbolischer Repräsentation über subsymbolische Repräsentation bis hin zur Fähigkeit, Muster unterscheiden zu können. Hier ist die Grenze erreicht, von der man gerade noch von „modest representation“ sprechen kann.³² Man könnte sagen: Je *computationaler* ein System im klassischen Sinne ist (d.h. Manipulation von internen Repräsentationen), desto repräsentationalistischer ist es.

Ich halte dagegen an dem in Kapitel 3.1.3 definierten Repräsentationsbegriff ‚Repräsentation B‘ fest, der den Bereich der „modest representations“ mit einschliesst und auch für dynamische Systeme geeignet ist, wie dort gezeigt wird.

Im Gegensatz zu Clark & Toribio (1994) leugnet Eliasmith (1996) sogar die explanative Kraft dynamischer Systeme. Zwar bieten dynamische Systeme schöne Metaphern und interessante Diskussionen und geben dadurch eine normative Richtung innerhalb der Kognitionswissenschaft vor. Sie verfügen aber weder über ein Modell noch über eine Definition bezüglich des Verhaltens oder der Kognition des Menschen.

Dem muss aber entgegengehalten werden, dass es bisher überhaupt noch keine umfassende Theorie der Kognition gibt. Auch Repräsentationalisten haben eine solche noch nicht hervorgebracht. Und dafür, dass sich die Dynamizisten erst seit etwa 1990 mit Kognition ausführlich beschäftigen, ist ihr Erfolg sehr beachtlich, da es bereits dynamische Modelle gibt, die bestimmte Bereiche der Kognition beschreiben.

Die „Decision Field Theory“ (DFT) von Busemeyer & Townsend (1993) deckt einen breiteren Bereich von Phänomenen detaillierter ab als klassische Nutzenstheorien und erklärt sogar Paradoxien aus diesem Bereich (vgl. Van Gelder 1995, S. 362). DFT

³² Hier ist m. E. die ICA (independent cluster analysis) angesiedelt. Dies ist eine Methode, um aus einer unbekannt Menge von Eingangsdaten voneinander unabhängige Muster erkennen zu können (sogenannte „nichtlineare Merkmalsextraktion“). Diese Methode könnte möglicherweise als Naturprinzip verstanden werden ebenso wie Selbstorganisationsprozesse oder Autopoiesie („Selbstorganisationprozess lebender Organismen“).

arbeitet auf denselben Prinzipien wie die „Motivational oscillatory theory“ (MOT) von Townsend, welche wiederum nach den dynamischen Prinzipien des Watt-Reglers (vgl. Tabelle 2.3.2, S. 6) funktioniert. In MOT gibt es nur wenige Variablen für den jeweils aktuellen Zustand von Motivation, Befriedigung, Präferenz und Handlung (Bewegung). Mit nur vier Differentialgleichungen lässt sich beschreiben, wie sich diese Variablen über die Zeit als eine Funktion des gegenwärtigen Zustands des Systems verändern.³³ MOT steht in ähnlicher Weise zu Nutzenstheorien wie der Watt-Regler zum computationalen Regler. Kognition ist nicht die Manipulation von Symbolen, sondern Evolution des Zustandsraumes innerhalb eines dynamischen Systems. Aber es gibt nirgends im System irgendwelche diskreten internen Geschehnisse, die man als „Entscheidung“ charakterisieren könnte. Und dennoch entscheidet ein MOT-Agent. Entscheidungen in MOT ergeben sich als Verhalten eines Agenten unter dem Einfluss eines Hin und Her aus erwünschten und unerwünschten Ergebnissen sowie internen Wünschen und Motivationen (vgl. Van Gelder 1995, S. 362).

M. E. macht es aber doch nur Sinn, wenn bestimmte Entscheidungen über einen bestimmten Zeitraum Gültigkeit haben, um so die Zeit zu haben, diese Entscheidung in einer Handlung umzusetzen. Entscheidungen müssen also zumindest stabile Zustände innerhalb des dynamischen Systems sein. Und solche Zustände sind Repräsentationen im Sinne von Repräsentation B.

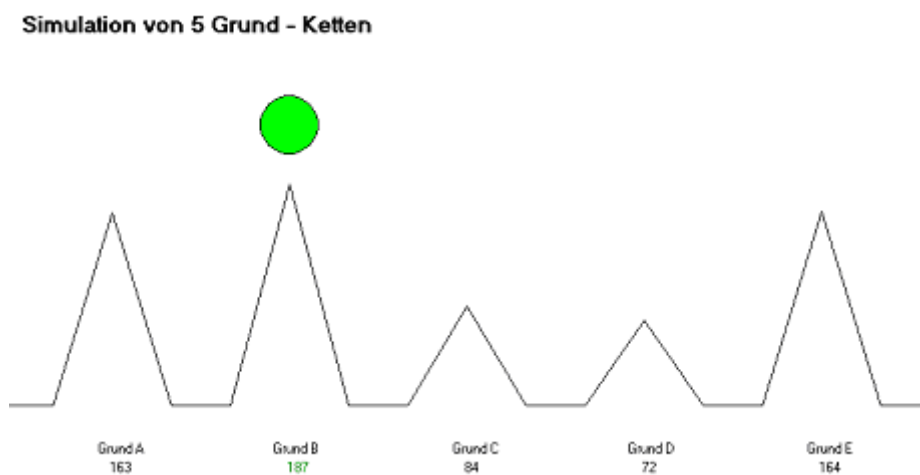


Abbildung 4.1.2.1: Prinzip der Entscheidungsfindung

³³ Van Gelder (1995) beschreibt diese vier Differentialgleichungen u.a. informell in Fussnote 9 auf S. 361 wie folgt: „[...] The change in motivation depends on how the current levels of motivation and of consumption compare with some standard level of motivation. [...] The change in one’s *preference* for the goal will depend on current motivation and one’s distance from the object of preference. [...] The change in consumption will depend on the level of preference, the level of consumption, and the distance from the object of preference. [...] How one moves toward or away from the object depends on one’s current level of preference for the object. [...]”

Dynamische Systeme scheinen für die Erklärung von Entscheidungen, die ständig von inneren und äusseren Faktoren bestimmt werden, sehr gut geeignet zu sein, weil sie kontinuierliche Systeme sind und eine Verarbeitung in Echtzeit leisten. Abbildung 4.1.2.1 veranschaulicht in vereinfachter Form das Prinzip der Entscheidungsfindung: sie stellt eine Momentaufnahme aller Gründe und deren Bewertung dar. Die höchste Präferenz resultiert aus dem Grund B, wonach die entsprechende Entscheidung fällt (mit einem Kreis markiert). Wären zwei Präferenzen gleich stark, würde in einem der nächsten Zeitpunkte bereits wieder eine eindeutige Entscheidung vorliegen, da sich zu jedem Zeitpunkt alle 5 Präferenzen in ihrer Bewertung ändern (wenn auch nur geringfügig). Aber solange eine Entscheidung vorliegt, d.h. solange ein Grund höchste Präferenz hat, liegt auch ein stabiler Zustand innerhalb dieses Systems vor, der diese Entscheidung repräsentiert und den man als Repräsentation B bezeichnen kann.

4.1.3 Low-High-Level-Cognition-Gap

Das grösste Problem scheint zu sein, die Kluft zwischen Low- und High-Level-Cognition zu schliessen. Antirepräsentationalistische Systeme sind der Low-Level-Cognition fähig. Brooks Mobots können Hindernissen ausweichen, auf bestimmte Ziele zugehen und Objekte greifen. Repräsentationalistische Systeme hingegen können höherwertige, schwierige High-Level-Cognition Aufgaben verrichten, indem sie über klassische interne Repräsentationen (Symbole) verfügen. Schachcomputer gewinnen gegen den Menschen und der Taschenrechner löst schwierige mathematische Aufgaben schneller und korrekter als der Mensch. Es gibt intelligente Planungssysteme und Expertensysteme, die automatisch Flugreservierungen vornehmen oder medizinische Diagnosen stellen. Diese Beispiele sprechen dafür, dass klassische Symbolsysteme für High-Level-Kognition gut geeignet sind.

Allerdings ist fraglich, ob damit bereits wirklich echte High-Level-Cognition erreicht ist oder nicht. Ich denke, dass der Mensch zu wesentlich höheren, komplexeren kognitiven Aufgaben in der Lage ist. Dazu gehört das Abrufen adäquaten Weltwissens in bestimmten Situationen ebenso wie eine angemessene Ausdrucksweise, das Verstehen einer Situation, eines Ausdrucks, eines Wortes, eines Bildes ebenso wie die Fähigkeit, sich in andere Situationen zu versetzen. Auch das Erkennen komplexer Zusammenhänge oder philosophisches Denken sind Merkmale hoher Kognition.

Bisher bringt aber *kein* künstliches System – weder antirepräsentationalistisch noch repräsentationalistisch – echte menschliche Kognition hervor. Dies liegt nicht nur daran, dass diese Systeme immer noch in den Kinderschuhen stecken, sondern dass die Komplexität menschlicher Kognition (aufgrund Informationsmangels innerhalb des Designprozesses) nicht „nachgebaut“ werden kann.

Es scheint daher prinzipiell unmöglich zu sein, die Kluft zwischen Low- und High-Level-Cognition zu schliessen. Andererseits stellt sich die Frage, ob es überhaupt möglich bzw. legitim ist, zwischen Low- und High-Level-Cognition zu unterscheiden und damit zu so gravierenden Schlussfolgerungen zu gelangen. Vielmehr scheint Kognition graduell zu sein – und durch die Evolution der Natur entstanden. Aber auch dann ist die Schlussfolgerung die gleiche: Da der Mensch Teil der Natur ist, kann er Kognition in letzter Perfektion nicht „nachbauen“.³⁴

Graduelle Repräsentation (oder „modest representations“) scheint für graduelle Kognition sehr gut zu passen. Für High-Level-Kognition wird bevorzugt klassische, symbolische Repräsentation verwendet. Mit abnehmender, expliziter kognitiver Leistungen wird auch weniger explizit-symbolische Repräsentation, sondern eher subsymbolische Repräsentation benutzt. Bis schliesslich auf unterster Low-Level-Kognitions-Ebene keine klassische Repräsentation mehr zu finden ist, sondern höchstens noch Repräsentation B. Repräsentation verschwindet so mit abnehmender kognitiver Leistung sowohl aus einem explanatorischen als auch aus einem ontologischen Blickwinkel.

4.1.4 Systemtheoretische Probleme

Alle (künstlichen) systemtheoretischen Modelle werfen Probleme mit ihrer Initialisierung und ihrer Parametrisierung auf. Ein entscheidendes Problem jedes Systems stellt die Bestimmung seiner Systemparameter dar. Aber selbst wenn alle relevanten Systemparameter bekannt wären, stellt sich die Frage nach deren initialer Wertebelegung. Je nach Initialisierung des Systems ergeben sich verschiedene Verhaltensmuster des Systems. Ferner müssen die Systemgrenzen und somit die Systemgrösse irgendwie festgelegt werden. Dies ist aber problematisch, weil vor diesem Schritt noch keine Anhaltspunkte dafür zur Verfügung stehen. Die Parameter- und Wertebestimmung ist nur über die Try-and-Error-Methode oder über Analogien

³⁴ Denn der Mensch hängt mit der Implementierung der künstlichen Intelligenz (Kognition) immer seiner eigenen hinterher. Ausserdem kann er nichts in der Welt jemals *vollständig* beschreiben und erklären, also auch nicht die Kognition.

möglich. Ein solches System kann daher nicht a priori komplett vorgelegt werden, sondern es werden immer empirische Ergebnisse für seine allmähliche Entwicklung benötigt. Es gibt auch keine vordefinierte Art und Weise, die Gleichungen festzulegen oder Daten zu interpretieren. Dadurch ist es schwierig, die Bedeutung der Parameter innerhalb eines Systems herauszufinden (vgl. Eliasmith 1996).

Kollektive Parameter tragen dem makroskopischen Verhalten eines hochdimensionalen Systems (Kognition) durch ein niedrigdimensionales Modell Rechnung. Dadurch kann mittels eines dynamischen Systemtheoriemodells mit nur wenigen Parametern ein hochdimensionales System mit vielen (emergenten) Eigenschaften beschrieben werden (z.B. die „Motivational oscillatory theory“ (MOT) in Abschnitt 4.1.2, S. 65).

Die dynamische Interpretation der Kognition ist aus diesen Gründen weder trivial noch den anderen Paradigmen wie Konnektionismus oder Symbolizismus vorzuziehen.

Allerdings treten diese Parametrisierungsprobleme nur auf, wenn das System von einem menschlichen Designer entworfen wird. Systeme, die wir in der Welt bereits vorfinden (bzw. als solche erkennen), sind evolutionär entstanden. Die Parameter und deren Belegung brauchen nur noch „abgelesen“ zu werden. Es haben sich auch Prinzipien in der Natur „gezeigt“, die vermuten lassen, dass sich Systeme auch selbst organisieren können, so dass keine Parametrisierung von aussen vorgenommen werden muss (vgl. Synergetik: Haken 1982). Diese ergibt sich vielmehr von selbst aufgrund der ständigen Veränderungen der Strukturen in der Welt. In dynamischen Systemen und selbstorganisierenden Karten³⁵ werden daher synergetische und selbstorganisierende Prozesse für die selbstständige Parametrisierung genutzt (vgl. Kelso 1995). Aus der selbständigen Organisation der Struktur des Systems ergeben sich emergente Eigenschaften. Dies sind Eigenschaften, die aufgrund einer Strukturveränderung auftreten. Ein System entwickelt sich in der Zeit und besteht aus Teilen, „so dass das Ganze mehr als die Summe der Teile ist“ (Jäger 1996, S. 152). Diese Differenz wird qualitativ mittels emergenter Eigenschaften beschrieben (vgl. dazu auch Abschnitt 4.3.3, S. 82ff).

Ein weiteres systemtheoretisches Problem betrifft die Komplexität des Systems und die Berechenbarkeit der Differentialgleichungen. Bereits einfache dynamische Systeme Modelle stellen an die Rechner, auf denen sie simuliert werden, hohe

³⁵ „Self organized maps“ sind neuronale Netze, die die *markantesten* Eigenschaften innerhalb eines Eingangsvektors am schnellsten lernen.

Anforderungen. Ungleich grössere Systeme mit vielen Differentialgleichungen werden zur Wetterprognose eingesetzt. Auch hier hängt die Prognosegenauigkeit und –länge hauptsächlich von der Rechenleistung eingesetzter Grossrechner ab.

Aus beiden Beispielen lässt sich ersehen, dass hohe menschliche Kognition zu komplex und unberechenbar für einen Rechner wird. Es gibt komplexe Probleme, die eine exponentielle Berechenbarkeitsfunktion aufweisen. Mit linearer Zunahme der Komplexität des Problems (abhängig z.B. von der Anzahl der Parameter eines dynamischen Systems) nimmt die Rechenzeit für dessen Lösung exponentiell zu. Natürliche Sprache, die sicher zur höheren Kognition des Menschen gehört, fällt bereits in diesen Bereich nicht-handhabbarer berechenbarer Probleme. Wie sieht es da wohl erst mit der menschlichen Kognition insgesamt aus? Es scheint so, als wenn selbst systemtheoretische Modelle, die einmal den Anspruch stellen könnten, die menschliche Kognition zu erklären, aus obigen Gründen nie in ihrer Vollständigkeit simuliert werden könnten. Dies scheint aber nur so. Denn aufgrund des Prinzips der Emergenz kann die Komplexität eines Systems von Anfang an klein gehalten werden (vgl. MOT-Modell in Abschnitt 4.1.2, S. 65). Dies ist erstaunlich, denn gerade aus der Dynamik des Systems ergeben sich solche emergente, qualitative Eigenschaften durch Bifurkationen (vgl. Abschnitt 3.2.3, S. 45). Und je komplexer das System ist, desto grösser kann seine Dynamik sein. Es wird also scheinbar Komplexität benutzt, um Komplexität zu verringern.

4.2 Inhaltliche Probleme

4.2.1 Kognition ohne Repräsentation

Im Folgenden wird auf Probleme eingegangen, die auftreten, wenn man Kognition ohne Repräsentation erklären will. Gegner des Antirepräsentationalismus kennen bestimmte kognitive Fähigkeiten des Menschen, die nicht ohne die Annahme von internen Repräsentationen möglich sind. Dazu gehören vor allem Planung, Memorierung, kontrafaktisches Schliessen, Benutzung abstrakter Entitäten und Sprache. Grundsätzlich lassen sich dabei zwei Kategorien von Problemklassen ausmachen: das Abwesende und das Abstrakte. Es ist ein Problem zu erklären, wie der Mensch mit der Umwelt zurechtkommt, ohne über interne Repräsentationen (oder Modelle über die Welt) zu verfügen. Ebenso schwierig ist es, zu erklären, wie der Mensch momentan

nicht vorhandene Sachverhalte in sein (künftiges) Verhalten mit einbeziehen kann. Kognitive Agenten können laut Clark & Toribio (1994) in „repräsentationshungrigen“ Fällen, in denen sie mit entfernten, nicht existenten oder hoch abstrakten Dingen zurechtkommen müssen, ohne interne Repräsentationen nicht auskommen. Dazu zählen z.B. das Nachdenken über Einhörner und Primzahlen usw. sowie schnelles kontrafaktisches Schliessen in Echtzeit und das Reagieren auf Input, dessen Erkennen abstraktes Denken erfordert.

Der Versuch, die menschliche Kognition ohne jegliche interne Repräsentation zu erklären, ist nach Eliasmith (1996) bereits im Behaviorismus gescheitert. Allerdings impliziert dies m. E. nicht notwendig das Scheitern neuerer antirepräsentationalistischer Ansätze (wie dynamischer Systeme). Denn diese neuen Ansätze können die weiteren Entwicklungen der wissenschaftlichen Forschung seither mit berücksichtigen.

Nach Clark & Toribio (1994) untersuchen Antirepräsentationalisten ausserdem nur die Fälle, bei denen die umgebenden Umweltinformationen dazu ausreichen, damit kognitive Agenten komplexes, aber von Repräsentationen unabhängiges Verhalten zeigen, wonach die These der Antirepräsentationalisten, dass Kognition ohne interne Repräsentation möglich ist, haltbar wäre.

Dem ist aber meiner Ansicht nach nicht so. Denn Busemeyer & Townsend (1993) zeigen mit ihrem DFT-Modell, dass vorausschauende, planende Entscheidungen innerhalb dynamischer Systeme möglich sind. Allerdings übersehen sie die implizite Repräsentation B in ihrem System. Diese Tatsache stützt wiederum die obige These von Clark & Toribio insofern, als Busemeyer & Townsend unbewusst zu „Repräsentationalisten“ werden, wenn sich in ihrem System implizite Repräsentationen befinden (was ich in Kap. 3.3.4, S. 55ff auch gezeigt habe).

Eliasmith (1996) behauptet, dass Mobots ohne interne Repräsentationen keine höheren kognitiven Leistungen erbringen können, sondern nur Probleme der Low-Level-Kognition lösen. Ausserdem würden sie auch kein komplexes Verhalten an den Tag legen. Erst wenn die interne (subsumption) Architektur komplexer würde, könnten sie flexibler werden und High-Level-Kognition leisten.

Ich denke auch, dass Repräsentation für High-Level-Kognition notwendig ist. Allerdings bestreite ich, dass Mobots mit Subsumptionsarchitektur für High-Level-Kognition jemals geeignet sind, da diese Architektur nicht flexibel sein kann, wenn sie vom Designer unveränderlich vorgegeben ist. Und weil diese sich nicht selbst verändern kann (wie z.B. die Gehirnstruktur), bleibt sie starr. Daher werden Mobots kaum hohe

Kognition erreichen, denn dazu muss sich das gesamte System selbständig und permanent der Umwelt anpassen können.

Mobots verfügen auch nicht über Variablen, die zur kurzfristigen Speicherung von Zahlenwerten dienen könnten, sondern über kontinuierliche (analoge) Signale, denen kein semantischer Gehalt zugeordnet werden kann. Brooks lehnt daher explizite Repräsentationen innerhalb seiner Mobots ab:

„We never use tokens which have any semantics that can be attached to them. The best that can be said in our implementation is that one number is passed from a process to another.” (Brooks 1991, S. 149).

Diese „Zahlen“ sind auch deshalb keine Repräsentationen, weil sie das Verhalten des Systems nur online, d.h. wenn es kontinuierliche Verbindung mit der Umwelt hat, nicht aber offline, d.h. wenn diese Verbindung unterbrochen ist, bestimmen können. Sobald die Verbindung zur Umwelt abbricht (z.B. durch einen Ausfall eines Sensors oder Leitungsbruch), bricht auch das gesamte System zusammen. Verfügte es über Repräsentationen, wäre dieses Problem zu umgehen.

Hierzu lassen sich drei Punkte anmerken. Erstens könnte es sein, dass aufgrund einer komplexeren Architektur der Mobots Repräsentationen möglich sind. Es könnten Ziele oder Motivationen, d.h. Dinge die diese Mobots machen oder erreichen sollen (z.B. Coladosen einsammeln), durch eine feste Verdrahtung von Bauteilen vorgegeben werden. Ferner könnte dadurch auch ein kleiner Teil der Welt fest repräsentiert sein (z.B. die Begrenzungen eines Fussballfeldes in einem Robocup-Turnier).³⁶ Zweitens könnten auch Mobots als dynamische Systeme aufgefasst werden. Aber aufgrund der festen Struktur der Subsumptionsarchitektur, d.h. aufgrund einer nicht-dynamischen Systemstruktur, kann sich innerhalb dieses Systems keine „freie“ Dynamik entwickeln. Die Plastizität, d.h. die Strukturveränderlichkeit, des Systems ist nicht gegeben. Innerhalb eines solchen Systems können sich daher auch keine „willkürlichen“, d.h. völlig aus den inneren und äusseren Einflüssen geformten, Systemzustände ergeben. Zwar kann jedes System als ein dynamisches System betrachtet werden. Damit aber dessen Eigenschaften ganz zur Geltung kommen können, muss ein System über eine plastische Systemstruktur verfügen. Je grösser die Plastizität eines Systems, desto dynamischer kann es sein (vgl. Abschnitt 3.2.2, S. 34 und Fussnote 44 auf S. 86). Drittens könnte man die Plastizität von Mobots erhöhen, indem man in sie neuronale Netze einbaut, also dynamische Systemarchitekturen zulässt. Damit würden ihnen auch möglicherweise höhere kognitive Fähigkeiten über implizite Repräsentationen gelingen.

³⁶ „Robocup“ ist die Fussballweltmeisterschaft für Roboter.

Mobots haben bisher aber erst eine kognitive Ebene erreicht, auf der es ihnen noch nicht einmal möglich ist, die XOR-Funktion zu repräsentieren. Höhere kognitive Fähigkeiten sind also für Mobots zumindest heute noch kein Thema.

Brooks „Subsumption Architecture“ ist nicht antirepräsentational. Diese Schlussfolgerung hängt davon ab, wie der Begriff „Repräsentation“ aufgefasst wird. Brooks behauptet, dass seine kognitiven Agenten, die aufgrund bestimmter Situationsbedingungen handeln, über keine klassischen Repräsentationen verfügen. Denn die Entitäten, die in Mobots vorzufinden sind („Zahlen“, d.h. rein quantitative Zustände), weichen nach Brooks von klassischen Repräsentationen zu sehr ab. Mobots verfügen z.B. auch über keine diskreten Variablen (vgl. Brooks 1997, S. 406). Daher ist es „Subsumption Architecture“ heute auch noch nicht möglich, eine Repräsentation der XOR-Funktion zu implementieren.³⁷

Brooks geht davon aus, dass Mobots nur in Interaktion mit der Umwelt mit dieser zurechtkommen, d.h. intelligentes Verhalten zeigen. Alles, was der Mobot dazu braucht, kommt aufgrund seiner Situiertheit in der Umwelt aus dieser. Durch die direkte körperliche physikalische Verbindung und Interaktion mit der Welt gelangt er zu Bedeutungen (s. dazu Kap. 3.1.2, S. 13f). Intentionalität als emergente Eigenschaft von Mobots wird diesen nur zugeschrieben. Sie selbst verfügen aber nicht über eine solche. Aufgrund emergenter Eigenschaften der Mobots schreibt man ihnen komplexere Strukturen zu als sie haben (s. dazu Abschnitt 4.3.3, S. 82f). Aus diesen Gründen braucht ein Mobot keine internen Repräsentationen.

Obwohl ich alle diese Gründe für ziemlich plausibel halte, interne Repräsentationen reduzieren zu können, bin ich nicht Brooks Ansicht, dass es keine solchen gibt oder solche für Kognition nicht gebraucht werden. Vielmehr sind auch in Mobots Repräsentationen zu finden. Roboter, d.h. festverdrahtete Maschinen, sind – ähnlich wie Prozessoren – mit überaus vielen Zielen und Repräsentationen ausgestattet. Denn der Designer gibt ihnen zumindest teilweise vor, wie sie sich die Welt, in der sie sich bewegen sollen, vorstellen sollen und was sie wie wann machen sollen. Sicherlich ist das Ergebnis der Handlungen von Robotern auch unmittelbar von der Situation abhängig, in der sie sich gerade befinden. Aber grundlegende Motivationen wie Gehbewegungen, oder das Erkennen bestimmter Gegenstände wie Coladosen sind vorprogrammiert. Der Roboter muss also z.B. eine Repräsentation einer Coladose (oder

³⁷ In einem E-Mail teilte mir Prof. Brooks – auf meine Anfrage, wie denn die XOR-Funktion in der Subsumption-Architektur seiner Mobots repräsentiert werden könne – mit, dass ich ihn genauso gut hätte fragen können, wie seine Kreaturen Schach spielen könnten. Ich interpretiere dies als eine klare Absage hinsichtlich der Fähigkeit von Mobots, High-Level-Cognition leisten zu können.

etwas Ähnlichem) in sich haben, da er diese sonst nicht erkennen und einsammeln kann.

Da Brooks Mobots nur über Low-Level-Cognition verfügen, sind sie sicher nicht repräsentativ genug, um daraus antirepräsentationalistische Argumente bezüglich menschlicher Kognition (= High-Level-Cognition) ableiten zu können.

Ein weiteres Problem mit Mobots könnte nach Eliasmith (1996) sein, dass die Probleme aus unteren Ebenen auf höhere Ebenen mitgeschleppt werden und dass, obwohl die Mobots über keine semantisch bewertbaren Tokens verfügen, sie diese Art der Tokens unterstützen können müssen, um interessantes Verhalten an den Tag legen zu können. Vermutlich müssen viele stabile Zustände in jedem kognitiven System vorhanden sein, auch wenn diese Zustände nicht unbedingt explizite, aber doch implizite Repräsentationen B sind. Hierzu gibt es bereits interessante Alternativen aus dem Bereich dynamischer Systeme, von denen manche auf solche Tokens verzichten (Beer & Chiel 1997), andere an solchen festhalten (Jäger 1996, vgl. Kap. 3.2.3, S. 50ff).

Beer & Gallagher (1992) zeigten, dass es innerhalb dynamischer Systeme auch ohne Symbole und ohne Verbindung zur Aussenwelt möglich ist, dass bestimmte Zustände aufrechterhalten bleiben. Ein offenes, d.h. mit der Umwelt verbundenes, System interagiert mit der Umwelt. Wird dieses System dann von der Umwelt abgetrennt (d.h. es wird zu einem (für sich ab-) geschlossenen System), hat es dennoch Einflüsse der früheren Umwelt aufgrund der Systemdynamik mittels stabiler Zustände in sich gespeichert. Beer unterscheidet zwischen intrinsischen computationalen Systemen, welche über interne Repräsentationen verfügen und unechten computationalen Systemen, welche über keine interne Repräsentationen verfügen. Beer verwendet anstatt symbolischer Repräsentation und Semantik die dynamische Systemtheorie (DST), welche autonome Agenten wesentlich besser beschreiben kann. Innerhalb der DST geht man von einem Zustandsraum und einer oder mehrerer Trajektorien durch diesen Raum aus und wendet für deren Berechnung diskrete und kontinuierliche Mathematik an.³⁸ DST analysiert die Evolution der Systemzustände über die Zeit hinweg. Obwohl solche Systeme reichhaltig über „interne Zustände“ verfügen, sind diese nach Beer & Gallagher (1992) keine Repräsentationen. Denn dynamische Systeme verfügen auch über die Fähigkeit, ein Verhalten zu zeigen, das sowohl mit als auch ohne Feedback von der Umgebung gleich bleibt. Dies zeigt, nach Meinung Beer & Gallaghers, dass auch nicht-repräsentationale innere Zustände das Verhalten des Systems unabhängig von der Umwelt (d.h. offline) bestimmen können. Anstatt interner Repräsentation von Zuständen der Umgebung des Agenten

³⁸ Vgl. Kap 3.2.3, S. 43ff.

erfolgt eine starke Kopplung zwischen den Umweltzuständen und den Antworten des Systems. Wenn diese Behauptung allgemeine Gültigkeit erlangen könnte, dann wären die Probleme der Memorierung und Planung, d.h. das Problem des „offline-reasoning“, beseitigt und man könnte (zumindest in diesem Bereich) ganz auf interne Repräsentationen verzichten.

Ich denke jedoch, dass sich diese Ablehnung nur auf explizite Repräsentationen bezieht, nicht aber auf implizite Repräsentationen. In diesem Modell von Beer & Gallagher sind „interne Zustände“ vorhanden, die man aber durchaus als implizite Repräsentationen, wie in Kapitel 3.1.3 (S. 21f) definiert, bezeichnen kann. Denn ist es möglich, dass sich unter den vielen „internen Zuständen“ innerhalb eines dynamischen Systems welche finden lassen (z.B. Attraktoren), denen man eine bestimmte Bedeutung aus der Umwelt beimessen kann. Damit könnte man sogar von expliziter Repräsentation sprechen, weil man ihnen ihre Bedeutung zuordnen kann. Allerdings müssten solche Zustände erst einmal erkannt werden. Denn es wird aufgrund der Kontinuität dynamischer Systeme bzw. wegen der permanenten Veränderungen des Systems schwierig, solche stabilen Zustände auszumachen (zu messen oder zu berechnen). Es passieren innerhalb dynamischer Systeme sehr viele und kontinuierlich Ereignisse, woraus sich viele stabile Zustände ergeben, welche vermutlich nie alle explizit gemacht werden können – aber dennoch einige davon. Und die, die davon noch nicht ausgemacht werden konnten, kann man Repräsentation B nennen, also als implizite Repräsentationen verstehen.

In diesem Sinne kann also innerhalb von dynamischen Systemen von impliziten und von expliziten Repräsentationen gesprochen werden. Dynamische Systeme sind damit keine antirepräsentationalistischen Systeme, sondern repräsentationalistische Systeme.

Das gleiche Argument kann gegen Keijzer (1998) vorgebracht werden, der auch Repräsentation zumindest im klassischen Sinne ablehnt. Er geht von einem erweiterten interaktionistischen Arbeitsrahmen aus, mit dem antizipatorisches Verhalten erklärt werden kann, ohne dabei repräsentationale Zielzustände vorzugeben bzw. interne Repräsentation zu verwenden. Ein durch interne Zielzustände geführtes Verhalten wird „antizipatorisch“ genannt. Dieses Verhalten ist nicht nur durch die unmittelbare Umgebung des Agenten bestimmt, sondern auch durch Ziele, die unter Umständen erst später erreicht werden. Repräsentationalisten nehmen an, dass dieses Verhalten nur über eine interne Repräsentation dieser Ziele (z.B. mittels einer Wissensbasis) realisierbar ist. Keijzer hingegen sieht diese Ziele über interne Kontrollparameter innerhalb

dynamischer Systeme realisiert, welche sich aufgrund von Selbstorganisationsprozessen des Systems und der Interaktion mit der Umwelt bilden. Allerdings können diese internen Kontrollparameter aufgrund der Dynamik des Systems nicht lokal geortet werden und somit sind sie nicht als Repräsentationen anzusehen:

„The concept of „internal control parameters“ is definitely not the same as that of „representation“.” (Keijzer 1998, S. 297)

Zwar können diese nicht geortet und damit nicht erkannt werden. Aber weil Parameter nur Sinn machen, wenn sie im Vergleich zur Dynamik des Systems relativ stabil bleiben, kann man m. E. auch hier von impliziter Repräsentation im Sinne unseres Referenzbegriffes von Repräsentation B sprechen.

Die Implementierung einfacher kognitiver Fähigkeiten wie Mustererkennung und Musterklassifizierung oder boolesche Operationen (wie NOT, OR, XOR) sind für neurodynamische Ansätze in der Kognitionswissenschaft kein Problem (vgl. Rumelhart 1986).³⁹ Dies zeigt sich in vereinfachter Form auch daran, dass die XOR-Funktion innerhalb dynamischer Systeme problemlos implementiert werden kann (vgl. Kap. 3.2.3, S. 48ff). Aber auch höhere kognitive Fähigkeiten wie Memorierung, Planung oder Vorhersage scheinen für Ansätze dynamischer Systeme (zumindest innerhalb eingeschränkter Bereiche) machbar zu sein: Planung kann durch die „Decision Field Theory“ und die „Motivational oscillator Theory“ von Busemayer & Townsend (1993) erklärt werden. Antizipatorisches Verhalten wird durch die „Behavioral Systems Theory“ von Keijzer (1998) erklärt. Die Fähigkeit dynamischer Systeme, kurzfristige Vorhersagen zu treffen (wie der Abschätzung des Abstandes zu einem Gegenstand, damit der Agent diesen treffsicher greifen kann), haben Rooij, Bongers und Haselager gezeigt (vgl. Rooij & Bongers & Haselager 2000). Ein Speichereffekt, der für Memorierung und die Emulation⁴⁰ genutzt werden kann, ergibt sich aus dem dynamischen Systemmodell von Beer & Gallagher (1992).

³⁹ Die grundlegende kognitive Fähigkeit des Menschen ist es, Objekte zu erkennen bzw. zu unterscheiden, also Mustererkennung und Negation. Explizit wird gesagt: „Dies ist eine Katze.“ Dies erfordert vorher Mustererkennung. Implizit wird auch gedacht bzw. „schwingt“ mit: „Dies ist eine Katze (aber alles andere nicht).“ Dies erfordert eine Anwendung der Negation.

⁴⁰ Mit Emulation ist hier die kognitive Vorab-Simulation noch nicht ausgeführter sensomotorischer Handlungen gemeint (vgl. Grush 1997). Interessant wäre auch, genauer zu untersuchen, ob nicht eine fruchtbare Verbindung zwischen Grushs Emulationstheorie und dynamischen Systemen zustande kommen könnte (was allerdings im Rahmen dieser Arbeit nicht geleistet werden kann). Dies würde mit den Vorteilen dynamischer Systeme hinsichtlich kontinuierlicher Echtzeitkognition mit den Vorteilen der Emulation dazu führen, dass beide Hauptprobleme hoher Kognition eliminiert werden könnten: Emulierende dynamische kognitive Agenten kämen demnach sowohl mit Abstraktem als auch mit Abwesendem zurecht.

Es stellt sich nun die Frage, ob und wie solche Theorien miteinander kombiniert werden können, um nicht nur einzelne kognitive Eigenschaften erklären zu können, sondern eine „Gesamtkognition“. Eine Möglichkeit könnte sein, abstrakte Entitäten als emergente Eigenschaften eines dynamischen Systems aufzufassen, falls dieses bereits über die Fähigkeit der Memorierung und Emulation der Umwelt verfügt. Abstrakte Entitäten könnten als stabile Zustände (Attraktoren) oder als plötzlich auftretende qualitative Veränderungen (Bifurkationen) innerhalb eines dynamischen Systems betrachtet werden. Kontrafaktisches Schliessen (also Schliessen mit Sachverhalten, dessen Zutreffen in der Zukunft für möglich (Antizipation) oder unmöglich (Fiktion) gehalten wird) könnte dann möglich sein, wenn die Fähigkeit zur Vorhersage und abstrakte Entitäten verfügbar sind. Aber dies sind Bereiche, in denen noch erhebliche Forschungsarbeit geleistet werden muss, um solche Postulate überprüfen zu können.

High-Level-Kognition kann also über die Theorie dynamischer Systeme zumindest angegangen werden, wenn auch damit „echte“ menschliche Kognition aufgrund deren Komplexität vermutlich nie erreicht werden kann (vgl. Kap. 4.1.3, S. 66ff). Aus dem bisher Gesagten lässt sich hinsichtlich dynamischer Systeme ableiten, dass sie zumindest Teile höherer Kognition erklären können. Es ist damit festzuhalten, dass bottom-up-Ansätze (zu denen dynamische Systeme gerechnet werden) nicht unbedingt nur über Low-Level-Cognition verfügen müssen.

4.2.2 Repräsentation und Sprache

Die repräsentationalistische Redeweise (auf höherer (z.B. konzeptueller) Systemebene) ist sicher sinnvoll. Denn auf der obersten, psychologischen Erklärungsebene fällt es uns am leichtesten, Erklärungen, Beobachtungen und Begründungen informell zu beschreiben. Allerdings sind die inhaltlichen Abstände der einzelnen Beschreibungsschritte so gross, dass im theoretischen Gesamtgebäude zu stark abstrahiert und vereinfacht wird. Sicherlich lassen sich dadurch schnell und einfach kohärente (und vor allem auf den ersten Blick intuitiv plausible) Modelle der Kognition entwickeln. Dies scheint aber auf jeder anderen Beschreibungsebene auch möglich zu sein, wenn auch mit grösserem Aufwand. Je mehr die Beschreibungsebene sich der physikalischen oder formalen nähert, desto schwerer und unwahrscheinlicher wird es, vollständige Beschreibungen zu finden, die das Bild, das sich uns zeigt, ganz abdecken. Dennoch ergeben sich aus detaillierteren Beschreibungsschritten detailliertere Informationen hinsichtlich des Gesamtsystems. Werden genug solcher neuer

Informationen gefunden, kann sich die Gesamtfunktion des Systems auf ganz andere Weise erklären lassen. Es ist auf diese Weise sogar möglich, dass komplexe funktionale Einteilungen aufgrund anderer Erklärungen anders vorgenommen werden können und dadurch die Arten der Beziehung zwischen den groben Beschreibungsschritten auf oberster Ebene anders ausfallen könnten. Dies wiederum könnte dazu führen, dass die von uns zuvor auf den ersten Blick für einfach und plausibel gehaltene Theorie ins Wanken gerät oder sogar verworfen werden muss.⁴¹

Wir können nicht auf interne Repräsentationen verzichten, wenn es darum geht, einem anderen Menschen das Verhalten eines kognitiven Systems zu erklären, weil die notwendige Eigenschaft der Effizienz der Kommunikation uns dazu zwingt. Aus konzeptueller Sicht ist auf unterster (physikalischer) Ebene eines kognitiven Systems eine solche Verwendung interner Repräsentationen hingegen nicht notwendig und auch gar nicht mehr möglich, weil uns das Verständnis aufgrund unserer hohen Kommunikationsebene (Sprache) dafür fehlt: Wir brauchen gar nicht auf die physikalische Ebene des Gehirns zu schauen oder zu wissen, was da vor sich geht, um mit jemanden kommunizieren zu können. Und wir können uns während des Kommunizierens nicht um die internen stabilen Zustände (oder Repräsentationen B) innerhalb des dynamischen Systems „Gehirn“ kümmern, sonst kämen wir nicht zum Sprechen. Auch die Systeme selbst kommunizieren auf unterster (physikalischer) Ebene mittels interner Repräsentationen B, also mittels *für sich* genommen bedeutungsloser Interaktionen, da nicht im Sinne der Bedeutung ($M \rightarrow B$) interpretiert (vgl. dazu Abschnitt 3.1.2, S. 14). Erst wenn die Bedeutung B (potentielle Bedeutung) interpretiert wird, wird sie zu einer aktuellen Bedeutung. Und Handlungsbedeutung ergibt sich erst auf höchster kognitiver (konzeptueller) Ebene (der Sprache ?) durch das Zuschreiben interner Repräsentation durch Interpretation (vgl. Davidson 1984 und Lewis 1983).

Wir verwenden also ständig Repräsentationen. Dies ist auf sprachlicher Ebene unvermeidlich, um effektiv kommunizieren zu können. Wenn wir darüber *sprechen*, was wir z.B. planen werden, dann sprechen wir über Repräsentationen. Aber auch auf nicht-sprachlicher Ebene verwenden wir Repräsentationen B, um planen zu können. Denn es gibt keine direkte Verbindung (Relation, d.h. Referenz von Begriff auf Gegenstand) zwischen Sprache und Welt (Wirklichkeit), sondern nur eine indirekte. Damit ist nicht nur gemeint, dass die Sprache die Welt nicht abbildet, sondern dass

⁴¹ Ich liege da auf einer Linie mit Daniel Dennett, der vertritt, dass jedes Verhalten auf verschiedene Weise intentional erklärt werden kann und es gibt auch keine Möglichkeit festzustellen, welche die richtige ist (vgl. Dennet 1987).

vielmehr die Wirklichkeit grundsätzlich nicht adäquat beschreibbar ist. Die Welt kann nicht direkt sprachlich (von einem Dritten) im Computer vollständig repräsentiert werden. Vielmehr müssen Computer und Mensch ein relevantes Modell der Welt entwerfen bzw. nur die Information verfügbar halten, die sie brauchen (vgl. „Pengi“ in Abschnitt 2.3.2, S. 8 und Agre & Chapman 1987). Ausserdem müssen sie sich dieses Wissen *selbst* aneignen – es kann ihnen nicht einprogrammiert werden.

Ich denke auch, dass der Mensch (das menschliche Gehirn) viel zu komplex ist, um dessen Entwicklung sowie dessen Funktion und damit die Verbindung bzw. Beziehung zwischen Gedanken und Gehirn auf hoher kognitiver Ebene (d.h. mittels psychologischem Vokabular) erklären zu können (da dieses Vokabular nicht alle Erklärungsebenen dazwischen einfangen bzw. abdecken kann, sondern eben nur die oberste (ungenaueste bzw. oberflächlichste)). Insbesondere muss die initiale Interaktion mit der Welt mit einbezogen werden. Aus diesen Gründen ist es nicht nur sinnvoll, sondern auch notwendig, zuerst einfache(re) neuronale Organismen und deren Verhalten und Interaktion mit der Umwelt zu untersuchen (vgl. Clark 2001). Der synergetische Effekt der Verbindung dieser Ergebnisse könnte dann hinreichen, schliesslich die menschliche Kognition auf neurobiologischem Substrat erklären zu können (vgl. Haken 1982).

Sprache und Welt sind nicht direkt verbunden. Vielmehr ist „Sprache im Kopf“ indirekt mit der „Sprache in der Welt“ verbunden über ein „Erden“ (Grounding), d.h. über eine Interaktionsbeziehung auf materieller Ebene an der molekularen (materiellen, physikalischen) Schnittstelle (Systemgrenze) zwischen Mensch und Welt (vgl. die Abschnitte 3.1.3, S. 20 und 3.3.4, S. 55ff). Der Austausch von „Sprachtokens“ (Ausdrücken) erfolgt somit nicht direkt mental, sondern indirekt und rein physikalisch. Aussagen (Tokens) der Sprache könnten daher möglicherweise als „Auslagerungen“ oder „physikalische Realisierungen“ von Attraktoren bzw. höheren stabilen Systemzuständen (in der Welt, d.h. ausserhalb des kognitiven Systems) verstanden werden.

4.3 Ontologische Probleme

4.3.1 Repräsentation und Bedeutung

Nach Eliasmith (1996) repräsentieren einzelne Neuronen nichts. Menschen aber schon. Irgendwo dazwischen werden also Repräsentationen erzeugt. Daher sollte man nach Eliasmith diese nicht einfach ignorieren oder abschaffen, nur weil sie auf neuronaler Ebene nicht sichtbar sind. Dabei darf man aber ontologische und methodologische Ebene nicht verwechseln. Es ist durchaus denkbar, dass wir über Repräsentationen sprechen, weil es pragmatisch sinnvoll ist, diese Entitäten anzunehmen. Auf explanatorischer Ebene ist für die Konstitution von Repräsentation eine Zuschreibung von Bedeutung $M \rightarrow B$ nötig, wodurch explizite Repräsentation entsteht. Implizite Repräsentation kann gar nicht als Repräsentation erkannt werden, ohne ihr Bedeutung $M \rightarrow B$ von aussen zuzuschreiben. Damit wird aus dieser impliziten Repräsentation eine explizite. Je weiter wir uns von der intentionalen (konzeptuellen) Ebene auf die physikalische Ebene des Systems begeben, desto weniger können wir von expliziten Repräsentationen sprechen, weil wir diese immer schlechter erkennen, da sie immer weniger symbolisch sind. Symbole erkennen wir nur auf konzeptueller Ebene, weil sie uns dort aufgrund ständigen Gebrauchs vertraut sind. Es sind eher implizite Repräsentationen im Sinne der nicht-intentionalen Bedeutungen B anzutreffen, die aber erst unserer Interpretation bedürfen, um zu expliziten Repräsentation zu werden: Wenn ein Neurowissenschaftler aufgrund von Beobachtungen innerhalb des Gehirns solche stabile Zustände findet (implizite Repräsentation B), dann muss er feststellen, in welchen Situationen diese immer auftreten. Aufgrund solcher Beobachtungen kann er schliesslich der Repräsentation B eine Bedeutung beimessen, wodurch diese zu einer klassischen expliziten Repräsentation mit einer ständigen Bedeutung wird.

Auf sprachlicher (konzeptueller) Systemebene sind Repräsentationen meist explizit. Implizite Repräsentationen liegen auf dieser Ebene z.B. vor, wenn man die Bedeutung „zwischen den Zeilen“ erkennen muss.

Gehirne funktionieren aber auch ohne dass ihnen irgendwelche expliziten Repräsentationen zugeschrieben werden, also auch ohne Interpretation von aussen. Es ist daher nicht notwendig, dass das Gehirn über explizite Repräsentationen verfügt.

Allerdings verfügt es als dynamisches System notwendigerweise über implizite Repräsentationen, wie wir mehrfach festgestellt haben (s. die Abschnitte 3.1.3 (S. 20), 3.3.4 (S. 55ff) und 4.2.1 (S. 69ff)).

Fazit: Damit etwas explizit repräsentiert, muss es über einen „geformten“ Träger (physikalische Struktur) und über Inhalt (Bedeutung $M \rightarrow B$) verfügen. Es muss also etwas geben, das diese Repräsentation als solche erkennt (intentionales Wesen). Es gibt also keine explizite Repräsentation ohne Rezipienten oder zuschreibendem Wesen. Dort, wo es kein intentionales Wesen gibt, gibt es implizite Repräsentation B aufgrund eines „geformten“ Trägers (physikalische Struktur) und Inhalt (Bedeutung B). Aber beides, physikalische Form und semantischer Inhalt fallen zusammen bzw. sind physikalisch dasselbe (vgl. Abschnitte 3.1.3 (S. 20), 3.3.4 (S. 55ff) sowie 4.3.3 (S. 83f)). Diese Implikation nehmen Antirepräsentationalisten der Kognitionswissenschaft (insbesondere Dynamizisten) zum Anlass, nur noch Form (Strukturen, und damit Bedeutung B) zu untersuchen und Inhalt (Bedeutung $M \rightarrow B$) zu ignorieren. Und sie merken gar nicht, dass sie dabei ständig implizite Repräsentationen B untersuchen.

4.3.2 Repräsentation und Intentionalität

Markman & Dietrich (2000) geben für perzeptuelle und motorische Prozesse zu, dass eine verkoppelte Repräsentation mit sensomotorischen Systemen die Semantik der Repräsentationen begründen könnte (vgl. dazu Abschnitt 3.2.1, S. 33). Allerdings gilt dies nach Markman & Dietrich nicht für High-Level-Kognition. Denn die Fähigkeit, abstrakte Konzepte zu repräsentieren, erfordert (neben dem Wissen um deren Bedeutung) einen zweiten Aspekt: ihre funktionale Rolle, d.h. ihre Beziehung zu anderen repräsentationalen Elementen. Innerhalb dynamischer Systeme ist es schwierig, die gemeinsamen Eigenschaften kognitiver Agenten zu erklären:

"[...] it seems clear that, on plausible assumptions about how to individuate psychological capacities, internally equivalent systems must have the same psychological capacities. Hence, it would appear that wide content cannot be relevant to characterizing psychological equivalence." (Dancy & Sosa 1992, S. 443).

Nur identische dynamische Systeme können das gleiche Verhalten aufweisen.⁴² Aber aufgrund der internen Dynamik dynamischer Systeme ist es sehr unwahrscheinlich, dass zwei dynamische Systeme äquivalentes oder gar identisches Verhalten aufweisen. Daher ist es auch unwahrscheinlich, dass zwei kognitive dynamische Systeme die gleichen psychologischen Fähigkeiten haben. Ausserdem ermöglicht (bzw. bestimmt daher) gerade „wide content“ gleiche psychologische Fähigkeiten. Denn wir befinden

⁴² Man könnte möglicherweise sogar sagen: Es gibt keine ontologische Identität (z.B. aufgrund der Erkenntnisse aus der Quantentheorie bzw. unendlich kleiner Teilchen und der daraus resultierenden kombinatorischen Vielfalt).

uns alle in derselben Welt bzw. unterliegen denselben naturgesetzlichen Prozessen, denen wir uns anpassen müssen. Gerade über die Interaktion von Agenten in einem sinnvollen Zusammenspiel in der Gemeinschaft ergeben sich durch implizite Konvention wichtige Aspekte, die diese Gemeinschaft erst aufrechterhalten und welche somit auch die gemeinsamen Eigenschaften der Individuen konstituieren.

Aber auch „narrow content“ (d.h. intentionale Kausalität bzw. mentale Verursachung) kann relevant sein. Intention ist immer mehr als Repräsentation, denn Intention ist immer aktiv. Repräsentation hingegen kann auch passiv sein (implizite Repräsentation). Vielleicht könnte man unter einer Intention auch eine aktive (bewusste), d.h. explizite Repräsentation verstehen. Und da innerhalb dynamischer Systeme ontologisch sowohl implizite als auch explizite Repräsentationen vorhanden sind, gibt es in dynamischen Systemen auch Intentionen. Intentionale Zustände können aber nach Fodor nur real sein, wenn jedem intentionalen Zustand eindeutig ein neuronaler (physikalischer) Zustand entspricht, durch den er realisiert ist. Und wenn ein intentionaler Zustand Z1 einen anderen intentionalen Zustand Z2 verursacht, dann folgt aus den Gesetzen der Neurobiologie, dass die Realisierung von Z1 die Realisierung von Z2 verursacht (vgl. Fodor 1987).

Ich denke, dass dynamische Systeme Fodors Bedingungen erfüllen. Die Behauptung, dass die Beziehung zwischen mentalen Repräsentationen (Intentionen) kausal sei, könnte durch Ansätze dynamischer Systeme bestätigt werden. Denn wenn man für die Kognition ein dynamisches System annimmt, dann ist es durchaus denkbar, dass eine Repräsentation B_{t1} eine andere Repräsentation B_{t2} kausal (innerhalb des dynamischen Systems, d.h. auf physikalischem Wege – wie in den Abschnitten 3.1.3, (S. 20), 3.3.4 (S. 55ff) und 4.3.3 (S. 83f) geschildert) verursacht, womit die kausale Geist-Materie-Richtung erklärt werden könnte. Dynamische Systeme realisieren intentionale Zustände auf physikalischer Ebene. Attraktoren sind stabile physikalische Systemzustände, die sich durch die Dynamik innerhalb des Systems ergeben. Sie entsprechen intentionalen Zuständen. Die Gesetze der dynamischen Systemtheorie erklären, wie ein stabiler Systemzustand Z1 einen anderen stabilen Systemzustand Z2 verursacht. Auf diese Weise leuchtet auch ein, warum nicht jedem intentionalen Zustand ein neuronaler Zustand entspricht. Ein Attraktor überdauert in seiner Lebenszeit viele verschiedene neuronale Zustände. Attraktoren können über sehr lange Zeit aufrechterhalten bleiben und somit als Speicher fungieren. Sie existieren auf einer höheren Systemebene als auf neuronaler Ebene.

Dynamische Systeme können je nach Komplexität und Art ihrer Systemstruktur intentionale Systeme sein. Wann ein System zu einem intentionalen wird, hängt vor allem auch von seinem Eingebundensein in eine intentionale Umgebung ab. Alle intentionalen Systeme sind aber dynamische Systeme.

Es gibt also eine Alternative zu Fodors repräsentationaler Theorie des Geistes (RTG): die dynamische Systemtheorie. Intentionale Zustände sind real im gleichen Sinne wie Attraktoren innerhalb eines Systems real sind. Real können intentionale Zustände nach Fodor nur sein, wenn sie physikalisch realisiert sind. Die RTG ist nicht mehr die einzige Theorie, die uns sagt, wie sie realisiert sein können. Auch die Dynamische Systemtheorie (DST) kann uns dies sagen: intentionale Zustände sind physikalisch als Attraktoren realisiert. Dynamische Systeme können also intentionale Systeme sein.

Fazit: Gerade weil sich Gehirne in einer "embodied embedded situation" befinden, kann ein Gehirn zwar über dieselben psychologischen Fähigkeiten verfügen wie ein anderes Gehirn, muss es aber nicht. Ausserdem können mentale Inhalte als qualitative Eigenschaften physikalischer Strukturen aufgefasst werden, die sich aufgrund stabilen Systemverhaltens (Attraktoren) und aufgrund plötzlicher qualitativer Veränderungen (Bifurkationen), d.h. als Synergieeffekte (deterministischem Chaos mit anschließender Selbstorganisation) innerhalb eines dynamischen Systems ergeben (vgl. Kapitel 3.2.3, S. 43ff und 4.3.3, S. 82ff). Auf diese Weise sind mentale Repräsentationen B bzw. Intentionen kausal auf andere mentale Repräsentationen B bzw. Intentionen wirksam. Und da Repräsentationen B bzw. Intentionen nichts anderes als physikalische Zustände sind, wirken Intentionen auch auf die physikalische Ebene. Repräsentationale Systeme können intentionale Systeme sein. Jedes intentionale System ist ein repräsentationales System.

4.3.3 Emergenz oder nicht-reduktiver Physikalismus

Es gibt erstaunliche Prinzipien in der Natur: Emergenz und Selbstorganisation. Ergeben sich Systemeigenschaften aus den physischen Teilen des Systems (und den Prozessen im System), die nicht mechanisch erklärbar sind, dann sind dies emergente Eigenschaften (vgl. Beckermann 1999, S. 217). Keijzer (1998) gibt ein Beispiel dafür, wie qualitativ neue Eigenschaften aus einem quantitativen System entstehen:

„To understand how a car works, you will have to figure out how a collection of car parts is assembled so that together they come to exhibit a totally new property: automotive power. You cannot deal with this problem by assuming that this “automotive power” somehow has to reside within the parts.” (Keijzer 1998, S. 275)

Aufgrund des Emergenzprinzips kann die dynamische Systemtheorie eine Brückenfunktion zwischen physikalischen und mentalen Erklärungen innehaben (vgl. Jäger 1996, S. 168). Höhere Dimensionalität des Systems wird durch eher emergent-qualitative Ausdrücke wie „Attraktor“ beschrieben (vgl. auch Kelso 1995), während auf physikalischer Ebene dasselbe Phänomen in Differentialgleichungen quantitativ berechnet wird. Emergente Eigenschaften sind mehr als die (quantitativen) Teile des Systems: sie sind qualitative Eigenschaften. Nach Jäger (1996) ergeben sich qualitative Eigenschaften aus einer explanatorischen Reduktion. Zunächst werden physikalische Phänomene beobachtet und auf systemtheoretischer Ebene quantitativ erfasst. Anschliessend werden die Gleichungen des Systems analysiert und qualitative Aussagen darüber getroffen und schliesslich als emergente Phänomene klassifiziert (vgl. Jäger 1996, S. 168). Möglicherweise könnte durch eine solche Theorienreduktion auf dynamische Systeme die Beziehung zwischen Gedanken (intentionalen Inhalten) und Gehirnmasse (neurodynamisches System) erklärt werden (vgl. Clark 2001, S 8: Lösung durch DS, vgl. Jäger 1996). Es findet dabei aber nur eine methodische (konzeptuelle) Reduktion statt, keine ontologische (d.h. es wird nicht Mentales auf Physisches reduziert).

Aber wie es scheint, haben wir bereits heute auch eine mechanistische Erklärung für diese emergenten, qualitativen Eigenschaften eines physikalischen, quantitativen Systems: Dynamische Systeme sind quantitativ-physikalische Systeme. In ihnen treten aber Veränderungen auf (Bifurkationen, vgl. Kapitel 3.2.3, S. 45), die wir als qualitativ einstufen. Ich denke, in Anlehnung an die Erläuterungen zum Referenzbegriff von Repräsentation in Kapitel 3.1.3, S. 20ff, dass qualitative Merkmale mit quantitativen auf der physischen Ebene zusammenfallen bzw. qualitative Merkmale nur durch Zuordnung entstehen, wobei die Zuordnung selbst bereits wieder auf physikalischer Ebene erfolgt. Qualitative Eigenschaften sind m. E. nichts anderes als Verknüpfungen zwischen quantitativen Eigenschaften. Anders gesagt: mentale Inhalte sind physikalische Beziehungen zwischen physikalischen Zuständen.

Man könnte also sagen, dass gar keine Reduktion stattfindet, da die qualitativen stabilen Zustände nur klassifiziert werden, d.h. das Wechseln physikalischer stabiler Zustände wird nur konzeptuell gekennzeichnet. Die auffälligste – vielleicht sogar die

einzig – Eigenschaft, über die sowohl mentale Inhalte als auch physikalische Systeme verfügen, ist deren Struktur. Daher liegt es nahe, innerhalb von Systemen diese beiden Dinge über ihre Struktur zusammenzubringen. Das Ergebnis ist, dass die konzeptuelle Ebene (qualitativ-diskret) mit der physikalischen Ebene (quantitativ-kontinuierlich) innerhalb eines dynamischen Systems mit genügend grosser Systemplastizität aufgrund ihrer Systemdynamik zusammenfällt (vgl. Abschnitte 3.2.2, S. 35ff und 4.3.2, S. 80ff).

Mentales ergibt sich kausal aus Physischem als physikalische Bezeichnung und der physikalischen Bezugnahme auf dieses physikalisch Bezeichnete: Intentionalität ist die Fähigkeit, solche Bezeichnungen durchzuführen und wieder zu erkennen. Konzepte oder Begriffe zu erwerben oder zu erkennen ist nichts anderes als rein physikalische Mustererkennung bzw. Musterklassifikation. Ein kleiner Eindruck davon, wie dies vor sich gehen könnte, wird in Abschnitt 3.1.3 auf S. 19f vermittelt. Mentales und Physikalisches sind dasselbe, nur jeweils von unterschiedlichen Perspektiven gesehen (aus konzeptueller oder aus physikalischer Sicht). Diese unterschiedliche Sichtweise entsteht auf physikalischer Ebene mit zunehmender Komplexität und Dynamik der physikalischen Strukturen innerhalb bestimmter Systeme (evtl. sogar plötzlich aufgrund von Bifurkationen, vgl. Abschnitt 3.2.3, S. 45).

Der so entstandene qualitative Unterschied zur physikalischen Sichtweise wird Intentionalität genannt. Dieser qualitative Unterschied, der nur aus der physikalischen Ebene hervorgeht entspricht dem Abstand zwischen Mentalem und Physischem und bewirkt die Möglichkeit intentionaler Zuschreibungen. Mentales geht kausal aus dem Physischen hervor. Mentales ist physikalisch erklärbar. Sowohl auf Mikro- als auch auf Makroebene können wir alles Physikalische (über die DST) erklären.

Dennoch kann es auf Mikroebene einen Dualismus geben, denn niemand weiss, worauf der Physikalismus seine Theorie letztlich begründen kann (z.B. weiss niemand, ob Materie eine duale (mental-materielle) oder eine monadische Substanz ist).

4.3.4 Zusammenführung antirepräsentationalistischer und klassischer Ansätze

Die Ansicht Markman & Dietrichs (2000), dass sich Ansätze der „Embodied Embedded Cognition“ hauptsächlich auf perzeptuelle und motorische Prozesse konzentrieren und High-Level-Kognition nicht erklären können, scheint also heute noch richtig zu sein, auch wenn Ansätze dynamischer Systeme gewisse Erfolge diesbezüglich zu verzeichnen haben. Zwar geben die Autoren zu, dass das Problem, wie Symbole nur mittels syntaktischer Manipulation Bedeutung erlangen können (Symbol Grounding

Problem), sich mit zunehmender Veränderung der Form und Implementierung von Repräsentationen, sogenannten „modest representations“, verflüchtigen könnte (vgl. Abschnitt 3.2.1, S. 33 und Harnad 1990 sowie Plunkett & Sinha 1992).⁴³ Sensomotorische Systeme könnten High-Level-Kognition aber deshalb nicht erklären, weil sie die funktionale Rolle abstrakter Konzepte, d.h. die Beziehung abstrakter Konzepte zu anderen repräsentationalen Elementen, nicht erkennen können. Dieser Aspekt wurde bereits in Abschnitt 4.3.2 (S. 80ff) näher untersucht.

Daher sollten klassische Ansätze symbolischer Repräsentation nicht abgeschafft, sondern erweitert werden, um die neuen Ansätze zu integrieren. Dadurch verändert sich die Sicht auf Repräsentation: nicht alle Repräsentationen sind permanent, symbolisch (diskret mit Referenz), amodal (d.h. resultieren nicht aus den Sinnesmodalitäten wie Sehen, Hören, Schmecken, Riechen und Tasten) oder von sensomotorischen Systemen des Agenten unabhängig. Da viele Aspekte der Welt stabil bleiben, kann die Anzahl interner Repräsentationen reduziert werden, indem externe Repräsentationen interne ersetzen. Dies ist aber nur dann möglich, wenn diese externen Repräsentationen welche sind, die nur dann gebraucht werden, wenn man an ihnen „vorbeikommt“ (z.B. Verkehrsschilder, Knoten im Taschentuch) (vgl. Abschnitt 2.3.2 über „Situating Action“, S. 7). Das Anbinden von Repräsentationen an spezielle Modalitäten kann als Grundlage für die grosse Flexibilität in kognitiven Prozessen und für abstrakte Konzepte dienen (perceptual symbol systems), weil modale Repräsentationen unmittelbar mit der Welt verbunden sind. Repräsentationen unterliegen kontinuierlichen Veränderungen entsprechend der Veränderungen der Umwelt, was die Wichtigkeit diskreter Symbole untergräbt und die Angemessenheit dynamischer Systeme hervorhebt.

Es scheint, dass CRUM und DST durchaus eine fruchtbare Verbindung miteinander eingehen könnten, um Kognition besser erklären zu können: Dynamisch-konnektionistische Netzwerke sind für Low-Level-Cognition auf physischer Ebene (z.B. Mustererkennung aus sensorischem Input oder Bewegung von Muskeln) aufgrund der Einbeziehung von Kontinuität, d.h. eines permanenten Zeitbezugs und der Interaktion mit der Umwelt besser geeignet als computational-repräsentationalistische Modelle. Für High-Level-Cognition auf konzeptueller Ebene (z.B. Erinnern, Planen, Vorstellen etc.) sind hingegen computational-repräsentationalistische Systeme adäquater. Daher sollten beide Ansätze vereint werden (vgl. Abschnitte 4.3.2 (S. 81f), 3.2.3 (S. 50ff) und Thagard 1999 sowie Jäger 1995).

⁴³ Der Begriff „Symbol Grounding Problem“ wurde 1980 von John Searle eingeführt. Vgl. dazu auch das Beispiel des Chinesischen Zimmers von John Searle (vgl. Searle 1992). Es wird diskutiert, ob Computer etwas verstehen oder nicht.

4.4 Zusammenfassung

Systematisch beschreibbare Strukturen und Prozesse sind Systeme. Kognition weist eine systematisch beschreibbare Struktur auf und beinhaltet systematisch beschreibbare Prozesse. Also ist Kognition ein System. Alle Systeme sind auch dynamische Systeme (weil sie wie alles in der Welt der Zeitlichkeit unterliegen). Kognition ist ein dynamisches System.⁴⁴ Dynamische Systeme sind repräsentational (was mit Hilfe des Referenzbegriffes aus Kapitel 3.1.3 (S. 16ff) im Laufe der Arbeit gezeigt wurde). Also sind alle kognitiven Systeme repräsentational. Oder anders gesagt: Es gibt keine antirepräsentationalistischen Systeme. Daher ist es sinnlos, Antirepräsentationalist zu sein.

Kognition ohne Repräsentation ist nicht möglich, weil sonst die Probleme „Abwesenheit“ und „Abstraktes“ nicht beseitigt werden können (s. bes. Abschnitt 4.2.1, S. 69ff). Dennoch kann Kognition mittels der dynamischen Systemtheorie erklärt werden. Kognition ist ein Phänomen, das sich mechanistisch aus entsprechend organisierten und komplexen dynamischen Systemen ergibt (s. Abschnitt 4.3.3, S. 82ff).

Die Fähigkeit, sich in der Welt zurechtzufinden, ohne die Welt dabei im Sinne klassischer Repräsentationen (d.h. Träger (Implementierung), Bedeutung (Inhalt) und direkte Beziehung (direkte Relation)) zu repräsentieren, wird durch dynamische Systeme möglich. Es gibt wohl (auf konzeptueller Ebene) die Unterscheidung zwischen Form und Inhalt, aber (auf physikalischer Ebene) keine direkte Relation zwischen Inhalt und Welt. Vielmehr stehen systeminterne Zustände auf der höchsten Systemebene mit Zuständen auf immer niedrigeren Systemebenen in kausaler Verbindung, so dass schliesslich auf niedrigster Systemebene die Schnittstelle zur Umwelt des Systems zu finden ist (dies dürfte die molekulare, d.h. die rein physikalisch beschreibbare Systemebene sein). *Hier* geschieht die Interaktion mit der Umwelt und nicht auf einer der höheren Ebenen. Es kann auf den ersten Blick also gar nichts innerhalb des Systems für etwas ausserhalb des Systems stehen, d.h. es scheint so, als ob es gar keine Repräsentation innerhalb des Systems geben könnte. Doch der Schein trügt.

Es gibt gerade in dynamischen Systemen (sich aus der Systemdynamik ergebende) stabile Zustände auf physikalischer Ebene, die wir auf konzeptueller Ebene

⁴⁴ Dynamische Systeme können unterschiedlich starke Systemdynamiken aufweisen. Dies liegt vorwiegend an ihrer Systemstruktur. Je stärker sie sich selbständig verändern kann (d.h. je grösser die Plastizität des Systems ist), desto stärker ist die Systemdynamik. Je grösser die Systemdynamik eines Systems ist, desto kognitiver kann es sein, d.h. desto besser kann es High-Level-Kognition durchführen. Kognitive Systeme weisen also eine entsprechende Plastizität ihrer Systemstruktur auf (vgl. dazu auch Abschnitt 3.2.2, S. 34ff und Haken 1982).

als Repräsentationen bezeichnen würden. Damit fallen Inhalt und Form bzw. konzeptuelle Ebene mit physikalischer Ebene auf dieser zusammen. Es gibt somit innerhalb von kognitiven Agenten keine klassische Repräsentation mit einer vertikalen Semantikrelation (d.h. einer direkten Relation zwischen mentalem Inhalt und Gegenstand in der Welt), sondern nur dynamisch erzeugte Stabilitäten, d.h. nur etwas Prozessuales, nichts Statisches! Selbst die neuronale Struktur im Gehirn ist nicht stabil, sondern plastisch, d.h. zumindest innerhalb bestimmter Rahmenbedingungen dynamisch: sie verändert sich ständig. Jedes neuronale Netz kann aufgrund genügender Plastizität ihrer Struktur und der daraus möglichen Systemdynamik als dynamisches System betrachtet werden. Damit kann auch das dynamizistische Vokabular auf Gehirne angewandt werden. Die Gehirne der Menschen haben sich im Laufe der Evolution zu optimalen plastischen dynamischen Systemen entwickelt, die die höchste Kognition hervorbringen, von der wir momentan empirisch wissen. Aufgrund ihrer hohen Plastizität können sie viele interne Repräsentationen haben und so menschliche Kognition verrichten. Watt-Regler-artigen dynamischen Systemen fehlt hingegen jegliche Plastizität. Dadurch können sie nicht über genügend interne Repräsentationen verfügen, um hohe kognitive Fähigkeiten zu erlangen. Vorstellungen oder Begriffe sind nicht neuronale Muster, sondern Attraktoren oder andere stabile Systemzustände (bzw. Systemverhalten) auf noch höherer Systemebene. Hierarchische Systemstrukturen (Systemebenen) ergeben sich aus der Dynamik dynamischer Systeme (vgl. Abschnitt 3.2.3, S. 43ff). Innerhalb neuronaler Strukturen gibt es nicht nur statische, verteilte Repräsentationen (Muster) (wie in Abschnitt 3.2.2 (S. 34ff) gezeigt), sondern aufgrund der Dynamik neuronaler Netze auch dynamisch-prozessuale Muster auf vielen hierarchisch verbundenen Ebenen. Vorstellungen sind Attraktoren, die sich aufgrund der prozessualen Strukturen, und diese wiederum aufgrund der Gehirnstruktur ergeben, sich aber auch gegenseitig beeinflussen und verändern. Vorstellungen oder Erinnerungen sind vor allem keine permanente physikalisch erkennbare Materie, sondern nur strukturinhärente, sich aus der Sequentialität (zeitlichen Kontinuität) der Prozesse ergebende Zustände, die innerhalb des Systems aufrechterhalten werden. Sie sind zwar keine Repräsentationen im klassischen Sinne, aber Repräsentationen im Sinne von ‚Repräsentation B‘ aus Kapitel 3.1.3 (S. 16ff).

5 SCHLUSSFOLGERUNGEN

- *Antirepräsentationalismus nur in Abgrenzung zu einem Repräsentationalismus möglich*

Antirepräsentationalisten lehnen Repräsentation für Kognition ab. Dazu muss aber festgelegt werden, was unter Repräsentation zu verstehen ist. Anhand des Vergleichs der Verwendung von Repräsentation für Kognition innerhalb antirepräsentationalistischer Ansätze mit einem zuvor definierten Repräsentationsbegriff kann entschieden werden, ob der jeweilige Ansatz als antirepräsentationalistisch gelten kann oder nicht.

Es zeigte sich, dass keiner der in dieser Arbeit vorgestellten sogenannten „antirepräsentationalistischen“ Ansätze als antirepräsentationalistisch bezeichnet werden sollte, da alle Ansätze implizite oder explizite Repräsentationen verwenden.

- *Ontologischer und methodologischer Antirepräsentationalismus*

Ich schliesse mich der Schlussfolgerung von Chemero (1999) an, dass man zwei Arten von Antirepräsentationalismus unterscheiden muss. Der methodologische (empirische) Antirepräsentationalismus geht davon aus, dass Kognition ohne einen Rekurs auf Repräsentation erklärbar ist. Der ontologische (metaphysische) Antirepräsentationalismus behauptet, dass es keine Repräsentation gibt.

- *Eine neue Art von Repräsentation*

Innerhalb dynamischer Systeme kann man von einer neuen Art von Repräsentation sprechen. Nicht mehr eine direkte Relation zwischen dem Inhalt einer mentalen Repräsentation und dem repräsentierten Gegenstand in der Welt ist von entscheidendem Interesse, sondern, wie der Inhalt einer mentalen Repräsentation im Gehirn implementiert ist und wie das Gehirn mit der Welt in Verbindung steht.

Stabilen Zuständen (z.B. Attraktoren) innerhalb eines dynamischen Systems können Bedeutungen zugeordnet werden und so als explizite Repräsentationen betrachtet werden. Aufgrund ihrer Entstehungsgeschichte innerhalb des Gesamtgefüges Mensch-Umwelt können diese Zustände aber auch implizite Bedeutung haben und

daher auch als implizite Repräsentationen angesehen werden. Aus diesem Grund sind dynamische Systeme weder methodologisch noch ontologisch antirepräsentational.

- *Antirepräsentationalismus als gemässiger Repräsentationalismus*

Die am häufigsten verwendete Art von Repräsentationen in der Kognitionswissenschaft sind Symbole. Diese haben keine extensionale, sondern intensionale Bedeutung. Diese kann explizit zugeordnet werden. Sie kann aber auch implizit auf Grund des Gebrauchs des Symbols vorhanden sein. Symbole sind diskrete, lokale Einheiten, die in klassischen Symbolsystemen manipuliert werden. Konnektionisten betrachten nicht mehr Symbole, sondern Subsymbole. Diese sind auf viele Neuronen verteilt. Daher ist ihr konzeptueller Inhalt von aussen nicht mehr erkennbar, obwohl implizit repräsentiert. Jedes Neuron, das an diesem Subsymbol beteiligt ist, verändert durch seine Aktivationsstärke den konzeptuellen Inhalt des Symbols. Auf diese Weise tragen neuronale Netze (z.B. Gehirne) unscharfen Begriffen und Ambiguitäten inhärent Rechnung, was in symbolischen Systemen gar nicht oder nur mit erheblichem Aufwand realisierbar ist. Die dynamische Systemtheorie schliesslich versucht ganz auf Repräsentation zu verzichten. Allerdings gelingt ihr das nicht, weil es auch in dynamischen Systemen stabile Systemzustände gibt, die ein bestimmtes Systemverhalten repräsentieren (z.B. Attraktoren). Je mehr Parameter (Dimensionen) das dynamische System hat, desto komplexer wird es. Dadurch kann es über viele solche stabilen Zustände (Repräsentationen) verfügen. Die Dimensionalität eines dynamischen Systems ist direkt proportional zur Repräsentationalität dieses Systems.

Man sieht, dass die erfolgreichsten Paradigmen der Kognitionswissenschaft – symbolische, konnektionistische und dynamische Systeme – als repräsentationalistisch eingestuft werden können. Man könnte auch sagen, je mehr Interaktion die Ansätze mit der Umwelt zulassen, desto repräsentationalistisch „gemässiger“ sind sie, weil sich dadurch die Art der Repräsentation (Form, Inhalt, Implementierung) verändert.

- *Vom Inhalt zur Form*

Innerhalb der Kognitionswissenschaft zeichnet sich ein methodologischer Wandel ab. Da die Inhalte mentaler Repräsentationen (intensionaler Gehalt im Sinne von Frege, nicht die extensionale Bedeutung) innerhalb dynamischer Systeme als stabile

Systemzustände physikalisch implementiert sind, werden nicht mehr die Inhalte mentaler Repräsentationen, sondern die Formen bzw. Strukturen innerhalb des Gehirns und mathematischer Modelle davon (Theorie dynamischer Systeme) untersucht. Auf diese Weise könnte die Kluft zwischen Mentalem und Physischem verringert oder sogar überwunden werden.

Kognition wird als dynamisches System betrachtet, dessen Struktur man erforscht. Ein Indiz dafür, dass Inhalt und Form identisch sind bzw. Bedeutung (d.h. der intensionale Gehalt einer Form) als reale, natürliche (z.B. physikalische) Struktur betrachtet werden kann, ist, dass jegliche Struktur für irgendeine andere Struktur potentiell kausal relevant ist. Anders gesagt: Jede Struktur X bedeutet potentiell für eine andere Struktur Y etwas – unabhängig davon, ob sich Y darüber bewusst ist oder nicht. Ob sich Y darüber bewusst ist oder nicht, hängt von der Komplexität der Struktur Y ab. Dies liegt am Komplexitätsunterschied der Strukturen (Organismen). Menschliche Kognition verfügt über eine komplexe Struktur (im Gegensatz z.B. zu der eines Bakteriums). Auch dynamische Systeme können sehr komplex sein und über eine Vielzahl stabiler Systemzustände auf vielen hierarchisch verschiedenen Systemebenen verfügen und daher menschliche Kognition erklären. Wir können aber nicht wissen, ob wir mit Hilfe der dynamischen Systemtheorie die menschliche Kognition jemals erschöpfend erklären oder menschliches Verhalten exakt vorhersagen können.

- *Klassische Symbolsysteme sind mit dynamischen Systemen vereinbar*

Dynamische Systeme werden von Dynamizisten als antirepräsentationalistisch bezeichnet. Dynamische Symbolsysteme sind jedoch ein Beweis dafür, dass auch dynamische Systeme etwas repräsentieren können. Stabile Zustände innerhalb dynamischer Systeme können als Repräsentationen betrachtet werden. Explizite Repräsentation gibt es, weil man diesen Zuständen Bedeutung zuordnen kann. Implizite Repräsentation gibt es aufgrund dessen, weil es stabile Zustände gibt, die das Systemverhalten repräsentieren.

Dynamische Systeme können daher weder in einem explanatorischen noch in einem ontologischen Sinne als antirepräsentationalistisch gelten. Zwar sind in dynamischen Systemen keine klassischen Symbole oder Subsymbole zu finden – und in diesem Sinne sind sie nicht repräsentational, aber man könnte die stabilen Zustände innerhalb eines dynamischen Systems als dynamische Symbole bezeichnen.

In dynamischen Systemen kann auch klassische Symbolverarbeitung durchgeführt werden, wenn auch nicht mit klassischen, sondern mit dynamischen Symbolen. Aufgrund der veränderten Sicht auf Repräsentation innerhalb dynamischer Systeme lassen sich sogar einige Probleme des klassischen Repräsentationalismus lösen. Auch Ansätze der *situated action* versuchen Symbolverarbeitung zu integrieren. Vielleicht ist es durch diese Symbiose möglich, derartige Synergieeffekte zu erreichen, dass der Ansatz neurodynamischer Symbolsysteme (NDSS) die menschliche Kognition und das Gehirn eines Tages erklären kann.

Eine der bekanntesten repräsentationalistischen Theorien des Geistes (RTG) ist jene von Jerry Fodor. Es stellt sich die Frage, ob Antirepräsentationalisten diese Theorie widerlegen oder eine plausiblere Theorie anbieten können, wonach Kognition ohne Repräsentation auskommt.

Es werden zumindest zwei Arten unterschieden, einem Fodorschen Repräsentationalismus zu widersprechen. Man kann behaupten, dass mentale Repräsentationen im Sinne Fodors nicht hinreichen, um intentionale Zustände zu realisieren. John Searle argumentiert, dass Syntax für Semantik nicht ausreicht. Allerdings trifft dieses Argument die RTG Fodors nicht, weil die mentalen Repräsentationen im Sinne Fodors bereits über einen bestimmten semantischen Inhalt verfügen, bevor sie syntaktisch verarbeitet werden.

Die zweite Möglichkeit ist, die Notwendigkeit von mentalen Repräsentationen zu leugnen. Daniel Dennett behauptet, dass Intentionen, die in einem System nicht explizit repräsentiert sind, dennoch das Verhalten dieses Systems kausal beeinflussen. Daher müssen intentionale Zustände nicht explizit repräsentiert sein – ganz im Gegensatz zu Fodor, dessen Motto lautet: „Keine intentionale Verursachung ohne explizite Repräsentation.“ (Fodor 1987, S. 25).

Beide Theorien, die RTG Fodors und die intentionale Sicht Dennetts, weisen Probleme auf, die durch den neuen Ansatz dynamischer Systeme möglicherweise umgangen werden können. Ich meine (auch wenn ich dieses Thema in Abschnitt 4.3 (S. 79ff) nur kurz angesprochen habe), dass es mit Hilfe dieses von Dynamizisten als antirepräsentationalistisch eingestuften Ansatzes möglich ist, intentionale Zustände mit physikalischen gleichzusetzen. Dabei werden Elemente verschiedenster philosophischer Richtungen zusammengeführt: die Token-Identität des anomalen Monismus Davidsons, Fodors kausale Semantiktheorie (über die physikalische Implementierung intensionaler Bedeutung haben Intentionen kausale Wirkung), Dennetts instrumentalistische Sicht der intentionalen Rede, Wittgensteins Sprachspiele und Gebrauchstheorie der Bedeutung,

Rortys Leugnen eines Spiegels der Natur im Kopf, Merleau-Pontys Theorie des intentionalen Bogens und der maximalen Vernunft und Heideggers In-der-Welt-sein. Einige dieser Elemente diese Theorien sind innerhalb des Ansatzes dynamischer Systeme in einer Gesamtheorie kohärent vereinbar.

Dabei muss der Anspruch, dass dynamische Systeme antirepräsentationalistische Systeme sind, aufgegeben werden. Es wäre eine interessante Aufgabe (im Rahmen einer weiteren Arbeit), die repräsentationalistische Theorie des Geistes von Jerry Fodor mit der Theorie dynamischer Symbolsysteme von Herbert Jäger (bzw. mit der Theorie dynamischer Systeme allgemein) zu vereinen. Dabei könnte sich herausstellen, dass Fodors RTG zu grossen Teilen innerhalb dynamischer Systeme Eingang findet, wobei die Schwächen der RTG durch inhärente Prinzipien dynamischer Systeme eliminiert werden.

FAZIT: Repräsentationalismus plausibler als Antirepräsentationalismus

Meine Untersuchungen bestätigen die These von Clark & Toribio (1994), dass Antirepräsentationalisten sich meist auf klassische, explizite Repräsentation, nicht aber auf implizite Repräsentation beziehen. Damit ist ein radikaler Antirepräsentationalismus, der besagt, dass Kognition ganz ohne Repräsentation möglich ist, nicht haltbar. Ausserdem scheinen kognitive Fähigkeiten, die Abstraktes oder Abwesendes betreffen, ohne die Speicherung von System- oder Umweltzuständen, d.h. Repräsentationen, nicht möglich zu sein. High-Level-Kognition erfordert repräsentationale Zustände. Jede Theorie, die für sich in Anspruch nimmt, menschliche Kognition zu erklären, kann auf stabile Zustände, die über eine gewisse Zeit Systemzustände oder Umweltzustände repräsentieren, nicht verzichten. Die Untersuchung einiger antirepräsentationalistischer Systeme bestätigte diese Hypothese. Insbesondere (die von Dynamizisten als antirepräsentationalistisch bezeichneten) dynamische Systeme (und alle Systeme können als dynamisch betrachtet werden) verfügen über Zustände, die als Repräsentationen bezeichnet werden können. Es ist also hinsichtlich der Erklärung von Kognition eher an einem Repräsentationalismus als an einem Antirepräsentationalismus festzuhalten. Abbildung 5.7.1 auf S. 93 zeigt abschliessend, wie ich die verschiedenen Ansätze der Kognitionswissenschaft unter Einbeziehung der Ergebnisse dieser Arbeit sehe.

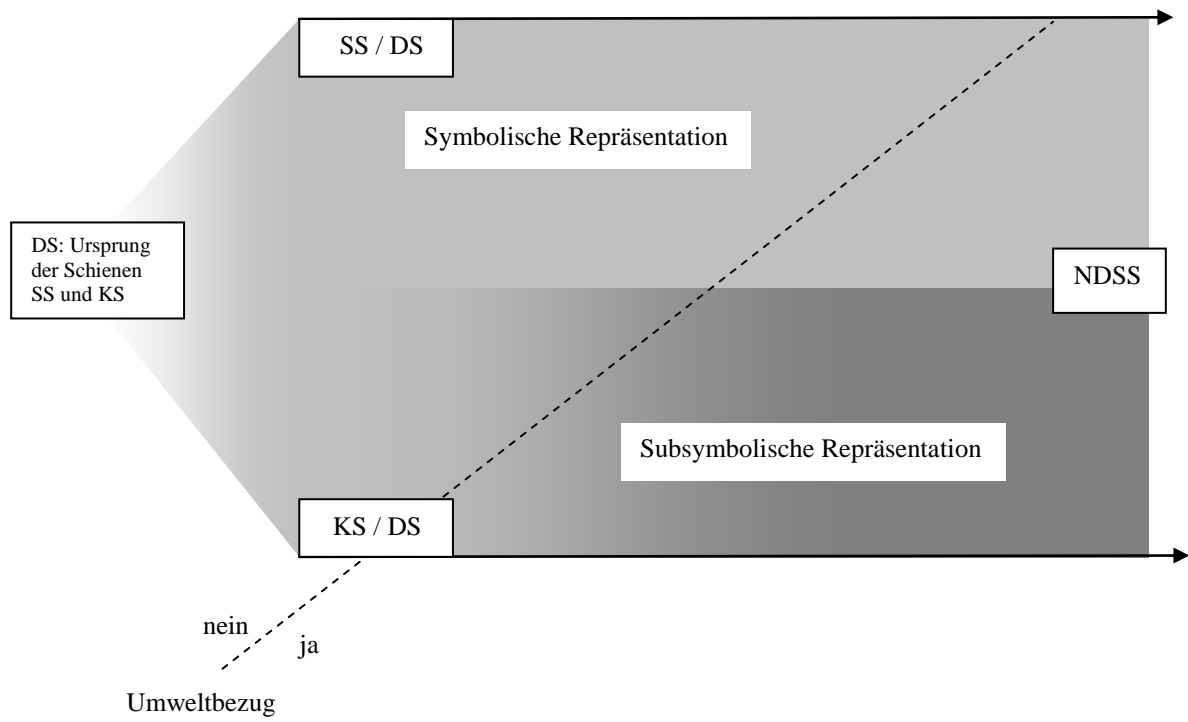


Abbildung 5.7.1: Einordnung der kognitionswissenschaftlichen Ansätze unter Einbeziehung der Ergebnisse dieser Arbeit (die unterschiedliche Schwärze zeigt die Verschiedenheit der verwendeten Repräsentationen an: hell entspricht eher symbolischer bzw. expliziter Repräsentation und dunkel eher subsymbolischer bzw. impliziter Repräsentation)

Literaturverzeichnis

- Agre, P.E., & Chapman, D. (1987). Pengi: An implementation of a theory of activity. In: *Proceedings of the Sixth National Conference on Artificial Intelligence*. 268-272. Menlo Park, CA: American Association for Artificial Intelligence.
- Anderson, J. R. (1983). *The architecture of cognition*. Cambridge, MA: Harvard University Press.
- Argyris, J., & Faust, G., & Haase, M. (1995). *Die Erforschung des Chaos*. Vieweg-Verlag.
- Barsalou, L.W. (1999). Perceptual symbol systems. In: *Behavior and Brain Sciences* 22, 577-660.
- Beckermann, A. (1999). *Analytische Einführung in die Philosophie des Geistes*. De Gruyter. Berlin, New York.
- Beer, R., & Chiel, H. (1997). The brain has a body: Adaptive behaviour emerges from interactions of nervous system, body and environment. In: *Trends in Neuroscience*, 20(12), 553–557.
- Beer, R., & Gallhager, J. C. (1992). Evolving Dynamical Neural Networks for Adaptive Behavior. In: *Adaptive Behavior* 1, 91-122.
- Borchard, C. (2000). *Herausforderung: Biocomputer*. Olms.
- Boros, J. (1999). Repraesentationalismus und Antirepraesentationalismus. Kant, Davidson und Rorty. In: *Deutsche Zeitschrift für Philosophie*, Berlin 47, 1999/4, 539-551.
- Bossel, H. (1989). *Simulation dynamischer Systeme: Grundwissen, Methoden, Programme*. Braunschweig, Wiesbaden: Friedrich Vieweg & Sohn.
- Brooks, R. 1991. Intelligence without representation. In: *Artificial Intelligence* 47, 139-159.
- Brooks, R. 1997. Intelligence without representation. In: *Haugeland 1997, Mind Design II*. MIT Press.
- Busemeyer, J., & Townsend, J. (1993). Decision Field Theory: A Dynamic-Cognitive Approach to Decision Making in an Uncertain Environment. In: *Psychological Review*, 1993, 432-459.
- Cann, R. (1993). *Formal semantics: An introduction*. Cambridge: Cambridge University Press.

- Chemero, A. (1999). Empirical and Metaphysical Anti-Representationalism. In: Riegler, A., & Peschl, M., & Stein, A. (Hrg.) (1999). *Understanding Representation in the Cognitive Sciences. Does Representation Need Reality?* NY: Kluwer.
- Clancey, W.J. (1997). *Situated Cognition: On Human Knowledge and Computer Representations*. Cambridge: Cambridge University Press.
- Clark, A. (1993). *Associative Engines: Connectionism, Concepts and Representational Change*. MIT Press.
- Clark, A. (2001). *Mindware – An Introduction to the Philosophy of Cognitive Science*. Oxford: Oxford University Press.
- Clark, A., & Toribio, J. (1994). Doing without representing. In: *Synthese* 101, 401-31.
- Clauß, G. (1999). *Statistik für Soziologen, Pädagogen, Psychologen und Mediziner: Grundlagen*. 3. Aufl. Thun und Frankfurt am Main: Verlag Harri Deutsch.
- Cummins, R. (1989). *Meaning and Mental Representation*. Cambridge, MA: MIT Press.
- Cummins, R. (1996). *Representation, Targets and Attitudes*. Cambridge, MA: MIT Press.
- Damasio, A. (1994). *Descartes' Error*. New York: Grosset, Putnam.
- Dancy, J., & Sosa, E. (eds.) (1992). *A Companion to epistemology*. Oxford: Blackwell.
- Davidson, D. (1984). Belief and the Basis of Meaning. In: Davidson, D. (1984). *Inquiries into Truth and Interpretation*. Oxford University Press. 141-154.
- Dennett, D. (1987). *The Intentional Stance*. Cambridge, MA: MIT Press.
- Descartes, R. (1972). *Meditationen mit sämtlichen Einwänden und Erwiderungen*. Hamburg: Meiner.
- Dreyfus, Hubert, L. (1998). Merleau-Ponty's Critique of Mental Representation: The Relevance of Phenomenology to Scientific Explanation: Intelligence Without Representation. *Houston Studies in Cognitive Science*.
- Eco, U. (1972). *Einführung in die Semiotik*. München: UTB, Fink.
- Eliasmith, C. (1996). The third contender: A critical examination of the dynamicist theory of cognition. In: *Philosophical Psychology*. Vol. 9, No. 4, 441-463. Reprinted In: Thagard, P. (ed) (1998). *Mind Readings: Introductory Selections in Cognitive Science*. MIT Press.
- Fodor, J. A. (1975). *The Language of Thought*. Cambridge, MA: Harvard University Press.
- Fodor, J. A. (1987). *Psychosemantics*. Cambridge: MIT Press.

- Fodor, J. A., & Pylyshyn, Z. W. (1988). Connectionism and cognitive architecture: A Critical Analysis. In: *Cognition*. 28, 3-71.
- Franklin, S. (1995). Artificial minds. MIT Press.
- Freeman, W., & Skarda, C. A. (1990). Representations: who needs them? In: McGaugh, J., Weinberger, J., & Lynch, G. (1990). Brain Organization and Memory. Guilford Press.
- Frege, G. (1892). Über Sinn und Bedeutung. In: *Zeitschrift für Philosophie und philosophische Kritik*, 25-50. Wiederabgedruckt In: Patzig, G. (Hrg.) (⁴1975). Funktion, Begriff, Bedeutung. Göttingen: Vandenhoeck und Ruprecht. 40-65.
- French, R. M., & Anselme, P. (1999). Interactively converging on context-sensitive representations: A solution to the frame problem. In: *Revue Internationale de Philosophie*, 3, 365-385.
- Gärdenfors, P. (2000). Conceptual spaces: the geometry of thought. Cambridge, MA: MIT Press.
- Grush, R. (1997). The architecture of representation. In: *Philosophical Psychology* 10, 5-23.
- Guastello, S. (1995). Chaos, catastrophe, and human affairs: applications to nonlinear dynamics to work, organizations, and social evolution. New Jersey: Erlbaum.
- Haken, H. (1982). Synergetik. Eine Einführung. Heidelberg: Springer.
- Harnad, S. (1990). The Symbol Grounding Problem. In: *Physica D* 42, 335-346.
- Haugeland, J. (1985). Artificial Intelligence: The very Idea. Cambridge, MA: MIT Press.
- Haugeland, J. (ed.) (1997). Mind Design II. Cambridge, MA: MIT Press.
- Hebb, D. O. (1949). The organization of behaviour. New York: Wiley.
- Heidegger, M. (1927). Sein und Zeit. 13. unv. Aufl. 1976. Tübingen: Max Niemeyer Verlag.
- Kant, I. (1787). Kritik der reinen Vernunft. Neuhrsg. von ²1787 d. Heidemann, I. (Hrg.) Reclam. Stuttgart.
- Jäger, H. (1995). An Introduction to Dynamic Symbol Systems. In: Hallam, J. (ed.): Hybrid Problems, Hybrid Solutions. Proceedings of the AISB-95. IOS Press/Ohmsha, Amsterdam 1995, 109-120.
- Jäger, H. (1996). Dynamische Systeme in der Kognitionswissenschaft. In: *Kognitionswissenschaft* 5, 151-174.
- Keller, R. (1995). Zeichentheorie. Tübingen: Francke.

- Keijzer, F. A. (1998). Doing without representations which specify what to do. In: *Philosophical Psychology* 11, 269-302.
- Kelso, S. (1995). *Dynamic Patterns: The Self-Organization of Brain and Behavior*. Cambridge, MA: MIT Press.
- Krieger, D. J. (1998). *Einführung in die allgemeine Systemtheorie*. München: Fink.
- Law, L-C. (1993). Symbolic processing vs. situated action: A dialectical synthesis? (Research report No. 24). München: Ludwig-Maximilians-Universität, Lehrstuhl für Empirische Pädagogik und Pädagogische Psychologie.
- Lewis, D. (1983). *Radical Interpretation*. In: Lewis, D. *Philosophical Papers I*. 1983. Oxford University Press. 108-121.
- Marcus, G. F. (2001). *The Algebraic Mind*. Cambridge, MA: MIT Press.
- Markman, A. B., & Dietrich, E. (2000). Extending the classical view of representation. In: *Trends in Cognitive Sciences*, Vol. 4, No. 12, December 2000, 470-475.
- Maturana, H., & Varela, F. (1987). *Der Baum der Erkenntnis*. München: Goldmann Verlag.
- Merleau-Ponty, M. (1967). *Phenomenology of Perception*. London: Routledge & Kegan Paul.
- Metzinger, T. (1998). Personen als dynamische Repräsentationssysteme. In: Sandkühler, H. J., (1998). *Repräsentation, Denken und Selbstbewußtsein*. Bremen: Zentrum f. philosophische Grundlagen der Wissenschaften, 73-124.
- Minsky, M., & Papert, S. (1969). *Perceptrons*. Cambridge, MA: MIT Press.
- Newell, A. (1980). Physical symbol systems. In: *Cognitive Science*, 4, 135-183.
- Newell, A., & Simon, H. A. (1981). Computer science as empirical inquiry: Symbols and search. In: *Communications of the Association for Computing Machinery* 19, 113-26. Reprinted In: Haugeland, J. (ed) (1981) *Mind Design*. MIT Press.
- Pasemann, F., (1996). Repräsentation ohne Repräsentation – Überlegungen zu einer Neurodynamik modularer kognitiver Systeme. In: Rusch, G., Schmidt, S. J., & Breidbach, O. (eds.) (1996) *Interne Repräsentationen – Neue Konzepte der Hirnforschung*. Frankfurt: Suhrkamp. 42-91.
- Plunkett, K., & Sinha, C. (1992). Symbol Grounding or the Emergence of Symbols? Vocabulary Growth in Children and a Connectionist Net. In: *Connection Science*, Vol. 4, Nos. 3&4, 1992, 293-312.
- Port, R. F., Van Gelder, T. (eds.) (1995). *Mind as Motion: explorations in the dynamics of cognition*. Cambridge, MA: MIT Press.

- Quine, W. V. (1960). *Word and Object*. Cambridge, MA: MIT Press.
- Riegler, A., & Peschl, M., & Stein, A. (Hrg.) (1999). *Understanding Representation in the Cognitive Sciences. Does Representation Need Reality?* NY: Klouwer.
- Ritter, J. (1992). *Wörterbuch der Philosophie. Band 8*. Basel: Schwabe.
- Rorty, R. (1979). *Philosophy and the Mirror of Nature*. Princeton: Princeton University Press.
- Rosenblatt, F. (1962). *Principles of neurodynamics*. New York: Spartan.
- Rumelhart, D. E., & McClelland, L. J. & the PDP Research Group. (1986). *Parallel Distributed Processing. Explorations in the Microstructure of Cognition. Volume 1: Foundations*. Cambridge, MA: MIT Press.
- Russell, B. (1918/19). The Philosophy of Logical Atomism. In: *The Monist* 28/29.
- Scheerer, E. (1992a). Repräsentation. In: Ritter, J. (1992). 790-798.
- Scheerer, E. (1992b). Repräsentation. In: Ritter, J. (1992). 834-846.
- Searle, J. (1980). Minds, brains and programs. In: *Behavioral and Brain Sciences* 3, 417-424.
- Searle, J. (1992). *The Rediscovery of Mind*. Cambridge, MA: MIT Press.
- Smolensky, P. (1986). Foundations of Harmony Theory. In: Rumelhart, D. E. (1986), *Parallel Distributed Processing. Explorations in the Microstructure of Cognition. Volume 1*. 194-281.
- Smolensky, P. (1988). On the proper treatment of connectionism. In: *Behavioral and Brain Sciences* 11, 1-23.
- Smolensky, P. (1990). Tensor product variable binding and the representation of symbolic structures in connectionist systems. In: *Artificial Intelligence* 46, 159-216.
- Stufflebeam, R. S. (1998a). Representation and computation. In: W. Bechtel & G. Graham (Eds.) *A companion to cognitive science*. Oxford: Basil Blackwell. 636-648.
- Stufflebeam, R. S. (1998b). Whither internal representations? In defense of antirepresentationalism and other heresies. (Abstract).
URL=<http://www.uno.edu/~rstuffe/pages/dissertation.html>
- Thagard, P. (1999). *Kognitionswissenschaft – Ein Lehrbuch*. Stuttgart: Klett-Cotta Verlag.
- Thelen, E., & Smith, L. B. (1994). *A dynamic systems approach to the development of cognition and action*. Cambridge: MIT Press.

- Van Gelder, T. J. (1990). Compositionality: A connectionist variation on a classical theme. In: *Cognitive Science*, 14, 355-384.
- Van Gelder, T. J. (1995). What might cognition be if not computation? In: *Journal of Philosophy* 92, 345-381.
- Van Gelder, T. J. (1997). Dynamics and Cognition. In: Haugeland, J. (ed) (1997). *Mind Design II*. 421-450.
- Van Gelder, T. J. (1998). The dynamical hypothesis in cognitive science (with commentaries). In: *Behavioral and Brain Sciences* 21, 615-657.
- Van Rooij, I.J.E.I., & Bongers, R.M., & Haselager, W. F. G. (2000). The dynamics of simple prediction: Judging reachability. In: *Proceedings of the 22nd Annual Conference of the Cognitive Science Society*, 2000, 535-540.
- Varela, F., & Thompson, E., & Rosch, E. (1991). *The Embodied Mind*. Cambridge, MA: MIT Press.
- Vera, A., & Simon, H. (1993). Situated action: A symbolic interpretation. In: *Cognitive Science* 17, 4-48.
- Wittgenstein, L. (¹²1999). *Philosophische Untersuchungen*. Werksausgabe Band 1. Frankfurt: Suhrkamp Verlag.