Ein abstraktes Modell menschlicher Intelligenz– Die Konstruktion eines künstlichen kognitiven Agenten

Simon Strübbe

Inhaltsverzeichnis

1 Einleitung

Was ist Intelligenz? Ist diese Frage einerseits so alt wie die Philosophie selbst, stellte sie sich in den fünfziger Jahren des 20. Jahrhunderts mit der Entwicklung frei programmierbarer Computer in einem neuen Gewand. Ein Computer ist fähig komplizierte mathematische Berechnungen auszuführen, zu denen vorher nur der Mensch fähig war. So kam im Rahmen der in den fünfziger Jahren neu eingeführten wissenschaftlichen Disziplin der künstlichen Intelligenz die Frage auf, ob ein Computer imstande ist, jegliche Form der menschlichen Intelligenz nachzubilden. Woraus nährte sich diese Hoffnung? Schon vor dem Aufkommen frei programmierbarer Computer zeigte Allen Turing, dass schon durch die Kombination einfacher Berechnungsschritte alles berechenbar ist, was aus mathematischer Sicht als berechenbar gilt. Er dachte sich mit den Turingmaschinen ein einfaches Modell eines Automaten aus, und zeigte, dass man eine universelle Turingmaschine konstruieren kann, welche das Verhalten aller Turingmaschinen simulieren kann. Um einen anderen Automaten (oder auch sich selbst) simulieren zu können, übergibt man der universellen Turingmaschine eine formale Beschreibung des Automaten und dessen Input. Auf Grundlage dieser Idee sind frei programmierbare Computer entstanden, denn Turing zeigte, dass man Automaten bauen konnte, deren Verschaltung fest vorgegeben ist, und die Berechnung nur auf der Programmebene ablaufen kann. Praktisch alle heutigen Computer können als universelle Turingmaschinen verstanden werden, mit der Einschränkung eines endlichen Speichers und endlicher Rechengeschwindigkeit.

Hatte man mit der Einführung des Computers ein Instrument alle nur denkbaren Berechnungen auszuführen, so musste man das menschliche Gehirn nur als Automaten verstehen, der auf physikalischen Prinzipien beruht und von der Evolution geformt wurde, um zu postulieren, dass ein Computer jegliche Form des menschlichen Denkens nachahmen könnte. Die frühe Euphorie der fünfziger Jahre fand allerdings schnell ein Ende, da man einerseits nicht wusste, wie das Gehirn seine Berechnungen vornimmt, und andererseits die Rechenkraft des menschlichen Gehirns exorbitant unterschätzt hatte.

Heute im Jahr 2016, wo dieses Buch entstanden ist, hat sich die Situation etwas verändert. Es ist heute gängig die Rechenleistung des menschlichen Gehirns abzuschätzen. Das menschliche Gehirn hat ca. 100 Milliarden Neuronen und ca. 100 Billionen Synapsen. Eine einfache Abschätzung, welche die Verbindungsstärke zwischen den Neuronen, welche in den Synapsen gespeichert ist, als eine einzige Zahl darstellt, kommt auf eine Rechenleistung von 20

Billiarden Rechenoperationen pro Sekunde, wenn man berücksichtigt, dass ein Neuron seinen Zustand 200 mal in der Sekunde neu berechnet. Heutige Supercomputer durchbrechen diese Rechenleistung.

Damit die künstliche Intelligenz ihr Versprechen einer intelligenten Maschine einlösen kann, müsste man neben der reinen Rechenleistung auch wissen, wie das menschliche Gehirn funktioniert. Hier tappt die Wissenschaft weitestgehend im Dunkeln. Relativ leicht zugänglich sind die Inputs und Outputs des menschlichen Gehirns, so dass man heute einerseits ein gutes Verständnis über die ersten Rechenschritte hat, die unmittelbar nach der menschlichen Sensorik stattfinden, und andererseits weiß, wie der Mensch seine Motorsignale steuert. Nur was in der Mitte geschieht, was die eigentliche Intelligenz ausmacht, ist nach wie vor unklar.

Kann man sich einerseits durch neurologische Experimente dem Thema Intelligenz nähern, kann man anderseits einen Roboter konstruieren, und schauen, auf welche Probleme dieser Roboter in der Alltagswelt stößt, um diese Probleme dann aus technischer Sicht zu lösen. Beide Ansätze werden in der heutigen Forschung intensiv verfolgt.

Das vorliegende Buch ist anders entstanden. Es nähert sich dem Thema aus philosophisch, logischer Sichtweise und entwirft ein Grobschema von Intelligenz vom Reißbrett. Tritt das Buch hiermit in die Fußstapfen manch anderer Autoren[7], die die Quintessenz menschlicher Intelligenz erfassen wollen, so ist der vorliegende Ansatz umfangreicher. Gewöhnlich wird eine Eigenschaft des menschlichen Denkens genommen, wie die Eigenschaft Muster erkennen zu können, um diese dann auf das ganze Gehirn zu übertragen. Im Beispiel der Mustererkennung ergibt sich dabei ein hierarchisches Modell, welches auf unterer Ebene Muster in der Wahrnehmung erkennt und auf höherer Ebene wiederum Muster in der Verarbeitung dieser Muster.

Nicht unähnlich dem allgemeinen Verständnis von Intelligenz, werden zwei Eckpunkte ausgemacht, die das untere Ende und das obere Ende der menschlichen Informationsverarbeitung markieren. Am unteren Ende ist die direkte Interaktion mit der physischen Außenwelt angesiedelt und am oberen Ende der Mechanismus des Ich-Bewusstseins. Wäre es logisch, am unteren Ende mit der Beschreibung anzufangen und sich zum Ich-Bewusstsein vorzuarbeiten, ist das Modell genau andersherum entstanden. Am Anfang stand die Analyse der Selbstreferenz in der mathematischen Logik, um diese dann auf das Reflexionsvermögen des menschlichen Geistes zu übertragen. Von diesem Reflexionsvermögen ausgehend, wurde ein Weg gesucht, wie einfache Interaktionen mit der Umwelt schließlich die Eigenschaften dieses Reflexi-

onsvermögens hervorbringen. Erst grob, dann immer detaillierter. In diesem vorliegenden Buch wird das Modell erstmalig vorgestellt. Hierbei wird der Schuh allerdings wieder in der "richtigen" Reihenfolge aufgezogen, und nach einem kurzen Überriss, wo es hingehen soll, arbeitet sich das Buch vom unteren Ende zum oberen Ende der Informationsverarbeitung fort.

Der Ausgangsfrage folgend, soll der Einstieg über ein von Turing ausgedachten Experiments gefunden werden. Eine Maschine soll als dem Menschen ebenbürtig intelligent gelten, wenn ein Mensch durch sprachliche Kommunikation mit der Maschine nicht entscheiden kann, ob er es mit einem Menschen oder einer Maschine zu tun hat. Damit die äußere Form der Maschine keine Rolle spielt, sitzen die beiden Gesprächspartner in verschiedenen Räumen und kommunizieren über eine Tastatur.

Turing erkannte früh, dass durch die Sprache die komplette Bandbreite menschlicher Intelligenz repräsentiert wird. Mit Sprache lassen sich nicht nur Sätze über die physische Außenwelt formulieren, sondern auch über alle internen mentalen Zustände des Menschen oder einer fiktiven Maschine. Darüber hinaus ist Sprache so flexibel, dass mit ihr auch über die Sprachelemente selbst gesprochen werden kann. Es läge daher nahe, sich dem menschlichen Geist über ein Verständnis der Sprache zu nähern. Dies ist bislang nicht gelungen.

Betrachtet man die natürliche Sprache für sich, ohne darauf zu schauen, was mit ihr ausgedrückt werden soll, kann nur ihre grammatikalisch logische Struktur untersucht werden. Dies ist hinreichend geschehen, wobei man bei der Analyse der Sprache versucht, diese auf eine formale mathematische Sprache abzubilden: Die mathematische Logik. Das Hauptproblem hierbei ist, dass das sprach-verarbeitende System kein Verständnis über die Welt hat, also die Dinge, über die gesprochen wird, so dass das System über Interpretationsprobleme stolpert, die dem Menschen beim Benutzen der Sprache gar nicht bewusst sind.

Mehrere Jahrzehnte hat man mit mäßigen Erfolg versucht, diese Interpretationsprobleme in den Griff zu bekommen, indem man dem System eine Form von "Weltwissen" vorgibt. Dies geschah wiederum mithilfe der mathematischen Logik, indem man als eine Art Zusatzwissen logische Zusammenhänge formulierte, wie "Feuer ist heiß" oder "Wasser ist nass". Auch bei diesem Zusatzwissen, weiß das System nicht, was "Feuer" oder "Wasser" ist, da es nie Erfahrungen über die reale Welt gesammelt hat.

Vielversprechender erscheint da der Ansatz eine Maschine in Form eines Roboters mit der Welt interagieren zu lassen, um so Konzepte über die Welt

zu lernen, über die sich die Maschine dann unterhalten kann. Auch das hier beschriebene Modell setzt bei der Interaktion mit der Welt an.

Um einen kognitiven Agenten zu programmieren, wie eine solche Maschine im folgenden genannt werden soll, ist es nötig, dass gesamte Denken zu strukturieren. Hinter jedem kognitiven Agenten steckt ein mathematisches Modell, welches nach einer gewissen Logik aufgebaut ist. Genau hier unterscheidet sich das menschliche Denken von einem kognitiven Agenten. Ist das menschliche Denken in den ersten Schritten bei einer Problemlösung konfus, und wird erst bei weiterem Nachdenken strukturierter, so ist das Denken eines kognitiven Agenten von vornherein auf eine gewisse Art strukturiert. Und genau hier liegt die Kunst bei der Programmierung eines kognitiven Agenten, solche strukturellen Unterscheidungen im Aufbau des Modells zu finden, die bei näherer Betrachtung trennscharf sind, und um die der Mensch beim Denken ebenfalls nicht umhinkommt. Um diese Grundstrukturierung eines kognitiven Agenten geht es in diesem Buch.

Das Denken eines kognitiven Agenten wird dabei in aufeinander aufbauenden Phasen beschrieben. Die erste Phase beschreibt die Art der Konzepte, die ein Agent über die physische Welt lernen kann. Es wird gezeigt, dass man mit drei Konzepttypen auskommt, um die Interaktion mit der Welt zu beschreiben. Nicht nur lässt sich alles Wissen über die physische Welt unter diese Konzepttypen subsumieren, sie weisen zudem die Trennschärfe auf, dass auch das menschliche Denken nicht umhinkommt diese Unterscheidungen zu treffen. Bereits hier wird über den aktuellen Stand der künstlichen Intelligenz Forschung hinausgegangen. Hat man es gewöhnlich beim Konzeptlernen mit einem *einzelnen* Klassifizierungsproblem zu tun, bietet das Modell eine Zerlegung in drei ineinander geschachtelten Klassifizierungsproblemen an. Reduziert sich hierdurch einerseits der Gesamtaufwand des lernenden Systems, so zeigt sich die Stärke beim darauf folgenden Verallgemeinern der Konzepte.

Das Verallgemeinern von Konzepten, so wie andere mögliche Operationen mit physischen Konzepten ist Thema von Phase II. Phase III beschreibt daraufhin wie der kognitive Agent, dass bis dato entwickelte Denken reflektieren kann, also das Denken über das Denken. Dies geschieht über die Entwicklung mentaler Konzepte, die sich an der Grundstruktur der physischen Konzepte orientieren. Hierbei ist das Denken des kognitiven Agenten stets auf den Inhalt der Konzepte gerichtet, nicht auf deren syntaktischen logischen Struktur. Bauen heutige Metakognitionsmodelle auf der reinen syntaktischen Struktur der Wissensrepräsentation auf, erlaubt der hier entwickelte Ansatz eine andere Art der Reflexion. So wird in Phase III ein Modell des Ich-Bewusstseins

entwickelt, welches auf dieser Art der Reflexion aufbaut.

Phase IV beschreibt, wie die Elementarkonzepte über das eigene "Ich" zu einem Selbstmodell zusammengeführt werden. Hierbei steht die Selbstmodelltheorie von Thomas Metzinger Pate, dessen Ideen entscheidenden Einfluss auf das hier entstandene Modell hatten. Erst in der fünften Phase kommt ein Kommunikationsmodul hinzu, mithilfe dessen sich der kognitive Agent über die entwickelten physischen und mentalen Konzepte austauschen kann. Der Abschluss des Buches bildet eine Betrachtung über die Moral von Maschinen.

2 Bewusstsein und Ich-Bewusstsein

Ist Bewusstsein etwas Unerklärbares? Hat nur der Mensch bewusstes Erleben? Wie könnte eine Theorie des Bewusstseins aussehen? Auch wenn das vorliegende Buch die Kognition von "unten" her beschreiben will, von der Interaktion mit der physischen Welt ausgehend, soll kurz angerissen werden, wie das Phänomen Bewusstsein, dem man die höchste Form der Kognition zuschreibt, in diesem Buch behandelt wird.

Dem Bewusstsein haftet etwas Mystisches an, da es traditionell von Philosophen der objektiven Welt gegenübergestellt wird. Die Welt besteht demnach aus Wahrheiten, die unentdeckt vor sich hin schlummern, bis ein Subjekt von diesen Kenntnis nimmt. In diesem Moment scheint es zwei Seinsbereiche zu geben: Die objektive Wahrheit und das reine Wissen über dieselbe. So unvereinbar diese zwei Seinsbereiche sind, so nebulös ist hier der Begriff des Subjekts. Ist die Welt objektiv gegeben, so setzt das Wissen über dieselbe ein Subjekt voraus, welches die Welt bewusst wahrnimmt. Entwickelt man den Subjektbegriff aus der Gegenüberstellung von "Welt" und "Wissen von der Welt" heraus, so bleibt das Verständnis was ein "Ich" ist, ebenso unklar, so ausweglos es erscheint in der Gegenüberstellung den einen Part auf den Anderen zu reduzieren.

Der Weg, der hier gegangen wird, ist, den Begriff des Bewusstseins aus dem Elfenbeinturm zu holen und radikal zu trivialisieren. Hierbei müssen zwei Phänomene voneinander getrennt werden: Das Bewusstsein und das Ich-Bewusstsein. Das Bewusstsein wird dabei als eine Informationsverarbeitungsstrategie behandelt und das Ich-Bewusstsein als das Vermögen des Gehirns, zu beliebigen mental repräsentierten Inhalten einen Ich-Bezug herzustellen.

Hierzu wird der Begriff des Bewusstseins nicht der objektiven Welt gegenübergestellt, sondern dem Unbewussten. Es wird postuliert, dass gewisse

Inhalte einer besonderen Form der Verarbeitung zugeführt werden. Geschieht viel der Verarbeitung im menschlichen Gehirn parallel, so besitzt dieses ebenfalls eine serielle Informationsverarbeitung. Nicht nur, dass bei dieser seriellen Informationsverarbeitung das gerade Berechnete zur Weiterverarbeitung zur Verfügung steht, das Gehirn verfügt zusätzlich über ein Arbeitsgedächtnis, das Inhalte bereitstellt, die entweder gerade erzeugt wurden, oder für das gerade bearbeitete Problem von Wichtigkeit sein könnten.

Es ist geradezu auffällig, dass uns alle Sachen bewusst werden, die von Aufmerksamkeitsfiltern als wichtig erachtet werden, und dass andersherum nur die Sachen bewusst werden können, die diese Filter positiv durchlaufen haben. Aufmerksamkeit und Bewusstsein gehören somit zum selben Prozess.

Die Trivialisierung des Begriffs des Bewusstseins besteht demzufolge darin, dass das bewusste Denken die Innenwahrnehmung dieser seriellen Informationsverarbeitung ist. Hierbei soll zwischen zwei Arten der Information unterschieden werden: Information, die potentiell bewusst gemacht werden kann, und Information, die nicht dieses Potential hat.

Im weiteren Verlauf des Buches wird dem Denken auf die Art eine Struktur gegeben, dass Information nicht für sich alleine stehen kann, sondern dass Information nur in Form von Konzepten gedacht werden kann. Es wird die Aufgabe des Buches sein, die möglichen Konzepttypen offenzulegen, und deren logische Struktur zu analysieren. Demzufolge besteht die Information, die potentiell bewusst gemacht werden kann, einerseits aus dem internen Abbild der physischen Außenwelt und andererseits aus den Konzepten, die aus der Interaktion mit der Außenwelt abgeleitet werden. Die Konzepte unterteilen sich dann nochmal, ob diese sich auf die physische Außenwelt beziehen oder Konzepte sind, die das Arbeiten mit Konzepten beschreiben. Die erste Art von Konzepten beschreiben somit das Denken über die physische Welt (Phase I und II) und die zweite Art von Konzepten das Denken über das Denken (Phase III).

Die Information, die nicht bewusst gemacht werden kann, ist die Arbeitsweise von den Schaltkreisen, die erstens das interne Abbild der Außenwelt erstellen und zweitens die Arbeitsweise der Schaltkreise, die Konzepte erstellen. Z.B. kann das Gehirn nicht bewusst darauf zugreifen, wie in den ersten Arbeitsschritten der visuellen Verarbeitung Kanten im wahrgenommenen Bild herausgefiltert werden.

Ist der bewussten Verarbeitung nicht zugänglich, wie Konzepte gebildet werden, diese "poppen" einfach unverhofft im Bewusstsein auf, so kann das Gehirn das Hantieren mit Konzepten "beobachten". Z.B. kann es nachver-

folgen, wie Konzepte unter andere Konzepte subsumiert werden, oder zwei Konzepte verknüpft werden. Diese "Beobachtung" der Konzeptverarbeitung soll in dem hier beschriebenen Modell nichts Mystisches anhaften. Sie wird dadurch erklärt, dass das Beobachten der Konzeptverarbeitung von der Funktion her ein Kontrollmechanismus ist, und das, was dort stattfindet ein weiterer Klassifizierungsprozess ist: Das bereits erwähnte Bilden von Konzepten über die eigene Konzeptverarbeitung.

Das Phänomen Ich-Bewusstsein ist, dem hier vorgestellten Modell nach, das Auftreten zweier Umstände. Erstens das Herstellen eines Ich-Bezuges und zweitens, dass dieser Ich-Bezug in der seriellen Verarbeitung des Bewusstseins stattfindet. Es wird angenommen, dass Ich-Bezüge auch hergestellt werden können, wenn auf diesen nicht die Aufmerksamkeit beruht. Beruht auf ihnen die Aufmerksamkeit, wird der Ich-Bezug im Bewusstsein explizit dargestellt und erzeugt somit Ich-Bewusstsein, was bedeutet, dass das informationsverarbeitende System eine Repräsentation davon hat, in welchem Verhältnis das Dargestellte zu "mir" steht.

Auf welche Weise Ich-Bezüge hergestellt werden können wird noch ausführlich Thema dieses Buches sein. Es wird postuliert, dass es nur zwei Arten des Ich-Bezuges gibt, die sich logisch unterscheiden. Grundlage dessen ist ein Handlungsbewertungsschema (siehe Abbildung 1). Es beschreibt, wie ein Subjekt auf einen Objektbereich einwirken kann, und wie es eine entstandene Situation bewertet. Das Handlungsbewertungsschemata wird sowohl benutzt um das Denken über die physische Außenwelt zu beschreiben, als auch das Denken über das Denken. Im ersten Fall ist der Objektbereich, auf den das Subjekt einwirkt die physische Außenwelt, im zweiten Fall ist der Objektbereich das mentale Arbeiten mit Konzepten.

Die logische Form des Subjekts unterscheidet sich bei der Einwirkung auf einen Objektbereich von der anschließenden Bewertung. Das Einwirken auf einen Objektbereich wird am einfachsten an einem mathematischen Operator verdeutlicht. Ein Operator hat einen Objektbereich vor sich, an dem er eine Operation vornimmt, um den Zustand A in den Zustand B zu überführen. Wichtig hierbei ist, das sowohl der Objektbereich eine Struktur oder Eigenschaften aufweist, als auch der Operator durch Eigenschaften beschrieben werden kann. Das Subjekt, welches die Operation vornimmt, bleibt in dieser Darstellung eigenschaftslos. Dieser Subjekttyp soll im folgenden Pointersubjekt genannt werden, was zum einen die Eigenschaftslosigkeit zum Ausdruck bringt und andererseits auf einen Urheber der Handlung hinweist.

Bei der Bewertung einer Situation hat das hier auftretende Subjekt eine

andere logische Form. Das System bewertet eine Situation danach, ob diese gut oder schlecht für das System ist. Um eine solche Bewertung durchführen zu können, muss dieses berechnen, inwiefern die Situation das System beeinflusst. Hierzu bringt es Eigenschaften der Situation mit Eigenschaften des Systems in Beziehung. Das hier auftauchende Subjekt ist also eigenschaftsbehaftet.

Eine Bewertung einer Situation kann beim Menschen implizit oder explizit erfolgen, während eine Maschine nur die explizite Repräsentation kennen kann. Implizit heißt, dass die Bewertung durch eine Emotion oder Gefühl stattfindet, während die explizite Repräsentation genau beziffert, wie die Eigenschaften der Situation die Eigenschaften des Systems beeinflussen. Als Beispiel soll der "Schmerz" dienen. Wird dieser vom Menschen implizit wahrgenommen, durch ein unangenehmes Gefühl, kann eine Maschine nur beziffern, inwieweit die Verletzung das System beeinträchtigt.

Der Mensch oder ein möglicher kognitiver Agent hat es also nicht mit einer Form von Subjekt zu tun, sondern mit Zweien: Ein eigenschaftsloser Kausator von Handlungen und ein eigenschaftsbehafteter Betroffener von einer Situation. Es ist Aufgabe des informationsverarbeitenden Systems hieraus ein stimmiges Gesamtbild seiner Selbst zu erzeugen (siehe Abschnitt 13).

Das Handlungsbewertungsschemata wird im Folgenden als eines von drei Konzepttypen behandelt, wobei dieses Schema den anderen Konzepttypen übergeordnet ist. Es wird angenommen, dass zwei Arten der Informationsverarbeitung stets parallel stattfinden. Auf unterer Ebene das Arbeiten mit Konzepten und auf der höheren Ebene eine ständige Bewertung und Einordnung dieses Arbeitens mit Konzepten. Je nachdem auf was die Aufmerksamkeitsfilter gerichtet sind, kann diese höhere Ebene selbst zum Gegenstand der Betrachtung werden (siehe Abbildung 2). In dem Fall werden mit Hilfe des Handlungsbewertungsschemata Konzepte vom eigenen Denken explizit repräsentiert. Dies kann eine explizit dargestellte Bewertung des eigenen Denkens sein, oder eine Kategorisierung der Denkvorgänge. Damit Denkvorgänge kategorisiert werden können, muss über ein mentales Handlungskonzept festgestellt werden, welche Art der Verarbeitung das System vorgenommen hat.

Durch Richtung der Aufmerksamkeit auf die höhere Kontrollebene der Kognition, kann somit die Metakognition selbst Gegenstand der Kognition werden. Dies geschieht, wie oben beschrieben, durch Bildung von Konzepten über das eigene Denken. Da diese des Handlungsbewertungsschemata entspringen, tauchen in diesen die zwei logischen Formen des Subjekts auf. Hierdurch wird das eigene Subjekt Gegenstand der Betrachtung und erzeugt

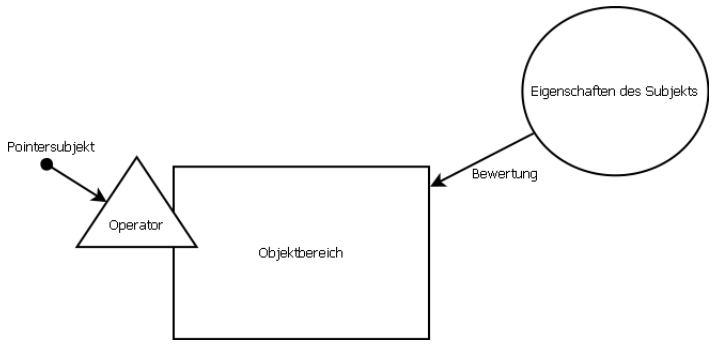

Abbildung 1: **Handlungsbewertungsschema:** Es gibt zwei Typen von Subjekt-Bezügen, die sich von der logischen Form her unterscheiden. Die Handlung wird durch einen Operator beschrieben, der auf den Objektbereich einwirkt. In diesem Konzept haben nur der Objektbereich und der Operator Eigenschaften, das Subjekt ist hier nur Bezugspunkt für die Handlung und wird Pointersubjekt genannt. Beim Bewertungskonzept werden Eigenschaften des Objektbereichs mit Eigenschaften des Subjekts verglichen, um zu prüfen, wie diese betroffen sind. Das Handlungsbewertungsschema wird sowohl zur Beschreibung einer physischen als auch mentalen Handlung herangezogen. Bei ersterer ist der Objektbereich die physische Außenwelt, bei letzterem ist der Objektbereich, das, was mental repräsentiert wird.

durch diesen Mechanismus Ich-Bewusstsein. Da es auch Metakognition von Metakognition geben kann, werden die gebildeten Konzepte über das eigene Denken wie jedes andere Konzept behandelt und es kann mit diesen weitergearbeitet werden. Auch hier ist die höhere Kontrollebene aktiv, so dass abermals die Aufmerksamkeit auf diese gerichtet werden kann und explizit wieder ein Ich-Bezug gebildet wird, der vom System selbst wieder als Konzept betrachtet wird. Wird die Aufmerksamkeit also entsprechend gelenkt, ist beliebig Metakognition von Metakognition möglich (siehe Abbildung 2).

Diese philosophischen einführenden Gedankengänge werden im Abschnitt 11.7 noch mal aufgegriffen, wenn das zugrunde liegende Modell erklärt wurde. Während das nächste Kapitel das Handlungsbewertungsschemata im Kontext einer aufgaben-zentrierten Verarbeitung betrachtet, wird es daraufhin konkret: Das Modell wird Stück für Stück aus direkt zugänglichen Teilen entwickelt, mit dem Ziel der algorithmischen Nachbildbarkeit.

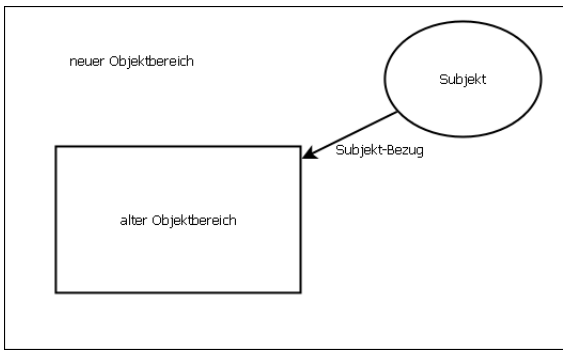

Abbildung 2: Denken über das Denken (Metakognition): Ein bestehender Objektbereich wird mit seinem Subjektbezug (samt Subjekt) zu einem Konzept, welches Teil eines neuen Objektbereichs wird, der der Kognition wieder zur Verfügung steht.

3 Handlung und Bewertung des kognitiven Systems und die aufgaben-zentrierte Verarbeitung

Was motiviert einen Menschen zu Handlungen? Was könnte eine Maschine zu Handlungen motivieren? Ist der Mensch quasi ständig mit Handlungen beschäftigt, so verwundert es vielleicht, dass es keine triviale Aufgabe ist, eine Maschine selbstständig zu Handlungen zu bewegen. Jede Handlung folgt aus einer Motivation heraus, doch per se hat eine Maschine keine Motivation überhaupt irgend etwas zu tun. Eine Motivation muss der Maschine erst mal vorgegeben werden.

Um herauszufinden, wie man einer Maschine eine Motivation verleihen kann, wird erst mal das Motivationssystem des Menschen untersucht. Prinzipiell wird jede Handlung des Menschen mit einer Bewertung versehen. Dies kann eine physische Handlung sein, oder ein Gedankengang, der hier als mentale Handlung verstanden wird. Darüber hinaus bewertet der Mensch jede entstandene Situation auch wenn diese nicht durch die eigene Handlung hervorgerufen ist. Aber wie ist der Mensch zu dieser Bewertung in der Lage?

Da der Mensch aus der Evolution heraus entstanden ist, sollte jede Motivation einen evolutionären Ursprung haben. Auch wenn es im einzelnen scheinbare Abweichungen davon gibt, soll dies für die Analyse angenommen werden. Es gibt ein Urziel des Menschen: Das Verbreiten und Überleben sei-

ner Gene. Das Problem was sich hieraus ergibt, ist, wie eine einzelne Handlung mit diesem Urziel in Verbindung gebracht wird. Angenommen man gäbe einer Maschine ebenfalls ein Urziel vor, so stellt sich hier das gleiche Problem.

Das Problem soll zunächst allgemein betrachtet werden. Auf der einen Seite stehen die möglichen Handlungen des Systems und auf der anderen Seite das Urziel. Diese beiden Seiten müssen sich annähern. Es wird angenommen, dass der Annäherungsprozess von beiden Seiten her stattfindet.

Um die möglichen Handlungen des Systems herauszufinden, muss ein anfänglicher spielerischer Lernprozess stattfinden. Dieser wird in Abschnitt 10 noch genauer betrachtet. Der spielerische Lernprozess gliedert sich dabei in zwei Phasen, die auch der Mensch durchlebt. In der ersten Phase interagiert das System nahezu wahllos mit der Umwelt, um das Verhalten der Umwelt und die eigene Einflussnahme herauszufinden. Z.B. schmeißt es wahllos Dinge auf den Boden oder versucht herauszufinden, ob Dinge lose oder fest zusammenhängen. In der zweiten Phase, wenn schon einige Interaktionskonzepte gelernt wurden, spielt das System anders: Es versucht Dinge aufzubauen. In dieser Lernphase wird spielerisch studiert, wie Handlungsergebnisse aufeinander aufbauen, ohne ein fürs Urziel relevantes Ergebnis zu erzielen.

Auf der anderen Seite muss vom Urziel aus eine Kaskade von Zwischenzielen aufgebaut werden. Dies hat die Evolution teilweise für den Menschen schon erledigt. Beispielsweise wird das Knüpfen von Freundschaft oder das Zusichnehmen von Nahrung vom Menschen positiv bewertet, ohne dass dem Menschen bewusst ist, dass dies dem Urziel dient. Hierbei muss, wie schon kurz erwähnt, zwischen impliziter und expliziter Bewertung unterschieden werden. Jede Emotion oder Gefühl ist eine Form von Bewertung, die hier als implizit bezeichnet wird. Sie kodiert für das System nicht nur, ob die Situation positiv oder negativ für das System ist, sondern auch in welcher Weise. Z.B. kodiert Angst, neben der negativen Bewertung, dass das System in Gefahr ist und erzeugt, die Bereitschaft von bestimmten Handlungen. Eine explizite Bewertung ist hingegen eine Repräsentation davon, wie das System durch die Situation beeinflusst ist. Wie also die Eigenschaften der Situation die Eigenschaften des Systems beeinflussen.

Der Begriff "Aufgabe" soll hier so definiert werden, dass ein (Zwischen-) Ziel vorliegt, und das System nach möglichen Handlungsketten sucht, um das Ziel zu erreichen. Ein kognitives System, ob Mensch oder künstlich, sei so aufgebaut, dass stets eine Aufgabe vorliegt, die vom System als zurzeit am relevantesten eingestuft wird, die es zu erledigen gibt. Hierbei ist zu beachten, dass sowohl das Lernen unter diese Definition fällt, als auch sämtliche

Freizeitbeschäftigung des Menschen. Selbst wenn der Mensch nur ausruhen will, was eine Schonung seiner Ressourcen darstellt, so stellt er sich z.B. die Aufgabe den Fernseher anzustellen. Auf die Art kann bei einem kognitiven System von einer aufgaben-zentrierten Verarbeitung die Rede sein.

4 Die Verschachtlung dreier Konzepttypen und das "klassische" maschinelle Lernen

Der Teil des hier vorgestellten Modells der Kognition, der sich mit dem Verarbeiten der Inputs und Outputs der physischen Welt beschäftigt, ist im wesentlichen eine Erweiterung des so genannten maschinellem Lernens. Beim maschinellem Lernen geht man gewöhnlich von einem sensorischen Input aus, der auf eine von verschiedenen vorgegebenen Klassen gemappt werden soll. Ein Problem des maschinellem Lernen ist also ein Klassifizierungsproblem. Hierbei hat man ein Trainingsset von Inputs und zugehörigen Klassen und wendet den Klassifikator nachher auf ungesehene Inputs an. Je nachdem wie gut der Algorithmus gelernt hat, kann dieser auch die ungesehenen Inputs den richtigen Klassen zuordnen.

Einen Fortschritt dieser Klassifizierer verdanken diese einer jüngst gemachten Entdeckung, dass es von Vorteil ist, ohne einen entsprechenden Output vorzugeben, erst mal stufenweise Features aus den Inputs zu extrahieren. Dies ist unter dem Namen "Deep Learning" bekannt geworden (siehe Abbildung 3 und 4). Features sind dabei statistische Häufungen von Mustern, die immer wieder auftreten, wie Kanten bei der visuellen Verarbeitung. Die Stufen ergeben sich dadurch, dass zunächst einfache Features, wie die eben genannten Kanten herausfiltert werden, um diese Kanten dann zu "höheren" Features zusammenzubauen, wie z.B. Rechtecken oder anderen Formen. Nach dieser Featureextraktion gestaltet sich das Klassifizierungsproblem relativ einfach, dass es meist ausreicht eine Klasse als gewichtete Summe dieser höheren Features darzustellen. Dabei hat z.B. ein Rechteck und eine Anzahl von Kreisen ein hohes Gewicht, wenn ein Auto erkannt werden soll, und Formen die nicht für ein Auto sprechen ein entsprechendes kleines oder sogar negatives Gewicht.

Alle dem Autor bekannten Verfahren des maschinellem Lernen haben dabei gemeinsam, dass stets ein einziges Klassifizierungsproblem für sich betrachtet wird. In dem hier vorgestellten Modell wird ein dreifaches Klassifi-

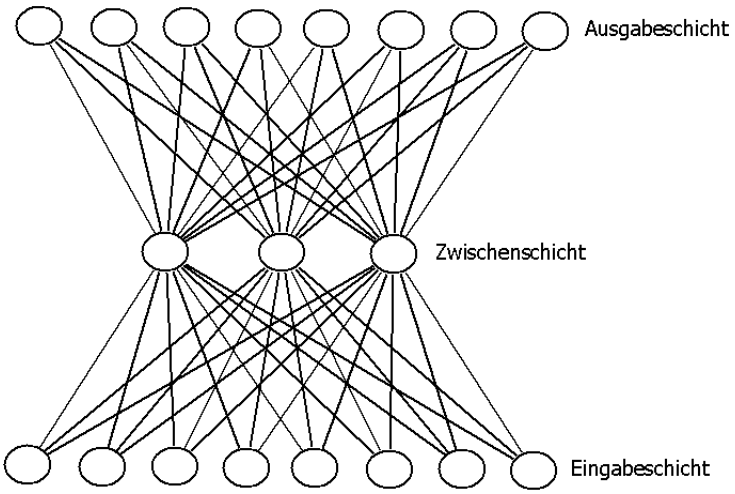

Abbildung 3: Die Abbildung zeigt schematisch einen Autodecoder. Die Eingabeschicht könnte z.B. ein Bild repräsentieren mit seinen einzelnen Pixelwerten. Die Aufgabe des Autodecoders ist es, die Eingabeschicht möglichst genau zu reproduzieren, d.h. dass an der Ausgabeschicht ein Bild entsteht, welches der Eingabeschicht möglichst ähnlich ist. Hierbei wird die Information der Eingabeschicht durch die Informationseinheiten der Zwischenschicht kodiert. Diese enthält weniger Informationseinheiten als die Eingabeschicht, so dass der Autodecoder gezwungen ist die Information zu komprimieren. Ein natürliches Bild lässt eine solche Komprimierung zu. Analysiert man die Kodierung der Zwischenschicht, so stellt man fest, dass diese kanten-ähnliche Features repräsentieren.

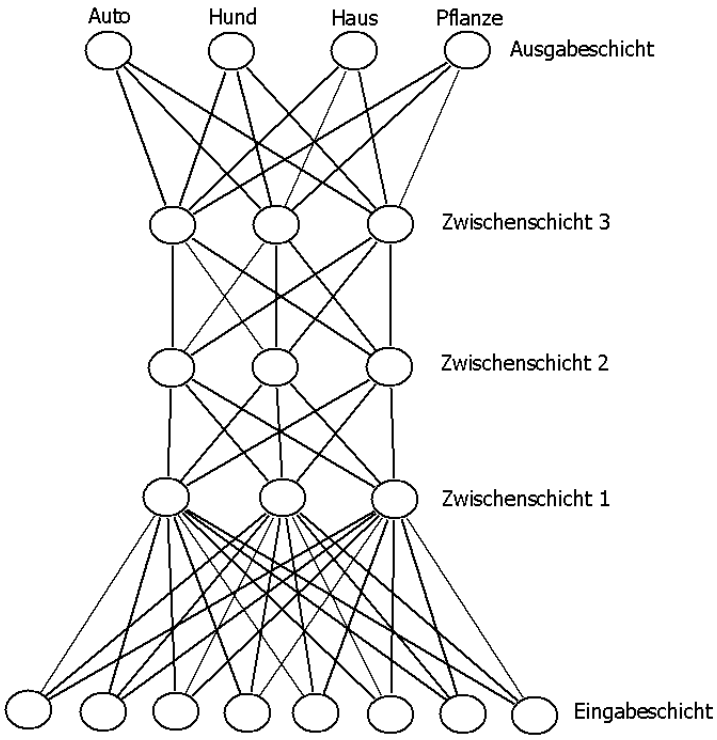

Abbildung 4: Beim Deep Learning werden mehrere Zwischenschichten aufeinander gesetzt, die immer abstraktere Features repräsentieren sollen. Jede Zwischenschicht enthält dabei weniger Informationseinheiten, als die vorangehende Zwischenschicht (in der Abbildung nicht gezeigt). Dadurch muss die Information von Schicht zu Schicht immer weiter komprimiert werden. Die Zwischenschichten werden einzeln trainiert, z.B. mithilfe eines Autodecoders (siehe Abbildung 3). Die letzte Zwischenschicht repräsentiert auf diese Weise sehr abstrakte Features, so dass durch eine Linearkombination dieser Features eine Einordnung vorgenommen werden kann, ob der gezeigte Input in der Eingabeschicht eines der vier vordefinierten Klassen "Auto", "Hund", "Haus" oder "Pflanze" zeigt.

zierungsproblem behandelt, wobei die Klassen ineinander verschachtelt sind. Die Verschachtlung löst sich dabei von Innen oder Außen auf, je nachdem was betrachtet wird. Bevor darauf eingegangen wird, wie rum sich die Verschachtlung auflöst, werden die Klassifizierungsprobleme kurz vorgestellt.

Im Gegensatz zum maschinellem Lernen werden die zu lernenden Klassen nicht Klassen sondern Konzepte genannt. Der Begriff Konzept taucht z.B. beim Lernen von sensomotorischen Konzepten auf[12]. Der Idee eines Konzepts nach, sind die verschiedenen Konzepte, die gelernt werden können nicht vorgegeben. Trifft der Agent auf ein Muster, so subsumiert er dieses unter ein bereits bekanntes Konzept oder erzeugt ein Neues, wenn das Muster von allen bekannten Konzepten zu sehr abweicht. Neben dem Subsumieren und der Neubildung von Konzepten gibt es eine Reihe von Konzeptoperationen, wie z.B. der Bildung eines Oberkonzepts aus zwei oder mehr Konzepten oder der Differenzierung eines Konzepts in zwei Neue. Die Idee des Konzepts ist nicht neu, was neu ist, dass die gesamte Kognition auf der Idee und den Möglichkeiten des Konzeptbegriffs aufgebaut wird. Selbst später, wo das Denken über das Denken beschrieben werden soll, geschieht dies mithilfe des Konzeptbegriffes. In diesem Abschnitt, wo das Denken über die physische Welt beschrieben wird, hat der Konzeptbegriff den Vorteil, dass z.B. das Bilden von Objekthierarchien zwanglos durch das Bilden von Oberkonzepten beschrieben werden kann. Bei der Beschreibung des Denkens über das Denken hat der Konzeptbegriff den Vorteil, dass alle Analysetools, die auf Konzepte angewendet werden können, ebenfalls bei der Metakognition Anwendung finden.

Zunächst werden die drei Konzepttypen so allgemein definiert, dass diese sowohl auf das Denken über die physische Welt als auch das Denken über das Denken zutreffen. Die Konzepttypen sind von Innen nach Außen:

1. Das Konzept einer in sich geschlossenen Sinneinheit.
2. Das Konzept der Relation zwischen zwei geschlossenen Sinneinheiten.
3. Das Handlungs-Bewertungskonzept.

Auf das Denken über die physische Welt bezogen, sind die drei Konzepttypen:

1. Das Konzept eines physischen Objekts.
2. Das Konzept der Relation (oder Interaktion) zwischen zwei physischen Objekten.

3. Das physische Handlungs-Bewertungskonzept.

Die Schachtelung ergibt sich wie folgt: Bei der Bewältigung einer Aufgabe (siehe Abschnitt 3) müssen Handlungen initiiert werden, um ein vom System als positiv bewertetes Ergebnis zu erzielen. Hierbei müssen die initiierten Interaktionen mit der physischen Welt die Interaktionen zwischen den physischen Objekten berücksichtigen. Um die Interaktion zwischen den physischen Objekten richtig zu benutzen, muss man ein Konzept der beteiligten physischen Objekte haben. Dies ist die Auflösung der Schachtelung von Außen.

Beim Lernen wird die Schachtelung von Innen aufgelöst. Zunächst werden Konzepte von physischen Objekten gelernt. Erst die Zerlegung der physischen Welt in Objekte lässt es zu, Relationen oder Interaktionen zwischen diesen zu beschreiben. Erst wenn die Interaktionen zwischen physischen Objekten verstanden sind, kann das kognitive System verstehen lernen, welche Handlungen zu welchen Ergebnissen führen.

Bevor es weiter geht, möchte der Autor eine Verwunderung ausdrücken. Warum gibt es in der heutigen Literatur zur künstlichen Intelligenz keine Veröffentlichungen, die sich mit dem Bilden von Konzepten über die Interaktion von physischen Objekten beschäftigen? Die Antwort liegt auf der Hand, wenn man berücksichtigt, dass sich die Forschung erst mal auf einzelne Klassifizierungsprobleme stürzt. Wolle man solche Konzepte bilden, müsste man als Vorbedingung erst mal durch eine Klassifizierung die Welt in physische Objekte aufteilen. Erst nach diesem Schritt könnte man die Interaktion zwischen solchen Objekten klassifizieren.

Einer These dieses Buches nach, können Relationen zwischen physischen Objekten nicht ohne Vorwissen aus einer statischen Szene abgeleitet werden. Es ist z.B. unmöglich in einem Bild, wo ein Glas auf dem Tisch steht, zu sagen, welchen Sinn diese Relation hat. Statische Relationen können nur aus dem Wissen über dynamische Relationen abgeleitet werden. Man muss in dem Beispiel wissen, dass ein Glas zu Boden fällt, wenn es nicht von einem Tisch unterstützt wird. Es läge demnach nahe, nach Interaktionskonzepten von physischen Objekten in der Literatur zu raumzeitlichen Features zu suchen. Aber auch hier werden einzelne nicht geschachtelte Klassifizierungsprobleme studiert. Z.B. wurden in der "CHOROCHRONOS"-Datenbank[8] dynamische Szenen zusammengetragen, die jeweils in vorgegebene Klassen unterteilt sind. Eine Klasse könnte das Winken einer Person sein. Die Algorithmen, die diese Datenbank zum Trainieren benutzen, zerlegen die Welt nicht zuerst in Objekte, sondern finden raumzeitliche Muster (Features), die

für das Winken und andere dynamische Aktionen charakteristisch sind, und klassifizieren die Inputs danach.

5 Phase I, die Strukturierung des Denkens über die Welt in drei Konzepttypen

Dieses Kapitel, welches die erste Phase des Aufbaus der Kognition beschreibt, gehört streng genommen noch nicht zur Kognition. Es werden keine kognitiven Vorgänge beschrieben, sondern die Tatsache, dass der Mensch nicht anders kann, als die physische Welt und die Interaktion mit ihr, mithilfe von drei Konzepttypen zu beschreiben. Dies ist zunächst eine philosophische Herangehensweise, die gleichzeitig die Frage aufwirft, ob es einem intelligenten Wesen überhaupt möglich ist, die physische Welt anders zu beschreiben. Auf diese Frage wird nochmal in einem kurzen philosophischen Abriss in Abschnitt 12 eingegangen. Auch die Mathematik soll kein Hilfsmittel sein über die drei Konzepttypen hinauszugehen, denn diese wird, der These dieses Buches nach, als Abstraktion von "natürlich" gelernten Konzepten beschrieben.

Zunächst soll die sogenannte sensomotorische Schleife betrachtet werden. Es ist in der künstlichen Intelligenz üblich, die Interaktion eines kognitiven Agenten mit der Welt als eine geschlossene Schleife zu behandeln. Der Agent nimmt hierbei die Welt mithilfe seiner Sensorik wahr, welches zu einer bestimmten motorischen Reaktion führt, welche die Welt verändert. Diese Veränderung wird wiederum von der Sensorik registriert und erzeugt neue motorische Reaktionen. Auf diese Weise erhält man eine geschlossene Schleife.

Die Interaktion mit der Welt soll in diesem Buch anders beschrieben werden. Die Idee der geschlossenen sensomotorischen Schleife beruht darauf, dass auf jede sensorische Wahrnehmung nach einem bestimmten Muster eine Reaktion berechnet werden kann. Hierbei verbirgt sich hinter der Halbschleife zwischen sensorischer Wahrnehmung und motorischer Reaktion die komplette Kognition des Agenten. Für einfach strukturierte Agenten ist diese Beschreibung zutreffend. Die Situation verändert sich aber, wenn der Agent seine Kognition selbst zum Gegenstand der Kognition machen kann, wenn dieser also zur Metakognition fähig ist. In diesem Fall kann der Agent nicht mehr durch ein einfaches Reiz-Reaktionsschema beschrieben werden. Kontrolliert der Agent durch Metakognition seine eigene Kognition, ist nicht sicherge-

stellt, dass überhaupt, oder wann, auf einen Reiz eine motorische Reaktion folgt.

Auch die andere Halbschleife zwischen Motorik und Wahrnehmung wird in diesem Buch anders behandelt. Es geht nicht darum durch die Motorik neue sensorische Reize zu schaffen, sondern durch motorische Interaktion Konzepte zu lernen, diese durch Kognition zu abstrahieren, und auf neue Situationen anzuwenden. Hierbei wird die Interaktion mit der Welt nicht durch die Sensomotorik eingeklammert.

Die Sensorik hat hier eine andere Aufgabe. Die Aufgabe der Sensorik besteht lediglich darin von der Außenwelt ein inneres Abbild zu schaffen. Hierbei werden keine Konzepte gelernt, wie die Sensorik dies zustande bringt. Es wird im weiteren angenommen, dass eine feste Verschaltung vorliegt, welche zu einer Übersetzung der Sensorinputs zu einem "korrekten" inneren Abbild der wahrgenommenen Welt führt. Korrekt soll in diesem Fall so verstanden werden, dass das innere Abbild die physischen Abstände zwischen den Objekten nicht eins zu eins wiedergeben muss, sondern lediglich korrekt mit dem inneren Abbild gearbeitet werden kann. Dies heißt, dass der kognitive Agent keine Ahnung davon hat, wie das innere Abbild geschaffen wird, sondern dieses als gegeben hinnimmt. Der Agent lernt z.B. nicht durch Interaktion mit der Welt, wie die zwei zweidimensionalen Bilder der Augen zu einem dreidimensionalen Bild verrechnet werden. Dies schließt nicht aus, dass ein Wissenschaftler herausfinden kann, wie diese Verrechnung vonstatten geht, dies ist aber ein anderer Blickwinkel. Der Wissenschaftler lernt dabei selbst Konzepte über einen naturwissenschaftlichen Gegenstand, wobei dieses Lernen des Wissenschaftlers wieder mit den hier beschriebenen Prinzipien behandelt werden kann.

Übersetzt die Sensorik die Außenwelt lediglich nach Innen, so wird die Interaktion mit der Außenwelt durch Handlung und Handlungsbewertung eingeklammert, indem physische Handlungs-Bewertungskonzepte gelernt werden. Die drei Konzepttypen, die gelernt werden können, sollen nochmal aufgelistet werden:

1. Das Konzept eines physischen Objekts.

2. Das Konzept der Relation (oder Interaktion) zwischen zwei physischen Objekten.

3. Das physische Handlungs-Bewertungskonzept.

Es wird nicht behauptet, dass der Mensch bei der Verarbeitung neuer Reize, dass Wahrgenommene sofort nach diesen drei Konzepten sortiert, sondern, dass diese Sortierung erst nach einigen analytischen Schritten vorliegt. Hierzu muss gezeigt werden, dass die drei Konzepttypen trennscharf und vollständig sind. Sie sind trennscharf, wenn es keine Überschneidungen zwischen den Konzepttypen gibt, so dass alles Wahrgenommene nur in eines der Konzepttypen passt. Vollständig sind die Konzepttypen, wenn sich ALLES, was über die physische Welt und die Interaktionen mit dieser gelernt werden kann, mithilfe dieser Konzepttypen beschreibbar ist. Lässt sich Gelerntes über die Welt nur mit einigem Aufwand in diese Konzepttypen zerlegen, so ist dies ein Zeichen dafür, dass es sich um fortgeschrittenes Wissen handelt. Fortgeschrittenes Wissen entsteht durch das Arbeiten mit diesen Konzepttypen, wobei Konzepte abstrahiert, logisch differenziert und vielfach verknüpft werden (siehe Abschnitt 7.3).

Der Vorteil mit diesen drei Konzepttypen zu arbeiten, besteht darin, dass Denken über die Welt in handhabbare Teile zu zerlegen. Diese Zerlegung soll es erst ermöglichen, Probleme, wie das Abstrahieren von Konzepten, einem Computer zugänglich zu machen. Ist es nur schwer möglich eine funktionierende Theorie der Abstraktion aufzustellen, so ist es um ein vielfaches leichter für jeden Konzepttyp ein eigenes Vorgehen zu entwickeln. Es wird später gezeigt, dass sich aus dem hier vorgestellten Modell zwanglos weitere logische Unterscheidungen ergeben, wie die einzelnen Konzepttypen abstrahiert werden können. Beim Konzept eines physischen Objekts gibt es drei logisch verschiedene Methoden diesem Eigenschaften zuzuordnen, die für eine Abstraktion verwendet werden können. Dies entspricht dem Ziel des Buches, die Kognition soweit zu zerlegen, dass ein Teilproblem letztendlich durch einen geschlossenen Ansatz gelöst werden kann, und gleichzeitig die Beschreibung der Kognition vollständig zu halten, d.h. keinen Teilaspekt der Kognition durch die Zerlegung auszuklammern.

5.1 Die Trennschärfe der drei Konzepttypen

In diesem Abschnitt wird die Trennschärfe der drei physischen Konzepte diskutiert. Auch wenn die drei Konzepte eine unterschiedliche logische Struktur haben, scheint es bei der Interaktion von Objekten und einer Handlung Überschneidungen zu geben. Logisch klar ist die Unterscheidung zwischen physischen Objekten und einer Interaktion von physischen Objekten. Bei ersterem geht es um eine in der Gesamtszene für sich geschlossenen Sinn-

einheit, wie ein Glas oder Tisch, welche später mit einem Symbol versehen werden kann. Die Beschreibung der Interaktion von Objekten setzt diese erste Heraustrennung von Objekten voraus. Die eigentliche Interaktion ist dann die Beschreibung des dynamischen Verhaltens, welche zwei Objekte miteinander eingehen. Hierbei reicht es stets zwei Objekte zu betrachten, die in besonderer Konstellation zueinander stehen.

Es soll hier schon erwähnt werden, dass man den einen Interaktionspartner später durch einen Quantor ersetzen kann. Der Begriff Quantor kommt aus der mathematischen Logik und beschreibt, ob eine Relation für JEDES Objekt zutrifft, oder ob EIN Objekt EXISTIERT, für das diese zutrifft. Quantoren werden explizit oder implizit ebenfalls im natürlichen Denken verwendet. Dies geschieht z.B. wenn ein Gegenstand sich bei JEDER Interaktion mit einem anderen Objekt leicht verformt. Dieses Objekt erhält später die Interaktionseigenschaft "weich".

Aufwendiger ist die Trennung von Objektinteraktionen und Handlungen. In jeder Handlung ist der Körper des kognitiven Systems involviert, der als Objekt mit anderen Objekten interagiert. Hierbei muss das System zwei Sichtweisen vom eigenen Körper voneinander trennen. Greift dieses z.B. ein Glas, betrachtet es dieses Greifen als Interaktion von zwei Objekten: Der Hand und dem Glas. Das Greifen des Glases und das Platzieren an einem anderem Ort wird aber gleichzeitig als Operation des Systems dargestellt. Im Unterschied zur Interaktion von Objekten gibt es bei der Operation oder Handlung des Systems ein Handlungsergebnis. Auch der Endzustand einer jeglichen Interaktion von Objekten kann als ein Ergebnis bewertet werden. Die entstandene Konstellation muss dafür vom System als bewertbares Ergebnis interpretiert werden. Später in Abschnitt 7.1 wird der Begriff "Kausalität" genauer betrachtet. Genaugenommen kennt die Physik den Begriff der Kausalität nicht per se. Die Physik kennt, wie das hier beschriebene Interaktionskonzept, nur Interaktionen von Teilchen. Erst wenn ein entstandener Zustand herausgenommen wird und diesem eine Bedeutung zugewiesen wird, kann analysiert werden, welche Faktoren kausal zu diesem Zustand geführt haben.

Im Gegensatz zur physikalischen Betrachtung der Situation, wo es nur Objektinteraktionen gibt, setzt sich das Handlungskonzept gerade so zusammen, dass es ein Kausator der Handlung gibt, der nicht weiter beschrieben wird. Wir wissen nur, wie man eine bestimmte Handlung ausführt, und haben keinen Zugriff auf die kausalen Zusammenhänge in unserem eigenem Gehirn, wie z.B. die Steuerung der Hand vonstatten geht. Dies führt dazu, dass

wir unsere eigenen Handlungen als Operationen auf einen Objektbereich beschreiben, mit einem Subjekt als Urheber, dem keine weiteren Eigenschaften zugeordnet werden.

So führt das Greifen des Glases auf zwei unterschiedliche Konzepttypen, einmal der Interaktion von Hand und Glas, wobei hier die gegenseitige Dynamik das Konzept ausmacht, und einmal auf ein Handlungskonzept, welches sich auf den entstandenen Zustand konzentriert, wobei sich das System hier als eigenschaftslosen Kausator beschreibt, der verschiedene Operationen in der Welt ausführen kann, die zu bestimmten Konstellationen führen.

Was später im Kapitel "Lernen" weiter ausgeführt wird, ist, dass das kognitive System zunächst lernen muss seine Motorsignale zu steuern. Auch hier soll dies mit Handlungskonzepten beschrieben werden. Der Agent lernt, wie in Abhängigkeit seiner Körperposition, die Motorsignale zu verschiedenen Effekten in der physischen Welt führen. Hierbei werden nicht die Motorsignale vom System explizit repräsentiert, sondern lediglich die physischen Effekte, die hervorgerufen werden können.

Das Handlungsergebnis muss nicht zwangsläufig eine vollkommen statische Konstellation beschreiben. Z.B. kann beim Bedienen eines Autos das Anfahren, Richtungsänderungen und das Anhalten als Handlungsergebnisse interpretiert werden. Allen Handlungsergebnissen gemein ist, dass sie Konstellationen beschreiben, auf denen weitere Handlungen aufbauen können. Hierzu muss die Konstellation eine Konstanz aufweisen. Diese Konstanz kann eine herbeigeführte statische Relation sein, wie das Abstellen eines Glases auf den Tisch, oder eine Konstanz in einem Vorgang, wie das Anfahren beim Auto. Hier ist die Konstanz, dass sich das Auto (Richtung Ziel) in Bewegung gesetzt hat.

Es soll noch erwähnt werden, dass das Handlungsbewertungskonzept in zwei Konzepte zerlegt werden kann. Eine Handlung ohne Bewertung macht zwar keinen Sinn, aber das kognitive System ist imstande jede Situation zu bewerten, was diese für das System bedeutet, ohne dass das System kausal auf diese eingewirkt hat.

5.2 Die Vollständigkeit der drei Konzepttypen

Die Vollständigkeit der drei Konzepttypen kann nicht (wie auch die Trennschärfe) mathematisch genau gezeigt werden. Man muss sich an Beispielen orientieren, die scheinbar nicht ins Schema passen. Das erste Beispiel ist die Relativität des Beobachters. Durch die Relativität des Beobachters soll ausge-

drückt werden, dass das kognitive System einen Körper hat, der als physisches Objekt verschiedene Positionen zu anderen Objekten einnehmen kann. Dieser Sachverhalt soll folgendermaßen durch die hier beschriebenen Konzepte erfasst werden: Eine selbst herbeigeführte Änderung der eigenen Position soll als Handlungskonzept beschrieben werden, wobei die neu entstandene Position zu Objekten das Handlungsergebnis ausmacht. Die räumlichen Relationen des eigenen Körpers zu anderen Objekten werden dabei durch statische Relationen beschrieben. Wie bereits erwähnt ergeben sich statische Relationen entweder als Handlungsergebnis, oder definieren sich dadurch, dass bestimmte dynamische Interaktionen in der vorliegenden Szene nicht eintreten.

Als zweites Beispiel soll die Wissenschaft der Physik betrachtet werden. Auch hier wird die Welt durch Objekte und deren Interaktionen beschrieben. Ein Gegenstand wird dadurch charakterisiert, dass dieser in Teilobjekte zerlegt wird, wobei die Interaktion der Teilobjekte die Eigenschaften des Gesamtgegenstandes ausmachen. Es ist freilich möglich sich eine fiktive alternative Physik auszudenken, die nicht auf der Interaktion von Objekten aufbaut, wie dies bei Zellulärautomaten[17] der Fall ist. Die Funktionsweise eines Zellulärautomaten spiegelt dabei, der hier vertretenen These nach, nicht unser Basisverständnis der physischen Welt wieder. Wir verstehen den Zellulärautomaten über den Umweg der Mathematik. Mathematik wird verstanden, indem natürlich gewachsene Konzepte abstrahiert werden (siehe Abschnitt 12). Damit ist das Konzept eines Zellulärautomaten kein Basiskonzept, sondern ein durch Abstraktion Gewonnenes.

Es soll weiter eine Fähigkeit des Menschen beschrieben werden, die hier De-Personalisierung des Handlungsbewertungskonzepts genannt werden soll. Beschreibt der Mensch z.B. technische Gegenstände, so beschreibt er diese nicht ausschließlich durch Objektinteraktion, sondern sagt z.B., dass ein Mikrochip auf den Speicher zugreift, was ein Handlungskonzept darstellt. Später in Kapitel 7.3 wird gezeigt, dass ein kognitives System Basiskonzepte aus zwei Gründen zu komplexeren zusammenfügt: Der erste Grund ist das Erstellen einer Lösungsstrategie für eine Aufgabe, der zweite ist das Beschreiben eines Sachverhaltes. So kann das kognitive System zur Beschreibung eines Sachverhaltes, wie der Beschreibung eines Mikrochips, de-personalisierte Handlungskonzepte heranführen.

6 Die subsymbolische Simulationsebene

Bevor das Arbeiten mit den Konzepttypen besprochen wird, soll etwas über den Aufbau des gesamten hier beschriebenen Modells gesagt werden. Die hier beschriebenen Konzepttypen, die physischen wie auch die später eingeführten mentalen, sind leere Schablonen. Die leere Schablone an sich hat keinen Inhalt sondern nur eine logische Struktur. Die logische Struktur legt dabei fest, welche Operationen wie mit der Schablone ausgeführt werden können. Die eigentliche Operation wird auf dem Inhalt der Schablone ausgeführt. Die Schablonen füllen sich dabei alleine durch Interaktion mit der physischen Außenwelt. Hat der Agent noch keine Interaktion mit der Außenwelt erfahren, wie dies beim Start des Agenten der Fall ist, soll es ihm nicht möglich sein, irgendeine Art der Kognition durchzuführen und etwa irgendwas über seinen eigenen Aufbau zu erfahren. Jede Operation mit Konzepten arbeitet auf dem Inhalt der jeweiligen Schablone, so dass beim Start keine Operationen möglich sind.

Dies heißt, dass z.B. beim physischen Objekt Glas, wenn dieses als Objekt erkannt wurde, nicht mit einem Symbol für dieses weitergearbeitet wird, sondern die Objektschablone mit dem sensorischen Input gefüllt wird, welches über das Glas gewonnen wurde. Nicht nur wird das Objekt Glas aus der Szenerie herausgetrennt betrachtet, es erhält durch die weitere Verarbeitung zusätzliche Eigenschaften, z.B. wie es mit anderen Objekten interagiert, oder wie es in einem Handlungsbewertungskonzept benutzt wird.

Um weiterzuarbeiten soll der Begriff "Szene" definiert werden. Es spielt dabei keine Rolle, ob die Szene eine aktuell beobachtete Außenwelt wieder gibt, oder ob eine Außenwelt durch eine innerlich simulierte Szene wiedergegeben wird. Die Szene soll eine Außenwelt beschreiben, wobei ihr Hauptmerkmal ist, dass sie diese nicht physikalisch korrekt wieder gibt. Beobachtet ein kognitives System die Außenwelt, so konstruiert es eine Szene, die aus Konzepten besteht. Es werden erstens gleichzeitig alle physischen Objekte erkannt, zweitens alle Objekt-Objekt-Interaktionskonzepte betrachtet, die mit diesen Objekten möglich sind, bzw. die statischen Relationen, die genau durch nicht Eintreten dieser dynamischen Interaktionen definiert sind und drittens die Handlungsmöglichkeiten, die die Situation bietet (oder genau nicht bietet). Eine innerlich simulierte Szene, wie eine aktuell beobachtete Szene, ist eine Rekonstruktion der Außenwelt, die alleine aus Konzepten über diese besteht. Führt das kognitive System also Handlungen in der Außenwelt aus, so ist dieses teilweise blind für die wirkliche Welt, da sich dieses alleine

an gelernten Konzepten orientiert. Nimmt das kognitive System z.B. einen Teller aus dem Schrank und stellt diesen auf dem Tisch ab, so berechnet es keine exakten Trajektorien. Stattdessen führt es Konzepte aus, wie das Öffnen des Schranks, das Greifen des Tellers, und das Ablegen des Tellers auf dem Tisch. Was bei dieser Handlungskette nicht in Konzepte gefasst ist, nimmt das System nicht wahr. Die Sensorik überwacht dabei die korrekte Ausführung des Konzepts, so dass nachjustiert werden kann, wenn z.B. der Teller nicht richtig gegriffen wurde.

Es soll was zu dem benötigten Rechenaufwand bei einer solchen Rekonstruktion einer Szene gesagt werden. Prinzipiell gilt, dass bei der Rekonstruktion einer Szene sowie bei allen später beschriebenen Konzeptoperationen, wie z.B. der Analogiebildung, der Rechenaufwand proportional zur Anzahl der gelernten Konzepte ist. Es ist eine erstaunliche Eigenschaft des menschlichen Gehirns, dass, gegeben ein bestimmtes Konzept, dieses parallel prüfen kann, welche überhaupt gelernten Konzepte mit diesem verknüpft werden können. Es wäre ein riesen Fortschritt herauszufinden, wie diese parallele Prüfung ALLER gelernten Konzepte im menschlichen Gehirn vonstatten geht. Solange dies im Dunkeln liegt, muss sich ein künstlicher kognitiver Agent auf die enorme serielle Rechenleistung moderner Computerchips verlassen und seinen gesamten Speicher Konzept für Konzept prüfen.

Klassische Sichtweise des Symbolischen vs. Subsymbolischen Studiert man die Literatur der künstlichen Intelligenz, so zerfällt der Bereich grob in zwei Lager: Die symbolische und subsymbolische Verarbeitung. Die symbolische Verarbeitung arbeitet mit der mathematischen Logik oder Erweiterungen dieser (siehe das bis heute verfolgte Projekt "CYC"[9]). Der Ansatz ist eine reine Symbolmanipulation, ohne dass das System weiß, was diesen Symbolen zugeordnet ist. Es wird mit logischen Verknüpfungen gearbeitet wie "Feuer ist heiß" oder "Wasser ist nass" ohne ein Verständnis von Feuer oder Wasser.

Auf der anderen Seite gibt es die subsymbolische Verarbeitung auch Konnektionismus genannt, da diese häufig mit Netzwerken wie künstlichen Neuronalen Netzen arbeitet. Diese werden mit sensorischen Daten gefüttert und müssen, wie bei dem Deep Learning erwähnt, Klassifizierungen auf den Daten ausführen. Einem einzelnen Knoten in einem solchen Netz kann nicht wie beim Symbolischen eindeutig etwas zugeordnet werden. Die Leistung des Netzes entsteht durch seine Gesamtheit.

Es wäre naheliegend folgende Kombination beider Ansätze zu verwenden: Auf unterer Ebene würde man Netze benutzen, um Objekte zu erkennen, diesen dann Symbolen zuordnen, um auf höherer Ebene nur noch mit diesen Symbolen zu arbeiten.

Der hier beschriebene Ansatz funktioniert anders. Verwandt mit der Logik ist hier, dass Schablonen verwendet werden, die eine logische Struktur haben. Mit dieser logischen Struktur kann allerdings nicht direkt gearbeitet werden, sie gibt nur vor, welche Konzeptoperationen möglich sind und wie die jeweilige Vorgehensweise ist. Die eigentliche Verarbeitung von beispielsweise zwei Konzepten findet auf dem Inhalt statt. Im Gegensatz zur oben vorgeschlagenen Kombination von beiden Ansätzen, verlässt das hier vorgeschlagene Modell nie die subsymbolische Ebene. Die in der Verarbeitung innewohnende Logik schreibt lediglich die Art der Verarbeitung vor. Dies macht nicht zuletzt einen Unterschied beim Denken über das Denken. Herkömmliche Metakognitionsansätze arbeiten ausschließlich mit der mathematischen Logik, also einem reinen Symbolansatz. Ihnen ist es nicht gelungen ein Modell des Ich-Bewusstseins aufzustellen, was in diesem Buch versucht wird.

Gödelsche Unvollständigkeit Die Gödelsche Unvollständigkeit ist ein Begriff aus der mathematischen Logik, der hier auf das Problem der symbolischen und subsymbolischen Verarbeitung übertragen werden soll. In der Logik unterscheidet man zwischen einem Aussagensystem, Axiomensystem genannt, und dem, worüber die Aussagen sprechen, ihre Semantik. Die Semantik wird in der Mathematik als Modell beschrieben. Z.B. hat man auf der Aussageseite die Axiome der Natürlichen Zahlen und als Modell die Natürlichen Zahlen selbst. Der Gödelsche Unvollständigkeitssatz[14] besagt nun, dass egal wie viele Aussagen man über die Natürlichen Zahlen hinzunimmt, man kann ihre Eigenschaften nie vollständig beschreiben.

Die mathematische Logik ist identisch mit der symbolischen Verarbeitung in der künstlichen Intelligenz. Wenn man annimmt, dass das worüber die Aussagen sprechen die physische Außenwelt ist, so kann man analog behaupten, dass ein logisches System die Außenwelt nie in allen Einzelheiten erfassen kann, egal wie kleinkörnig die logischen Aussagen sind.

Dem Problem entgeht man, wenn man nur auf der Semantik arbeitet, d.h., dass die Verarbeitung mit Simulationen der physischen Welt arbeitet. Ein Beispiel soll dies verdeutlichen: Einem symbolverarbeitenden System wird gesagt, dass jemand Kaffee in eine Tasse gießt und dabei etwas verschüttet.

Ein System, welches mit gelernten Konzepten über die Welt arbeitet, würde nachfragen, ob sich die Person die Hand verbrüht hat. Dies geht aus der Logik des Satzes nicht hervor, gehört aber zur Konstruktion der Szene, die alle möglichen Konzepte mit betrachtet.

Auch wenn die Konstruktion der Szene, wie oben beschrieben, nicht mit der wahren physischen Außenwelt übereinstimmt, so ist diese dahingehend vollständig, dass sie alles erfasst, was für das kognitive System wichtig sein könnte.

7 Phase II, das Arbeiten mit den drei physischen Konzepttypen

Die Beschreibung der Operationen, die mit Konzepten ausgeführt werden können, ist der umfangreichste Abschnitt. Es sollen nicht einfach nur alle möglichen Konzeptoperationen vollständig aufgelistet werden, sondern auch, für jeden Konzepttyp spezifisch, zumindest prinzipiell erläutert werden, was die Merkmale der jeweiligen Verarbeitung sind. Technisch kann jede Konzeptoperation mit einer überladenen Funktion realisiert werden, wobei die logische Form des jeweiligen Konzepts die Funktionsschritte festlegt. Es wären zunächst drei Grunddatentypen zu definieren für die drei physischen Konzepte. Später kommen analog drei Datentypen für die mentalen Konzepte hinzu. Bereits hier oder später kann auch ein Datentyp für die "Szene" angelegt werden, so wie sie oben definiert wurde. Durch logische Unterscheidungen kommen weitere Datentypen hinzu, z.B. kann eine Eigenschaft von einem physischen Objekt drei verschiedene Ursprünge haben. Auch für zusammengesetzte oder abgeleitete Konzepte, wie die statische Relation, ist es sinnvoll später Datentypen zu definieren. Das verwenden verschiedener Datentypen, auf die überladene Funktionen verschieden reagieren, spiegelt das Schablonenprinzip wieder, dass nur mit dem Inhalt einer Schablone gearbeitet wird, die eigentliche Schablone ist ein leerer Datentyp.

Weiter sind alle Konzeptoperationen so ausgelegt, dass die entsprechenden Funktion letztendlich geschlossene Lösungen repräsentieren. Dies ist noch nicht bei allen Funktionen der Fall. Der Autor ist sich aber sicher, dass das hier vorgestellte Modell im Abstrakten vollständig ist, und nahtlos ineinander greift, dass sich so die verbleibende Arbeit im Detail abspielt. Hierbei ist nicht ausgeschlossen, dass weitere grundsätzliche logische Unterscheidun-

gen gemacht werden müssen, so dass die einzelnen Operationen durch einen einfacheren Ansatz gelöst werden können.

Für einige Konzeptoperation mit spezifischen Konzepttyp existieren bereits Lösungen, da an ihnen seit langem geforscht wird, z.B. das Erkennen eines physischen Objekts. Andere Konzeptoperationen enthalten völlig neue Ideen. Z.B. ist dem Autor bis dato keine Lösung bekannt, wie die Vorgehensweise bei der Generierung eines Interaktionskonzepts ist (siehe Abschnitt 7.1.2).

7.1 Die Konzeptgenerierung

Zunächst wird die Konzeptgenerierung beschrieben. Das Modell ist so aufgebaut, dass das kognitive System Information nur in Form von Konzepten aufnehmen kann. D.h., dass das System nicht imstande ist mit Information umzugehen, welche nicht nach einem Konzept strukturiert ist. Dies sei bei dem Menschen ebenso der Fall. Wobei die drei vorgestellten Konzepte über die physische Welt die Grundkonzepte bilden aus denen komplexere aufgebaut werden können.

Es werden im Laufe des Buches verschiedene Begriffe, die im Alltag verwendet werden, an der jeweiligen Stelle, wo diese gebraucht werden, zum Zwecke des Modells genau definiert. Z.B. ist der Sensoreindruck "Rot" für sich keine Eigenschaft. Eigenschaften werden Objekten zugeordnet. D.h., dass erst "rote" oder "runde" Objekte erkannt werden müssen, um ihnen dann diese gemeinsame Eigenschaft zuordnen zu können.

Bei der Konzeptgenerierung, und nur hier, wird der Begriff "Feature" gebraucht. Ein "Feature" sei eine wiederholt auftretende Konstellationen in der Wahrnehmung. Auf die physische Außenwelt bezogen, sind dies Konstellationen von einzelnen Sensorpunkten. Ein Feature ist somit *unter* dem Konzept angesiedelt, genaugenommen bilden wiederholt auftretende Konstellationen die Grundlage für die Bildung eines Konzepts. Z.B. sind die wiederholt auftretenden Kanten in natürlichen Bildern Features, die zu Objekten zusammengebunden werden können. Hierbei stellt sich zunächst die Frage, wie und warum bestimmte Features zusammengebunden werden und andere nicht. Hierbei soll die aus der Psychologie bekannte Objektpermanenz[1] eine zentrale Rolle spielen. Geht es bei der Objekterkennung gerade darum, dass die Objektpermanenz erhalten bleibt, ist die gebrochene Objektpermanenz die Grundlage für das Aufstellen von Objektinteraktionen.

Man kann sich die philosophische Frage stellen, ob wiederholt auftreten-

de Konstellationen in der Wahrnehmung genau dann zu einem Feature werden, wenn die Konstellationen nicht zufällig sind. Sind die Kanten im Bild nicht zufällige Linien, sondern markieren gerade Objektgrenzen, so kann man behaupten, dass ein Feature immer eine dahintersteckende Bedeutung hat. Umgekehrt wäre es auch nicht sinnvoll Konzepte zu bilden aufgrund von Konstellationen, die wirklich zufällig sind und keine Bedeutung haben.

Beim Handlungsbewertungskonzept hat man es ebenfalls mit wiederholt auftretenden Konstellationen zu tun. Diese werden dort aber nicht Features genannt, sondern sind die Grundlage einer Blockbildung. Eine Handlung wird durch Blöcke in Teilhandlungen zerlegt, die im Laufe des Lernprozesses unterschiedliche Bewertungen erhalten können, je nachdem welche Blöcke an welchen Handlungen beteiligt waren.

Neben der Eigenschaft, dass Konzepte aus wiederholt auftretenden Konstellationen gebildet werden, treten zwei Vorgehensweisen bei allen drei Konzepten auf: Erstens muss unnötige Information abgestreift werden und zweitens muss zum Teil nicht sichtbare Information ergänzt werden, was Autovervollständigung genannt werden soll.

7.1.1 Die Generierung des Konzepts eines physischen Objekts

Über diesen Punkt wurde in der Literatur zur künstlichen Intelligenz wohl am meisten geschrieben. Der Konzeptidee folgend, sollen die möglichen Objektklassen nicht von vornherein feststehen. Üblich ist, dass ein Trainingsset verwendet wird, von Bildern als Input mit Objektklassen als vorgegebenen Output. Ist das Trainingsset gelernt, wird geprüft, ob das System Objekte auf unbekannten Bildern richtig in diese Objektklassen einordnen kann.

Die Konzeptidee ist, dass es kein Trainingsset gibt. Egal wie viele Objektklassen schon gebildet wurden, ein aktuell neuer Eindruck wird mit den bekannten Konzepten verglichen und entweder unter ein Konzept subsumiert oder, wenn der Unterschied, für das es ein Maß geben muss, zu groß ist, ein neues Konzept gebildet.

Bei der Objekterkennung reicht eine Schicht von Features meist nicht aus, um ganze Objekte zu klassifizieren. Beim Deep Learning (siehe Abbildung 4) werden mehrere Featureschichten übereinander gelegt, wobei auf unteren Schichten "einfache" Features erkannt werden, die zu immer komplexeren Features zusammengefügt werden. Die komplexeren Features repräsentieren dabei Formen, wie näherungsweise Kreisen oder Rechtecken. Auch wenn bis dato nicht im einzelnen klar ist, was jede Schicht genau macht, gibt

es ein recht aufschlussreiches Paper[13] zu dem Problem, was die einzelnen Schichten kodieren. Die Symmetrie der erkannten Features spielen dabei eine zentrale Rolle. Das Paper konnte zeigen, dass die Symmetrie einer Kante die einfachste Symmetrie ist, und das z.B. Autodecoder (siehe Abbildung 3) diese mit höchster Wahrscheinlichkeit zuerst finden. Auf höheren Ebenen spielt die Symmetrie weiterhin eine Rolle, da ein Deep Learning Netzwerk aus identischen Schichten besteht. Repräsentiert auf einer höheren Schicht ein einzelnes künstliches Neuron eine ganze Kante, werden die höheren Formen ebenso nach dem Symmetrieprinzip als erstes erkannt. Z.B. weist ein Kreis oder Rechteck, welche aus solchen Kanten gebildet werden, eine hohe Symmetrie auf.

Spielt Symmetrie beim Zusammenbinden von Features eine zentrale Rolle, so erklärt diese das Problem nicht vollständig, da nicht alle Objekte in der Natur eine hohe Symmetrie aufweisen. Es sollen drei weitere Punkte genannt werden, die beim richtigen Zusammenbinden von Features hilfreich sind. Als erstes soll die Objektpermanenz betrachtet werden. Diese bildet kein Hilfsmittel bei der Betrachtung eines statischen Bildes. Bewegt sich hingehen der Gegenstand, so kann dieser vom Hintergrund getrennt werden, und die Features, die den Gegenstand ausmachen zusammengebunden werden. Diese Figur-Hintergrund Trennung ist beim Erkennen eines Gegenstandes das Abstreifen unnötiger Information.

Als zweites kann das bereits gewonnene Wissen über Objekte benutzt werden, um Features richtig zusammenzubinden. Objekten ist gemein, dass diese ähnliche Formen oder Teilformen besitzen. Eine Form ist dabei noch kein Konzept von einem Gegenstand, das Formwissen, was nicht explizit repräsentiert wird, dient aber dazu die Features von unbekannten Objekten richtig zusammenzubinden.

Der dritte Punkt ist eine logische Überlegung. Ein Feature, wie eine Kante oder Ecke, kann nur einem Objekt zugeordnet werden. Hierbei gilt, wenn Feature A und Feature B zusammengehören und Feature B zu Feature C, so gehört auch Feature A zum selben Objekt wie Feature C. D.h., dass über die Features ein überschneidungsfreier transitiver Graph gelegt werden kann (siehe Abbildung 5).

Ein weiterer Punkt bei der Generierung des Konzepts eines physischen Objekts ist die Autovervollständigung. In diesem Buch wird die Leistung des sensorischen Apparats ein inneres Abbild der physischen Außenwelt zu schaffen kaum Beachtung geschenkt, da diese Leistung vom System nicht in Form von Konzepten abgespeichert wird. Bei der oben beschriebenen Ver-

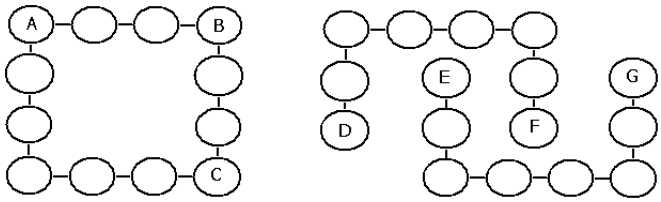

Abbildung 5: In der Abbildung stehen Kreise für Features und Kanten dafür, dass zwei Features zusammengebunden wurden. Die Abbildung soll verdeutlichen, dass das Zusammenbinden von Features transitiv ist. Ist Feature A mit Feature B, und Feature B mit Feature C verbunden, so ist auch Feature A mit Feature C verbunden. Obwohl Feature E räumlich näher an Feature D ist, wurden die Features anders zusammengebunden. Feature D ist mit Feature F, und Feature E mit Feature G verbunden. Die Transitivität ist ein logisches Hilfsmittel herauszufinden, welche Features zu einem Objekt zusammengebunden werden sollen. Siehe hierzu das Paper[16].

arbeitung wird stillschweigend vorausgesetzt, dass die Augen oder Kameras, die zweidimensionalen Bilder zu einem dreidimensionalen Gesamteindruck zusammensetzen. Hierbei gibt es Überdeckungen, so dass ein Gegenstand nie vollständig dreidimensional betrachtet werden kann. Hier kommt die Autovervollständigung ins Spiel. Es wird angenommen, dass teilweise oder nur aus einer Perspektive gesehene Objekte dreidimensional vervollständigt werden, aufgrund des Wissens, was man über den, oder ähnliche, Gegenstände hat.

Die Schablone "physischer Gegenstand" wird so mit dem dreidimensionalen visuellen Eindruck vom Objekt gefüllt, welches aus der Umwelt herausgetrennt wurde.

7.1.2 Die Generierung des Konzepts einer Interaktion von zwei physischen Objekten

Wie weiter oben schon erwähnt, kann eine statische Relation zwischen zwei Objekten nur aus dynamischen Interaktionen abgeleitet werden. Auf einem statischen Bild kann nicht erkannt werden, dass z.B. ein Glas auf einem Tisch eine Relation eingeht, ohne das Wissen über die dynamische Relation, dass das Glas ohne den Tisch zu Boden fällt. Die dynamischen Relationen sind somit die Grundrelationen aus denen die statischen abgeleitet werden. Dieses Ableiten wird im Kapitel über die Konzeptverknüpfung behandelt (siehe Abschnitt 7.3).

Da ein statisches Bild keine dynamischen Vorgänge enthält, benötigt man eine Abfolge von Bildern (Videos). Es läge nahe, dass es Stand der Wissenschaft wäre in Videos dynamische Interaktionen von Objekten zu studieren. Dies ist nicht der Fall. Der Standardfall ist, dass aus Videos raumzeitliche Features extrahiert werden, um dynamische Vorgänge, wie das Winken einer Person zu klassifizieren. Im Prinzip benutzt man hier die Methoden des Deep Learnings, nur das eine weitere Dimension, die Zeit, hinzugekommen ist.

In diesem Buch wird ein geschachtelter Aufbau von Konzepten beschrieben. D.h., dass beim Studium von dynamischen Vorgängen, die physische Welt bereits als komplett in Objekte zerlegt gedacht werden kann. Es reicht dabei die Interaktion von jeweils zwei Objekten zu studieren, da komplexere Interaktionen auf die Interaktion von zwei Objekten zurückgeführt werden können. Bei der Komplettzerlegung der Welt in Objekte ist zu beachten, dass selbst der "Boden" zu einem Objekt wird. Ein Zimmer kann dabei als Gesamtobjekt oder in seine Einzelobjekte zerlegt verstanden werden. Ebenso lassen sich fast alle Objekte, wie eine Flasche, in Teilobjekte zerlegen, wie dem eigentlichen Gefäß und dem Verschluss.

Gehen zwei Objekte eine dynamische Relation ein, so muss diese exklusiv nur diese beiden Objekte betreffen. Bei einigen Interaktionen kann der Interaktionspartner nicht ausgemacht werden, was unter dem Punkt "Autovervollständigung" diskutiert wird.

Da sich der Autor bei diesem Punkt an keiner wissenschaftlichen Veröffentlichung orientieren kann, wird ein Weg zu einer Lösung präsentiert, der sich von der mathematischen Logik her nähert. In der Logik kann Objekt A mit Objekt B über die Relation R miteinander in Beziehung stehen. Nur sind A und B punktartige Bezeichner, so dass die Relation R nicht wirklich viel über die gegenseitige Dynamik aussagen kann. Der erste Schritt war die Objekte A und B als Mengen zu verstehen, so wie ein Objekt aus einer Menge von sensorischen Daten besteht. Die Relation R würde dann diese Mengen punktweise miteinander verbinden. Hierzu wäre es nötig den einzelnen Punkten eines Objekts Eigenschaften zuzuordnen, so wie jeder Punkt einer Porzellantasse aus einzelnen Porzellanelementen besteht. Ein Problem ist, dass man an dieser Stelle schon den Objekten Eigenschaften zuordnen müsste, was nach dem hier vorgestellten Modell erst später geschieht.

Der nächste Schritt brachte eine Lösung. Zunächst muss festgestellt werden, dass die Welt komplett in Objekte zerlegt ist, und eine Relation gar nicht direkt sichtbar ist, sondern vom kognitiven System in die Welt hineininterpretiert wird. Als zweites stehen für die Bildung von Interaktionskonzepten zu

diesem Zeitpunkt nur die Features der Objekte zur Verfügung. Die Lösung ist nun, dass die beiden fraglichen Objekte A und B zu einem Gesamtobjekt O zusammengebunden werden. In dem zeitlichen Ablauf wird dann die Gesamtdynamik der Verbindung O und die Einzeldynamiken von Objekt A und B betrachtet. Damit von einer Interaktion die Rede sein kann, muss die Objektpermanenz eines oder beider Objekte gebrochen sein. Dass z.B. ein Objekt zu Bruch geht oder sich verformt. Die Objektpermanenz wird hierbei etwas erweitert. Ein Objekt bricht seine Objektpermanenz, wenn es z.B. seinen Bewegungszustand ändert. Bei einer Interaktion von Objekten kann diese Permanenz des Bewegungszustandes geändert werden, wenn z.B. eine Billardkugel an der Bande abprallt. In diesem Beispiel ist das eine Objekt die Billardkugel, das andere Objekt nicht der Billardtisch, sondern das Teilobjekt "Bande". In der dynamischen Interaktion von Billardkugel und Bande bleibt die Bande unverändert und die Billardkugel ändert seinen Bewegungszustand.

Bei einer Interaktion von Objekten ändern sich also entweder die Features des einen oder beider Objekte, oder eine andere Form der Permanenz, wie der Bewegungszustand, wird gebrochen. Zum Charakterisieren der Interaktion, aus denen später die Eigenschaften einer Interaktion abgeleitet werden, stehen hierzu die Art und Weise, wie sich die Features, oder eine andere Form der Permanenz ändert, zur Verfügung, oder, was hier neu eingeführt wird, eine Beschreibung der Gesamtdynamik. Eine Gesamtdynamik kann z.B. durch die sogenannte "Hankel Matrix"[12] beschrieben werden. Diese betrachtet die Dynamik eines Vorganges und liefert einen Wert, wie stark sich die beobachteten Parameter ändern.

Bevor etwas zur Exklusivität der beteiligten Objekte gesagt wird, soll die Autovervollständigung betrachtet werden. Jede Veränderung eines Objekts wird als Interaktionskonzept verstanden. Steht z.B. ein Tisch draußen an der Luft und auf ihm liegt ein Papier, welches plötzlich vom Tisch gefegt wird, so sucht sich, um die Schablone der Interaktion vollständig zu machen, das kognitive System einen Interaktionspartner. Für das Herunter-fegen des Papiers wird der Wind verantwortlich gemacht. Auf diese Weise wird ein nicht direkt sichtbares Objekt gebildet: "Der Wind". Der der Interaktionspartner für das herunter fliegende Papier ist.

Die Exklusivität hat viel mit der Autovervollständigung zu tun. Es bleibt erst mal die Frage offen, woran sich ein kognitives System orientiert, um einen exklusiven Interaktionspartner auszumachen. Im Billardbeispiel verändert die Kugel ihre Richtung, aber es ist die Leistung des kognitiven Systems

nicht die restlichen Kugeln oder andere Objekte hierfür verantwortlich zu machen, sondern das Objekt Bande.

7.1.3 Die Generierung des Handlungs-Bewertungs-Konzepts

Das Handlungs-Bewertungskonzept spielt nicht nur eine zentrale Rolle bei physischen Konzepten, es wird auch ein analoges mentales Handlungs-Bewertungs-Konzept eingeführt, welches die Basiseigenschaften des hier beschriebenen Konzepts übernimmt. Das mentale Handlungs-Bewertungs-Konzept stellt das Hantieren mit Konzepten für das kognitive System wieder als eigene Handlungen dar, als mentale Handlungen.

Zentral für das Handlungs-Bewertungs-Konzept ist, dass es nicht auf einzelnen Objekten oder Interaktionen von Objekten operiert, sondern auf einem (kompletten) Objektbereich. Der physische Objektbereich ist die weiter oben definierte "Szene". Zur Wiederholung, eine Szene, ist eine Ansammlung von physischen Objekten, bei der man ALLE möglichen oder nicht möglichen dynamischen Interaktionen berücksichtigt. Später werden zusammengesetzte Konzepte betrachtet, die hier ebenfalls Teil der Szene sein können. Statische Relationen sind entweder nicht eintretende dynamische Relationen oder das Ergebnis einer Handlung. Für die Szene, ob simuliert oder aktuell beobachtet, muss ein eigenständiger Datentyp definiert werden.

Neu hinzu kommt hier der Begriff der "Kausalität". Wie bereits erwähnt, kennt die Physik den Begriff der Kausalität nicht per se. Um von Kausalität sprechen zu können, muss zuerst eine Konstellation in einer Szene ausgezeichnet werden, um dann alle Faktoren berücksichtigen zu können, die zu der Konstellation geführt haben. Kausalität wird also von kognitiven Systemen in die Welt hinein interpretiert, die Physik kennt nur Teilchen und deren Wechselwirkungen, die per se keine Konstellation auszeichnen.

Eine Konstellation in einer Szene wird dadurch ausgezeichnet, dass ein kognitives System diese als gut oder schlecht bewertet. Liegt eine solche Konstellation vor, so wird "sternförmig" mit ihr gearbeitet. Sternförmig heißt, dass auf der einen Seite analysiert wird, welche Faktoren kausal zu der Konstellation geführt haben, und auf der anderen Seite wird analysiert wie Subjekte von der Konstellation betroffen sind. Selbst wenn es nur um *ein* Subjekt geht, kann die Konstellation für die Bewertung multifaktoriell sein. D.h., dass die Subjekteigenschaften auf verschiedene Weise betroffen sind.

Ein kausaler Faktor, wenn in der Szene nicht sogar der einzige, ist stets das Handlungssubjekt. Dieses wird, wie schon beschrieben, als Pointersubjekt

behandelt, da es keine weiteren Eigenschaften hat, als verantwortlich für die Handlung zu sein. Die Handlungs-Bewertungs-Schablone wird mit der durchgeführten Operation und der Wirkung auf den Objektbereich gefüllt, wobei die Wirkung zu einer Bewertung vom System führt. Im Gegensatz zum physischen Handlungs-Bewertungskonzept kennt das mentale Gegenstück keine äußere Kausalität. Jedes Hantieren, wenn auch teilweise automatisiert, schreibt sich das Subjekt als eigene mentale Handlung zu.

Eine physische Handlung kann nur durch Interaktionskonzepte realisiert werden. Dies sind Interaktionen des eigenen Körpers mit Objekten der Szene oder angestoßenen Interaktionen von zwei externen Objekten. Die verwendeten Interaktionen und die benutzten Objekte dienen später zur Charakterisierung der Handlung.

Startet die Betrachtung des Handlungs-Bewertungskonzepts mit der ganzen Szene, so ist hier das Informationsabstreifen die sternförmige Analyse. Diese weist Teile der Szene zurück, die keinen kausalen Einfluss hatten, und analysiert welche Beeinflussung des Subjekts wirklich relevant sind. Diese "Verschlankung" des Handlungs-Bewertungs-Konzepts kann das kognitive System auch im weiteren Verlauf noch beschäftigen. So ist es beim natürlichen Vorbild des Menschen üblich, dass im Nachhinein Beeinflussungen als irrelevant Beiseite geschoben werden.

Handlungen setzen sich üblich aus Handlungsketten zusammen, welche eine Abfolge von Handlungen sind (was eine Handlungskopplung ist wird in Abschnitt 7.3 eingeführt). Meist liegt nur eine Bewertung für die gesamte Handlungskette vor. Hier beginnt, das, was als "Blockbildung" bezeichnet wird. Liegen verschiedene Handlungsketten mit Bewertung vor, bei denen einige Blöcke (Einzelhandlungen) identisch sind, so erhält ein einzelner Block verschiedene Bewertungen. Auf die Art kann ein Block insgesamt als positiv bewertet werden auch wenn dieser in einer einzelnen Handlungskette zu einem negativen Ergebnis geführt hat. Die Blockbildung ist der einzige Weg einzelne mentale Handlungen zu bewerten, da mentale Handlungen nur über den Umweg der physikalischen Realisierung eine Bewertung erhalten können.

Es wurde im Abschnitt 3 herausgestellt, dass das kognitive System aufgaben-zentriert arbeitet. Hierbei gibt es eine Annäherung von zwei Seiten. Auf der einen Seite stehen die Handlungsmöglichkeiten des kognitiven Systems auf der anderen Seite die Ziele. Von der Seite der Ziele her muss eine Kaskade von Zwischenzielen aufgebaut werden. Von der Seite der Handlungen her muss gelernt werden, wie Handlungen aufeinander aufbauen. Hierzu muss das System lernen, was für eine Art von Konstellation in der Szene

ein Zwischenziel darstellt. Allgemein kann gesagt werden, dass Zwischenziele, auf denen weitere Handlungen aufbauen können, eine Art von Konstanz aufweisen müssen. Verflüchtigt sich sofort das Handlungsergebnis kann auf diesem nicht aufgebaut werden. Wie Ziele und Handlungen zusammengeführt werden, und wie die anfänglich zufälligen Motorsignale kontrolliert werden, wird im Abschnitt 10 über das Lernen nochmal diskutiert.

7.2 Die Konzeptverallgemeinerung, Subsumierung und Differenzierung

Bei der Konzeptverallgemeinerung zeigt sich die erste Stärke des Modells. Die Verallgemeinerung von Konzepten beruht auf gemeinsamen Eigenschaften und das Modell liefert auf natürliche Weise eine Zerlegung, aus welchen Quellen die Eigenschaften gewonnen werden können. Es wird behauptet, dass diese Zerlegung nach Eigenschaftsquellen vollständig ist. Wie schon bei der Trennschärfe und der Vollständigkeit der drei physischen Konzepttypen kann dies nicht mathematisch genau gezeigt werden. Sie ergibt sich daraus, dass kein Beispiel gefunden werden kann, das man nicht eindeutig einer der genannten Quellen zuordnen kann.

Die Konzeptverallgemeinerung, die Subsumierung sowie die Differenzierung von Konzepten beruhen auf dem Begriff der "Eigenschaft". Bei der Konzeptverallgemeinerung haben zwei Objekte oder zwei Konstituenten eines Konzepts die gleiche Eigenschaft. Bei der Subsumierung liegt bereits ein Konzept mit allgemeineren Eigenschaften vor, unter das das gerade betrachtete Konzept mit seinen spezielleren Eigenschaften subsumiert wird. Bei der Differenzierung wird ein Konzept in zwei Konzepte zerlegt, da eine Eigenschaft gefunden wurde, die die beiden Konzepte unterscheidet.

Der Begriff "Eigenschaft" ist philosophisch schwer zu greifen. Es gibt in der Literatur kein eindeutiges Verfahren, wie z.B. die gemeinsamen Eigenschaften zweier Objekte bestimmt werden können. Ein solches spezielleres Verfahren soll hier für das Konzeptmodell entwickelt werden, indem die Quellen bestimmt werden, aus denen Eigenschaften gewonnen werden. Der Begriff "Eigenschaft" definiert sich dann daraus, wie diese gewonnen wurde, d.h. eine Eigenschaft wird Interaktionseigenschaft genannt, wenn diese aus der Art und Weise gewonnen wurde, wie Objekte interagieren.

Zunächst soll geklärt werden, Wem oder Was man einer Eigenschaft zuordnen kann. Wie kurz angedeutet ist der sensorische Eindruck "Rot" keine

Eigenschaft. Eigenschaften werden Objekten oder hier Konzepten zugeordnet. D.h., dass erst zwei "rote" Objekte erkannt werden müssen, um diesen dann die gemeinsame Eigenschaft "Rot" zuordnen zu können. Eigenschaften entstehen so nicht auf der Feature-ebene, sondern erst später nach der Konzeptbildung. Sie definieren in welchem Verhältnis verschiedene Konzepte gleichen Typs stehen.

Allgemein kann gesagt werden, dass es für jedes Konzept drei Quellen gibt, aus denen Eigenschaften gewonnen werden können. Die drei Quellen repräsentieren die drei Konzepttypen, so dass für ein Konzept als Quelle der eigene Konzepttyp oder einer der anderen beiden Konzepttypen dienen kann. Um dies zu verdeutlichen wird dies an den einzelnen Konzepttypen gezeigt. Aus technischer Sicht muss den Datentypen, die für jeden Konzepttyp angelegt wurden, drei weitere Felder zugeordnet werden, entsprechend den drei Eigenschaftsquellen.

Die Beschreibung der Eigenschaftszuordnung beginnt mit dem innersten Konzept, dem Objekt, und schließt dem äußersten Konzept, der Handlung ab. Allgemein kann gesagt werden, das äußere Konzepte die Eigenschaften der inneren Konzepte klassifizieren, also neue Eigenschaftsklassen einführen. Während innere Konzepte die Eigenschaften von äußeren Konzepten differenzieren, also weitere Unterscheidungen innerhalb der Eigenschaft einführen. Werden zwei Konzepte gleichen Typs verglichen und als ähnlich befunden, hat dies zur Folge, dass diese bei einer Verknüpfung von Konzepten in ihrer Funktion her zunächst als identisch angenommen werden.

7.2.1 Eigenschaften von physischen Objekten

Dem oben Beschriebenen nach gibt es drei Quellen für die Eigenschaftsbildung von physischen Objekten. Eine Quelle kommt aus der Bildung des Konzepts selbst. Diese Quelle beschreibt, dass bei zwei Objekten etwas im sensorischen Input gleich ist. Aufgrund der Form können eine Tasse, ein Glas und eine Vase ein Oberkonzept bilden.

Die nächste Quelle ist die Interaktion von Objekten. Ein Schwamm und ein Backstein haben zwar dieselbe Form, verhalten sich aber anders bei der Interaktion mit anderen Objekten. Ein Schwamm verformt sich leicht bei der Interaktion mit anderen Objekten während das ein Backstein nicht tut. Der Schwamm würde die Interaktionseigenschaft "weich" bekommen, was nichts weiter als ein Wort für die leichte Verformbarkeit ist. Da das Relationskonzept dem Objektkonzept übergeordnet ist, sind die Interaktionseigenschaften von

Objekten neu gebildete Eigenschaften.

Die letzte Quelle ist das Handlungsbewertungskonzept. Hier geht es darum, welche Funktion die Objekte innerhalb einer Handlung haben. Während in obigen Beispiel sowohl eine Tasse als auch ein Glas benutzt werden kann, um aus diesen zu trinken, so tun wir dies nicht mit einer Vase. Das übergeordnete Handlungskonzept führt neue Objekteigenschaften wie "Trinkgefäß" oder "Gefäß zur Aufbewahrung von Blumen" ein. Die gemeinsame Form dieser Objekte führte zunächst zur Annahme der gleichen Funktion. Erst die verschiedenen Handlungen, die mit diesen ausgeführt werden, trennt die Objektklasse.

Der Clou an dieser Beschreibung von Eigenschaften von physischen Objekten ist, dass behauptet wird, dass jede Eigenschaft, die der Mensch über ein Objekt bildet, eindeutig einer der drei Quellen zugeordnet werden kann.

7.2.2 Eigenschaften von Interaktionen

Bei den Eigenschaften der Interaktionen und bei den Eigenschaften des Handlungsbewertungskonzepts muss die Schachtelung der drei Konzepttypen berücksichtigt werden. Eine Interaktion kann aus sich selbst heraus charakterisiert werden, indem bei zwei Interaktionen die dynamischen Eigenschaften gleich sind. Z.B. können zwei Objekte dynamisch gleich auf den Boden fallen (und zerschellen). So können diesen beiden Interaktionen eine gleiche Eigenschaft zugeordnet werden, die mit dem Symbol "herunterfallen" versehen werden kann.

Um Interaktionen von Objekten zu beschreiben, geht man von einer kompletten Zerlegung der physischen Welt in Objekte aus. Die Interaktion von Objekten ist dann die Schachtel, die über dieser Zerlegung liegt. Ein gleicher Vorgang, wie das zu Boden fallen eines Objekts, kann verschieden charakterisiert werden, je nachdem welche Objekte in die Interaktion eingesetzt werden. Beim zu Boden fallen ist *ein* Objekt immer der Boden, die Interaktionseigenschaft wird aber weiter differenziert, je nachdem welches Objekt zu Boden fällt: Eine Feder führt zu einem anderem Charakter der Interaktion als ein Stein.

Die dritte Quelle, welche eine Interaktion charakterisiert, ist wiederum das Handlungsbewertungskonzept. Auch hier ordnet man einer Interaktion verschiedene Eigenschaften zu, je nachdem, welche Rolle diese in einer Handlung spielt. Z.B. spielt das Aufschlagen eines Eies mit dem Messer, welches eine dynamische Interaktion darstellt, eine Rolle bei der Zubereitung eines

Essens. Da das Handlungskonzept dem Interaktionskonzept übergeordnet ist, führt es neue Eigenschaften ein. Die Interaktion zwischen Objekt und Werkzeug hat hier die neue Eigenschaft: "Essenszubereitungsinteraktion".

7.2.3 Eigenschaften von Handlungsbewertungskonzepten

Aufgrund der Schachtelung der Konzepttypen, soll hier von einer inneren und äußeren Charakterisierung die Rede sein. Die innere Charakterisierung ergibt sich daraus, welche Objekte und Interaktionen bei einer Handlung eine Rolle spielen, analog dazu, welche Objekte in eine Interaktion eingesetzt werden. Da sowohl das Objektkonzept als auch das Interaktionskonzept dem Handlungskonzept untergeordnet sind, differenzieren diese Handlungskonzepte. Um im obigen Beispiel zu bleiben, wird die Handlung "Essenszubereitung mithilfe eines Werkzeugs" durch die Objekte "Ei" und "Messer" und der unsanften Interaktion zwischen beiden, weiter konkretisiert bzw. differenziert.

Die äußere Charakterisierung ergibt sich aus der beschriebenen sternförmigen Analyse eines Handlungsbewertungskonzepts. Ein Handlungsbewertungskonzept startet mit der Beschreibung der kompletten Szene und der Einwirkung des Handlungssubjekts. In Folge der Analyse wird dieses verschlankt, indem die wirklich relevanten kausalen Einflüsse auf das Handlungsergebnis und die wirklich relevanten Beeinflussungen des Subjekts herausgearbeitet werden. Dementsprechend kann ein Handlungsbewertungskonzept nach kausalen Einflüssen und Beeinflussungen von Subjekten kategorisiert werden.

Im Physischen können kausale Einflüsse verschiedene Ursprünge haben. Sind zum einen Handlungssubjekte die kausale Ursache für eine entstandene Konstellation, wobei vereinfacht angenommen wird, dass diese statische Relationen in statische Relationen überführen (siehe Abschnitt 7.3.4), so spielen hier (ungewollte) dynamische Relationen eine Rolle. Die Konstellation ergibt sich kausal also aus den Handlungen und den (ungewollten) dynamischen Interaktionen.

Das Handlungsbewertungskonzept wird später auf mentale Handlungen übertragen. Im Mentalen gibt es keine dynamischen Interaktionen, alle Ergebnisse sind die Folge von mentalen Handlungen. Da das Hantieren mit Konzepten seriell abläuft, müssen die mentalen Handlungen für eine Kausalitätsanalyse umsortiert werden (siehe Abschnitt 11.4). Es wird behauptet, dass nach dem Prinzip der äußeren Charakterisierung alle Eigenschaften erfasst werden, die wir mentalen Handlungen zuweisen. D.h., dass wir kei-

ne weiteren mentalen Begriffe bilden, die aus dem Hantieren mit Konzepten hervorgehen, die nicht auf die innere oder äußere Charakterisierung von mentalen Handlungen zurückzuführen sind.

Dies gilt solange bis das später beschriebene Selbstmodell eingeführt wird (siehe Abschnitt 13). Im Selbstmodell werden die beiden Subjekttypen, das pointerartige Handlungssubjekt und das eigenschaftsbehaftete betroffene Subjekt, in *einem* Selbstmodell zusammengefasst. Diesem Selbstmodell können Eigenschaften zugeordnet werden, die aus der Summe von physischen und mentalen Handlungsbewertungskonzepten hervorgehen. Hier wird also dem Subjekt, welches in seiner Summe "Person" genannt wird, als ganzes Eigenschaften zugeordnet. Das ebenfalls später eingeführte sprachliche Interaktionsmodul mit anderen Subjekten sorgt dafür, dass Subjekte sich gegenseitig Eigenschaften zuordnen können. Dies ist freilich nur möglich, da menschliche Subjekte sich voneinander unterscheiden, was bei künstlichen kognitiven Systemen nicht zwingend der Fall sein muss, wenn diese auf persönliche Erfahrungen verzichten und alles Gelernte teilen.

7.2.4 Die Auswirkungen der Schachtelung auf das Verhältnis der Konzepttypen

Dadurch das innere Konzepte in die Schachtel der äußeren Konzepte eingesetzt werden, ergibt sich ein besonderes Verhältnis. In das Konzept einer Relation werden Objekte eingesetzt. Hierbei können in dieselbe Relation verschiedene Objekte eingesetzt werden. Ebenso verhält es sich bei der Handlung. In dieselbe Handlung können verschiedene Relationen und in diese wiederum verschiedene Objekte eingesetzt werden. Dies führt auf eine baumartige Struktur, dass zu einer einzelnen Handlung mehrere Relationen und zu einer einzelnen Relation wiederum mehrere Objekte gehören. Dies schließt nicht aus, dass eine andere Handlung, die eine andere Wurzel darstellt, sich derselben Objekte bedient. In diesem Fall kann ein Objekt auf verschiedene Weise benutzt werden. Eine völlige Trennung eines Objekts oder Relation aus der entsprechenden Handlung kann nur über eine Analogiebildung (Abschnitt 7.4) geschehen.

7.2.5 Klassische hierarchische Ontologien

In der künstlichen Intelligenz ist es üblich Objekte in sogenannten hierarchischen Ontologien zu ordnen. Hierbei werden z.B. die Objekte "Apfel", "Birne"

und "Kirsche" zum Oberbegriff "Obst" zusammengefasst. Hieraus ergibt sich eine Hierarchie von Objekten. Der Hierarchiegedanke wird meist über die Objekte hinaus weiter getrieben, so dass Szenen, wie eine Küchen- oder Strandszene, danach erkannt und klassifiziert werden, je nachdem welche Objekte vorzufinden sind[3].

Auch wenn sich teilweise durch Abstraktion Hierarchien von Konzepten bilden, so durchbricht das hier vorgestellte Modell den reinen Hierarchiegedanken. Ein Handlungsbewertungskonzept beruht darauf, dass Wirkungen in eine Szene eingehen, und dass Beeinflussungen von Subjekten aus dieser abgeleitet werden. Dies ist nicht durch eine Hierarchie von Ontologien darstellbar.

7.3 Die Konzeptverknüpfung

7.3.1 Die Baumstruktur von zusammengesetzten Objekten

Bei der Konzeptverknüpfung treten zwei weitere Vorgehensweisen zu Tage, wie der Mensch mit Konzepten umgeht. Die erste Beobachtung ist, dass es bei der Beschreibung eines Sachverhalts nicht bei der einfachen dreifachen Verschachtelung bleibt: Der Mensch bildet weitere tiefere Verschachtelungen um Dinge zu beschreiben. Beim Umgang mit einer Flasche behandelt dieser die Flasche einmal als Gesamtobjekt in der üblichen einfachen Verschachtelung, um z.B. das Verhalten der Flasche in einer Interaktion zu beschreiben. Er beobachtet etwa wie sich die in ihr enthaltene Flüssigkeit bei Bewegung der Flasche verhält. Möchte er an den Inhalt der Flasche, betrachtet er die Flasche als ein durch Konzepte zusammengesetztes Objekt. Das Konzept "Gesamtflasche" ist aus den Teilkonzepten "Flasche" und "Verschluss" aufgebaut. Die Flasche und der Verschluss gehen dabei eine statische Relation ein, die den Sinn erfüllt, dass die dynamische Relation des Hinauslaufens der Flüssigkeit nicht eintritt.

Ein aus Konzepten zusammengesetztes Gebilde ist vom Konzept her der Typ, der beim Gebilde den übergeordneten Konzepttyp bildet. Dies sei an mathematischen Termen veranschaulicht. Der Ausdruck "$(3 + 5) \cdot 4$" ist ein Produkt, da die Multiplikation zuletzt ausgeführt wird und so den Typ des Terms definiert, wohingegen "$3 + 5 \cdot 4$" eine Summe ist. Analog dazu kann ein aus Konzepten zusammengesetztes Gebilde entweder ein Objekt bilden oder eine Relation, je nachdem welches das übergeordnete Konzept ist. Ein zusammengesetztes Gebilde bildet auf diese Art eine Baumstruktur, wobei

die Wurzel den Konzepttyp des gesamten Gebildes bestimmt.

Die zweite Beobachtung beschreibt, wie der Mensch Handlungen mit der Welt verknüpft. Zerlegt er die Interaktion mit der Welt zunächst in die drei Konzepttypen, wird die Handlung nochmal gesondert betrachtet. Das, worauf die Handlung wirkt wird konzeptionell zu einem Gebilde zusammengefasst und der Handlung gegenübergestellt. Hierbei wird die Schnittstelle betrachtet, wie die Handlung an dem Gebilde angreift. Ist die Aufgabe z.B. ein Paket von A nach B zu stellen, so ist die Schnittstelle die Art und Weise wie das Paket gegriffen werden kann.

Hierdurch ergibt sich eine einheitliche Vorgehensweise, wie Aufgaben gelöst werden. Zuerst bildet der Agent ein Modell des Sachbereichs, der manipuliert werden soll, und stellt danach eine Lösungsstrategie auf, welche Handlungen er ansetzen muss, um sein Ziel zu erreichen.

7.3.2 Eigenschaften bei der Objektverknüpfung

Welche Funktion haben Eigenschaften von Konzepten? Sie erfüllen genau zwei Funktionen. Die erste Funktion ist, dass allgemeine und weniger allgemeine Eigenschaften existieren, die durch Abstraktion gebildet werden. Hierdurch entstehen Konzepthierarchien. Die eigentliche Aufgabe von Eigenschaften ist aber das richtige Verknüpfen von Konzepten. Eigenschaften werden aus wahrgenommenen Verknüpfungen heraus gelernt und müssen zu einem gewissen Grad abstrahiert werden, damit neue Verknüpfungen gebildet werden können.

Zwei der Eigenschaftsfelder eines jeden Konzepttyps geben eine gelernte Verknüpfung an. Das Eigenschaftsfeld, welches den eigenen Konzepttyp als Quelle benutzt, dient ebenfalls der Verknüpfung. Z.B. bilden ein Glas und eine Tasse aufgrund der Form ein Oberkonzept “Trinkgefäß”. Diese Zusammenordnung hat ebenfalls die Aufgabe Konzepte zu verknüpfen, indem aufgrund der Form von einer Tasse geschlossen werden kann, dass man aus ihr trinken kann, wenn dies vorher nur bei einem Glas beobachtet wurde.

Konzepte werden auf der subsymbolischen Ebene durch eine Simulation des Objektbereichs verknüpft. Würde man Konzepte nur symbolisch verknüpfen, indem die Objekte die passenden Eigenschaften besitzen, hätte man nicht folgendes Feedback: Zusätzlich zur Handlungsbewertung gibt es eine Bewertung, wie gut oder schlecht sich Objekte aufgrund ihrer Eigenschaften verknüpfen lassen. Dies ist eine neue Form der Bewertung, die zunächst nicht auf das Urziel hin bewertet wird, sondern ein reiner Lernprozess, der

eine solche Urzielbewertung nicht braucht.

Konzepte werden allein anhand ihrer Eigenschaften verknüpft, so dass bei der subsymbolischen Verknüpfung nur auf diese geachtet werden müssen. Da stets neuartige Verknüpfungen gebildet werden müssen, ist nicht klar, ob die abstrahierten Eigenschaften sich wirklich in dieser Form verknüpfen lassen. Das Verknüpfungswissen, stellt einen wichtigen Teil des Weltwissens dar, und ist vom Typ her ein Wissen über Eigenschaften. Genaugenommen werden so Eigenschaften von Konzepten weiter differenziert.

7.3.3 Aufbau eines Modells und Ableitung statischer Relationen

Der Begriff "Modell" soll hier so definiert werden, dass dieses aus zusammengesetzten Konzepten besteht und einen Sachbereich wiedergibt bzw. beschreibt. Es soll erst allgemein von Sachbereich die Rede sein, so dass der Sachbereich mentale oder physische Zusammenhänge beschreiben kann. Für ein Modell soll ein weiterer Datentyp eingeführt werden. Aus Bequemlichkeit in der Definition kann man zulassen, dass ein einzelnes Basiskonzept schon ein Modell ist.

Beschreibt der Sachbereich physische Zusammenhänge, so kann dies etwa die Beschreibung eines (technischen) Gegenstandes oder eines Organismusses sein, welches man nicht als Subjekt interpretiert, wie etwa eine Pflanze. Das Modell umfasst nicht die Beschreibung, wie mit diesem Gegenstand umzugehen ist, dies würde unter die "Lösungsstrategie" fallen. Daher wäre zu vermuten, dass nur die Konzepte eines physischen Objekts und Interaktionen zwischen diesen benutzt werden können. Eine physikalische Beschreibung würde so aussehen. Ein kognitives System hat zusätzlich die Möglichkeit "depersonalisierte Handlungskonzepte" zu benutzen. Diese sind die Übertragung von Handlungskonzepten auf den physischen Bereich, wo an sich gar kein Subjekt am wirken ist. Dies geschieht etwa bei technischen Geräten, wie einem Mikrochip, wo Beschreibungen zulässig sind, dass der Chip auf den Speicher zugreift, obwohl physikalisch gesehen nur Teilchen interagieren. Ebenso kann man von einer Pflanze behaupten, dass sie sprießt, was ebenfalls eine Handlung ist, wo das gleiche physikalische Argument gegen eine solche Beschreibung zu entgegnen wäre.

Eine statische Relation definiert sich, wie bereits erwähnt, entweder dadurch, dass diese ein Handlungsergebnis beschreibt, oder dass eine dynamische Relation nicht eintritt. Bei der Beschreibung eines Modells können depersonalisierte Handlungsergebnisse auftauchen, wie z.B. der Zustand, dass

die Daten von der Festplatte in den Arbeitsspeicher geladen wurden, so dass auch statische Relationen als Handlungsergebnisse auftauchen. Bei der Ableitung einer statischen Relation, wird ein dynamisches Basiskonzept zwischen zwei Objekten betrachtet. Man benötigt eine bereits gewonnene statische Relation, die das Nicht-Eintreten der Ersten beschreibt. Dies soll an Beispielen verdeutlicht werden: Man stellt z.B. ein Glas auf den Tisch, was ein statisches Handlungsergebnis beschreibt, damit dieses Glas nicht dynamisch zu Boden fällt. Die zugeschraubte Flasche verhindert, dass die Flüssigkeit herausläuft, wobei der Schraubverschluss eine statische Relation mit der Flasche eingeht. Der Steckverschluss eines Gartenschlauchs, statisch an einen Wasserhahn angeschlossen, verhindert das unkontrollierte verspritzen des Wassers aus dem Hahn. Man stellt sich mittig auf den Stuhl, wenn man seine Größe erhöhen will, um ein Umkippen des Stuhls zu verhindern. Ein Reifen wird mittig an ein Tischbein gelehnt, damit dieser nicht wegrollt. Bei den beiden letztgenannten Beispielen spielt die genaue relative statische Position von zwei Objekten eine Rolle damit eine dynamische Relation nicht eintritt. Auch wenn die dynamischen Relationen die erst gelernten Konzepte sind, hangelt sich das kognitive System von bereits abgeleiteten statischen Relationen zu weiteren statischen Relationen fort.

7.3.4 Die Stellung von statischen Relationen

An dieser Stelle soll geklärt werden, was es heißt, dass sich eine statische Relation aus einer Nicht-Eintretenden dynamischen Relation "definiert". Einer statischen Relation ordnet ein kognitives System einen "Sinn" zu. Im Gegensatz zu dynamischen Relationen (Interaktionen), die aufgrund der geltenden Gesetze in der Natur einfach geschehen, und dementsprechend nur gelernt werden können, hat die statische Relation ein logisches "Freifeld", in das ein Sinn der Relation eingetragen werden muss. Würde ein Glas auf dem Tisch nicht nur mit dem Tisch eine statische Relation eingehen, sondern im Prinzip zu allen Objekten der Szene, betrachtet ein kognitives System nur statische Relationen, die einen Sinn erfüllen. So werden in einer Szene nur statische Relationen hervorgehoben, die eine solche Sinnzuordnung haben. Die Sinnzuordnung geschieht, nach dem hier beschriebenen Modell, dadurch, dass die statische Relation ein Handlungsergebnis oder ein Nicht-Eintreten einer dynamischen Relation beschreibt. Im speziellen kann sie auch eine Konstellation in der Szene beschreiben, die vom System bewertet wird, aber nicht das Ergebnis einer Handlung eines Subjekts ist. Dadurch wird verhindert,

dass nicht bei N Objekten in der Szene N^2 statische Relationen beschrieben werden müssen, sondern nur diese, denen ein Sinn zugeordnet werden kann. Im Sinne des Modells führt man für statische Relationen einen eigenen Datentyp ein, der sich grundlegend vom dynamischen Pendant unterscheidet. Während dynamische Relationen durch das dynamische Verhalten zwischen zwei Objekten beschrieben werden, enthält die Schablone einer statischen Relation, die zwei fraglichen Objekte, eine Beschreibung der gegenseitigen Konstellation, und ein Feld in das auf die oben beschriebene Art ein Sinn eingetragen wird.

Zwischen- und Endergebnisse einer Handlung zeichnen sich dadurch aus, dass die gewonnene Konstellation eine gewisse "Konstanz" aufweist. Dies soll durch die Beschreibung der Konstellation durch statische Relationen geleistet werden. Statische Relationen haben nach dem oben Beschriebenen eine Sinnzuweisung. Diese Sinnzuweisung soll weiter im Hinblick auf die Lösung einer Aufgabe analysiert werden. Zuerst soll festgestellt werden, dass ein "Sinn" nicht völlig frei im Raum steht, sondern nur im Hinblick einer Aufgabe interpretiert werden kann. Stellt die statische Relation das Ergebnis einer Handlung dar, so bezeichnet die Sinnzuordnung, dass im Folgenden weniger Handlungsschritte vollbracht werden müssen, um ein Ziel zu erreichen. Hierbei wird ein besonderes Maß verwendet, um die Entfernung zum Ziel zu messen: Die Entfernung zu einem Ziel bemisst sich an der Anzahl an Einzelhandlungen, die noch nötig sind um das Ziel zu erreichen. Was als Einzelhandlung gilt, oder wie Handlungsketten in Blöcke zerlegt wird, wird im Abschnitt 7.3.6 über freie Handlungen nochmal diskutiert.

Die Sinnzuordnung einer statischen Relation, die eine entstandene Konstellation beschreibt, die nicht das Handlungsergebnis eines Subjekts ist, geschieht analog: Der Sinn entsteht dadurch, dass bemessen wird, inwieweit die Konstellation von einem Ziel entfernt ist. Kann die Konstellation nicht mit einem Ziel in Verbindung gebracht werden, so bewertet das kognitive System die Konstellation nicht und es gibt somit keine Sinnzuweisung.

Führt das Subjekt eine Handlung aus, ist es stets bemüht, ungewollte dynamische Relationen zu verhindern. In dem Sinne definieren Nicht-Eintretende dynamische Relationen eine herbeigeführte Konstellation, und weisen den statischen Relationen in dieser Konstellation so einen Sinn zu.

7.3.5 Aufbau einer Lösungsstrategie

Das Verknüpfen von Konzepten zu einer Lösungsstrategie ist das Zusammenführen von Handlungsmöglichkeiten zu einem Ziel, indem die Struktur des Objektbereichs berücksichtigt wird. Der Objektbereich kann erst mal allgemein der mentale oder physische Objektbereich sein. Beim Finden einer Lösungsstrategie werden explizit Konzepte verschiedenen Typs miteinander verknüpft. Ein Konzepttyp ist die Handlung des Subjekts, der andere Konzepttyp ist konzeptionelle Gebilde, welches der Handlung gegenübergestellt wird.

Um bei den hier vorgestellten Beispielen zu bleiben, greift z.B. ein kognitives System ein Glas vom Tisch, führt dieses zum Mund, und trinkt daraus. Hier werden mehrere Handlungen zu einer Handlungskette (siehe weiter unten) verknüpft. Das System muss dabei die verschiedenen statischen und dynamischen Relationen in der Szene berücksichtigen. Zunächst müssen die Eigenschaften des Glases zu einer Trinkaktion passen. Weiter muss die statische Relation zwischen Glas und Tisch und die dynamischen Relationen zwischen Hand und Glas, und Mund und Glas auf die richtige Weise berücksichtigt werden.

Der Objektbereich besteht dabei, wie im vorigen Abschnitt beschrieben, aus einem Modell. Das Hantieren mit einem Gegenstand, der durch ein Modell beschrieben wird, wird auf die gleiche Weise gelernt, wie jede Handlung: Eine Handlung führt zu einer Wirkung in der Szene. Das Modell vom Gegenstand, mit dem hantiert werden soll, erleichtert dabei die Kausalitätsanalyse. Anstatt durch try and error die Effekte des Gegenstandes herauszufinden, liefert das Modell kausale Zusammenhänge zwischen Handhabung und Effekt. Es wird postuliert, dass der Mensch "Mini-"Modelle von jedem Gegenstand lernt, welche alleine zum Ziel haben, Handlung und Effekt zu verknüpfen. Dies ermöglicht ihm den Umgang z.B. mit einem Toaster oder einem Mixer. Für eine Lösungsstrategie soll ein weiterer Datentyp eingeführt werden, der wieder aus definitorischen Gründen auch zulässt, dass eine Lösungsstrategie eine einzige Handlung umfasst.

Im physischen kann zwischen einer Handlungskette und einer Handlungskopplung unterschieden werden. Eine Handlungskette ist eine Folge von Handlungen, die entweder unabhängig voneinander sind, oder nur genau in dieser Reihenfolge ausgeführt werden können, da sie aufeinander aufbauen. Eine Handlungskopplung soll das beschreiben, was beim Menschen der motorische Kortex leistet. Der motorische Kortex ist nicht nur dafür zuständig den

eigenen Körper zu kontrollieren, sondern er lernt auch wie man mit Gegenständen operiert. Eine Handlung wurde logisch als Operator eingeführt, der auf einem Objektbereich operiert. Benutzt das kognitive System z.B. einen Tennisschläger, um einen Ball zu schlagen, so müsste das System mit zwei Operatoren arbeiten. Ein Operator beschreibt, wie mit der Hand der Tennisschläger manipuliert wird, der andere Operator, der auf dem ersten aufsetzt, würde beschreiben, wie der Tennisschläger den Ball manipuliert. Der motorische Kortex eines kognitiven Systems sei abstrakt so definiert, dass dieser aus den beiden Operatoren einen einzigen berechnet. Dadurch wird die Situation für das System so dargestellt, als würde man direkt auf dem Ball operieren, und Handlungen ausführen, wie den Ball nach rechts oder links schlagen. Diese Darstellung deckt sich mit dem Befund aus der Psychologie[4], dass der Mensch Objekte in sein physisches Körperbild aufnehmen kann, d.h. man benutzt den Schläger so, als wäre er Teil des eigenen Körpers. Dies soll hier eine Erklärung finden darin, dass zwei aufeinander aufbauende Operatoren zu einem Operator verrechnet werden.

Allgemein liegt immer eine Handlungskopplung vor, wenn ein Kraftausgleich zwischen einem Objekt und dem eigenen Körper vorliegt. Der Griff des Tennisschlägers geht einen Kraftausgleich mit der Hand ein: Alle Kräfte, die auf den Tennisschläger wirken, werden auf die Hand übertragen und ausgeglichen. Ein analoger Kraftausgleich liegt vor, wenn man von einem Stuhl aufsteht. Hierbei wird der Kraftausgleich zwischen Stuhl und Körper beim Aufstehen gelöst. Der abstrakte motorische Kortex lernt also modifizierte Handlungsoperatoren immer wenn ein Kraftausgleich vorliegt.

7.3.6 “Freies” Objekt und “freie” Handlung

Analog zur Idee der Handlungskopplung wird definiert, wann der Körper des kognitiven Systems und die Objekte in der physischen Welt als freies Objekt bezeichnet werden können. In der Regel kann der Körper und die Objekte der Welt nur als “quasi” frei betrachtet werden. Fliegt ein Objekt nicht frei durch den Raum, und unterliegt nur den physikalischen Gesetzmäßigkeiten für den freien Fall, liegt stets eine Kopplung des Objekts an anderen Objekten vor. Hiermit ist gemeint, dass das Objekt an andere Objekte durch einen Kraftausgleich gebunden ist. Wurde dies am Beispiel des kognitiven Systems veranschaulicht, welches mit seinem Körper auf einem Stuhl sitzt, so liegt ein Kraftausgleich stets vor, wenn sich das System durch den Raum über den Boden bewegt, da die Füße einen Kraftausgleich mit dem Boden herstellen.

Ebenso ist ein Glas auf dem Tisch ebenfalls nur quasi frei, da es mit dem Tisch einen Kraftausgleich herstellt. Manipuliert das System Objekte muss es diese Kraftausgleichskopplungen mitberücksichtigen.

Es wird behauptet, dass das kognitive System quasi freie Objekte als freie Objekte interpretiert, an denen es modifizierte Operatoren anwendet, die genau die Kraftausgleichskopplungen berücksichtigen. Es betrachtet seinen Körper also als freies Objekt, in dem es je nach Situation modifizierte Operatoren anwendet, um sich z.B. von der Sitzposition zu lösen, oder seinen Gang über den Boden zu steuern. Ebenso betrachtet es das quasi freie Glas als freies Objekt, indem es modifizierte Operatoren anwendet, um es von der Kopplung an den Tisch zu befreien. Nimmt es dabei das Glas in die Hand entsteht eine erneute Kraftausgleichskopplung zwischen Glas und Hand. Das Vorgehen des kognitiven Systems ist also ein Umgang mit quasi freien Objekten, wobei Kraftkopplungen gelöst und durch neue ersetzt werden. Damit das System in freien Objekten Denken kann, ist es also nötig, dass das System je nach vorliegender Kraftkopplung modifizierte Operatoren verwendet, und sich nicht bewusst ist, dass es solche verwendet, oder es dem System als selbstverständlich gilt, stets keine "reinen" Operatoren zu verwenden, sondern stets modifizierte. Das gebildete Handlungskonzept, berücksichtigt nicht diese Modifikationen, sondern speichert als Konzept das "Greifen" des Glases und das Abstellen an einen anderen Ort. Die vorige Kopplung an den Tisch wird vom System nicht berücksichtigt, da stets solche Kopplungen vorliegen. Der abstrakt eingeführte Motorische Kortex muss hierzu sämtliche modifizierte Operatoren lernen, damit das System in der Welt agieren kann.

Wurde beschrieben, wie quasi freie Objekte behandelt werden, so gibt es natürlich auch "feste" Objekte. Hiermit sind nicht nur feste Objekte gemeint wie die "Wand", die man nicht ohne weiteres aus der Verankerung, sprich der festen Kopplung mit anderen Wänden lösen kann, sondern auch zusammengesetzte Gebilde, wie weiter oben beschrieben. Eine zugeschraubte Flasche mit Inhalt, gilt zunächst für das System als Ganzes wie ein quasi freies Objekt. Will das kognitive System an den Inhalt der Flasche, so löst es das Gebilde in seine Einzelkonzepte auf: Die "Flasche" und den "Verschluss", die eine statische Relation eingehen. Das Aufschrauben des Verschlusses muss eine gelernte Handlung sein, die zum Inhalt hat ein Gebilde in quasi freie Objekte zu zerlegen. Konzeptionell hat das kognitive System es jetzt mit Konzepten von mehreren Objekten zu tun, die kein Gesamtkonzept mehr bilden.

Über die Idee der freien Objekte kann definiert werden, wie eine Hand-

lungskette in Blöcke oder Einzelhandlungen zerlegt werden kann. Eine "freie" Handlung sei so definiert, dass diese eine Operation an einem (quasi) freien Objekt ist. Das Ergebnis der Operation wird dabei wie weiter oben beschrieben (Abschnitt 7.3.4) durch statische Relationen beschrieben. Operiert das Subjekt an einem (komplexen) Modell, so kann diese Handlung im Regelfall nicht in Einzelhandlungen zerlegt werden, da Kopplungen innerhalb des Modells vorliegen. Da das Modell im Sinne der Wurzel eines Konzeptbaumes als Ganzes als ein freies Objekt behandelt werden kann, gilt die Gesamtoperation an dem Modell als ein Handlungsschritt. Freilich kann je nach Modell auch die Gesamtoperation im Spezialfall in Einzelhandlungen zerlegt werden. Z.B. kann die Operation sich mit Hilfe eines Toasters einen Toast zu machen in Einzelhandlungen zerlegt werden: Toast einstecken, Taster drücken, Toast entnehmen. Hierbei definiert das Input-Output Verhalten des Modells, ob eine solche Zerlegung möglich ist. Mithilfe der Idee der freien Handlung kann somit eine Handlungskette in Einzelhandlungen zerlegt werden.

7.4 Die Konzeptanalogie

Bei der Konzeptanalogie stehen die Konzepte in einem besonderen Verhältnis. Um ein Modell oder Lösungsstrategie aufzubauen, zieht das kognitive System ähnliche bereits gelernte Modelle oder Lösungsstrategien heran, um das Problem zu lösen. Es handelt sich also nicht um eine Subsumierung sondern die Konzepte stehen auf gleicher Stufe nebeneinander. Hierbei findet ein "Mapping" statt zwischen dem vorliegenden Konzept und dem herangezogenen Modell oder Lösungsstrategie. Beim Mapping muss berücksichtigt werden, dass nur Konzepte gleichen Typs aufeinander gemappt werden können. Das Mapping auf andere Konzepte stellt den einzigen Weg dar ein inneres Konzept aus der eigentlichen Schachtel zu lösen (siehe Abschnitt 7.2.4).

Um die Konzeptanalogie zu verstehen wird ein Beispiel betrachtet. Ein guter Schachspieler soll 50.000 aus-analysierte Stellungen im Kopf haben, um in einer aktuellen Stellung eine Lösungsstrategie entwickeln zu können. D.h., dass dieser 50.000 Lösungsstrategien heranzieht, um eine neue Lösungsstrategie aufzustellen. Hierbei verwirft er die meisten, da sich diese nicht auf die aktuelle Stellung mappen lassen. Die Lösungsstrategien die sich mappen lassen, werden daraufhin analysiert, ob diese für die aktuelle Situation relevant sind.

Auf dieselbe Weise soll ein kognitives System arbeiten, dass zum Aufbau eines Modells oder Lösungsstrategie im Prinzip alle gelernten Modelle oder

Lösungsstrategien heranzieht, von denen freilich die meisten sofort verworfen werden und dem System nicht bewusst werden. Dies zeigt ein weiteres mal die unglaubliche Leistung des Menschen, dass dieser alle gelernten Konzepte parallel prüfen kann, ob diese für die aktuelle Situation relevant sind. Prüft ein künstliches kognitives System diese gelernten Konzepte seriell, so wäre die benötigte Rechenleistung proportional zur Anzahl der gelernten Konzepte.

Das "Bauchgefühl" des Menschen, zu welchem Ergebnis eine Handlung führt, sei damit so erklärt, dass dieser eine gewaltige Anzahl an ähnlichen Konzepten heranzieht, welche danach gewichtet sind, wie relevant diese in dieser Situation sind.

7.5 Konkretisierung einer Lösungsstrategie

Dieser Punkt zielt darauf ab, dass ein kognitives System Probleme erst mit abstrahierten Konzepten löst. Um eine Lösungsstrategie umzusetzen, muss jedes Konzept wieder konkretisiert werden. D.h., dass jede Handlung, jede Interaktion von Objekten und die Objekte selbst auf konkret in der vorliegenden Szene vorhandene Möglichkeiten zurückgeführt werden müssen. Zu einem abstrakten Trinkgefäß muss z.B. ein Objekt in der Szene gefunden werden, welches diese Funktion einnehmen kann. Die Konkretisierung ist somit die Gegenoperation zur Abstraktion.

8 Der Masteralgorithmus

Der Masteralgorithmus steht über allen kognitiven Vorgängen und kann nicht vom System selbst beeinflusst werden. Ein kognitives System wurde so eingeführt, dass es aufgaben-zentriert arbeitet. Der Masteralgorithmus sortiert Aufgaben nach Relevanz und Dringlichkeit. Er bewertet zusätzlich welchen Benefit eine Aufgabe einbringt, oder den Nachteil eine Aufgabe nicht auszuführen. Der Masteralgorithmus kann technisch also durch eine Liste von Aufgaben dargestellt werden, wo jedem Eintrag die beschriebenen Attribute zugeordnet werden.

Das System entwickelt zwar durch Kognition die Aufgaben, wenn diese nicht durch das Urziel festgelegt sind, und bestimmt deren Relevanz, es hat aber keinen Einfluss auf die Ausführung des Masteralgorithmus selbst.

9 Grundantriebe eines kognitiven Agenten

Neben dem Urziel hat jedes kognitive System Grundantriebe, die es erst in Lage versetzt Aufgaben durchzuführen. Es muss zwei Dinge tun:

1. Es muss zunächst die Welt verstehen, in der es agiert.
2. Es muss seine Handlungsmöglichkeiten in dieser Welt herausfinden.

Diese beiden Punkte führen auf ein Lernen, welches sich über die gesamte "Lebensspanne" des kognitiven Systems ausdehnt. Der Masteralgorithmus ordnet Lernen als Aufgabe ein und bestimmt dessen Relevanz und Dringlichkeit. Er muss abwägen, ob es sinnvoller ist, erst weiter zu lernen, oder eine andere Aufgabe auszuführen.

10 Lernen

Auch wenn das Lernen ein Prozess ist, der sich über die gesamte "Lebensspanne" eines kognitiven Systems erstreckt, können neben dem Erlernen der Beherrschung des eigenen Körpers zwei anfängliche Phasen ausgemacht werden, die jedes natürliche kognitive System durchläuft.

10.1 Beherrschung des Körpers

Aufgrund von Gravitation, die eine bestimmte Richtung aufweist, führen verschiedene Lagen des Körpers bei gleichen Motorsignalen zu verschiedenen Effekten. Es wird angenommen, dass das kognitive System innere Sensoren hat, die die Lage des Körpers erfassen. Hierzu bedarf es keiner neuen grundsätzlichen Konzepte, der Körper wird als physisches Objekt wie jedes andere betrachtet, so dass die durchgeführte Aktion die Ausgangslage des Körpers mit erfassen muss und diese als Zustand der gesamten Szene mit beschreibt.

Es ist davon auszugehen, dass das kognitive System anfänglich zufällige Motorsignale erzeugt und dessen Wirkung auf die physische Welt beobachtet. Diese können noch nicht zu irgendwelchen Zielen führen, die dem Urziel dienen. Allgemein führen die anfänglichen spielerischen Lernprozesse auf keine Ziele, die irgendeine Relevanz haben. Hier müssen anfangs Ziele vorgegeben werden, die alleine die Beherrschung des eigenen Körpers und die Beherrschung des Umgangs mit der Umwelt zum Ziel haben. Die spielerische Phase

des Lernens sei durch diesen Charakter der Ziele definiert. Wobei die später folgende lebenslange Lernphase zum Ziel hat Aufgaben von Relevanz besser zu lösen.

10.2 Erste Lernphase

In der ersten Lernphase lernt das kognitive System nicht wie Handlungen aufeinander aufbauen. Es lernt alleine die dynamischen Interaktionen in der Welt. Es schmeißt, einem Kleinkind gleich, alle möglichen Gegenstände auf den Boden um zu sehen was passiert. In dieser Phase benimmt sich das kognitive System wie der Elefant im Porzellanladen. Kein Objekt ist vor ihm sicher. Es kümmert das kognitive System nicht, dass es durch seine Aktionen aus der Sicht eines Erwachsenen mehr Arbeit macht als vorher.

10.3 Zweite Lernphase

In der zweiten Lernphase geht man davon aus, dass das kognitive System genügend Konzepte über das dynamische Verhalten von Objekten gelernt hat, um anders spielerisch zu lernen. In dieser Phase lernt es, wie Handlungen aufeinander aufbauen, um ein Gesamtziel zu erreichen. Es baut Dinge wie Türme auf oder führt andere Handlungsketten aus, die zu einer Konstellation führen, die die Charakteristika eines Handlungsergebnisses haben. Auch wenn das erreichte Ziel ein spielerisch definiertes Ziel ist, lernt es, wie Handlungen und Ziele zusammengebracht werden.

10.4 Lebenslanges Lernen

Ist die spielerische Phase vorbei, schließt sich ein lebenslanges Lernen an, was zum Ziel hat relevante Aufgaben besser zu erfüllen. Hierzu abstrahiert das System Konzepte, subsumiert Konzepte unter andere und baut Modelle auf, alles zu einem besseren Verständnis der Welt, in der es agiert.

11 Phase III, die Reflexion und Steuerung mentaler Vorgänge

Das Denken über das Denken beinhaltet die Phase III des Aufbaus eines kognitiven Agenten, wo beschrieben wird, wie mentale Vorgänge Zwecks einer

Steuerung klassifiziert werden. Die Klassifizierung geht von der Beobachtung aus, dass mentale Vorgänge als eigene Handlungen betrachtet werden, so dass das Handlungsbewertungskonzept auf das Mentale übertragen wird. Auf diese Weise ergibt sich eine andere Behandlung der Metakognition, als diese klassisch üblich ist.

11.1 Klassische Metakognition

Die Beschreibung der klassischen Metakognition[6] geht von der "klassischen" sensomotorischen Schleife aus, wie diese kurz schon beschrieben wurde. Die sensomotorische Schleife repräsentiert eine geschlossene Schleife zwischen Sensorik und Motorik. Die Sensoren nehmen Information über die Umwelt wahr, welche nach einem festen Algorithmus in motorische Signale umgewandelt werden, die wiederum die Umwelt beeinflussen. Die geänderte Umwelt wird wieder von den Sensoren wahrgenommen und führen nach dem gleichem Algorithmus zu erneuten Motorsignalen. Tritt hierbei ein Problem auf, schaltet sich eine übergeordnete Schleife ein: Die Metakognitionsschleife. Diese überträgt einige Informationen aus der sensomotorischen Schleife in die übergeordnete Schleife. Zusätzlich werden Informationen über die Art des Problems in diese Schleife übertragen, so dass die Metakognitionsschleife eine Problemanalyse betreiben kann. Wurde das Problem gelöst, fließt die Information wieder zurück in die sensomotorische Schleife, wobei diese, entsprechend der Problemlösung, angepasst wird.

Etwas Wünschenswertes fehlt an dieser Vorgehensweise: Die kognitiven Vorgänge der Metakognitionsschleife können nicht wieder selbst zum Gegenstand höherer Kognition werden. Metakognition von Metakognition schließt sich bei dieser Anordnung der Schleifen aus. Darüber hinaus sind alle bekannten Ansätze dieser Art reine Symbolmanipulationsansätze. Diese beruhen auf der mathematischen Logik, genauer auf der Prädikatenlogik. D.h., dass die Metakognitionsschleife, nicht wie im Abschnitt 6 beschrieben, auf dem Inhalt der Symbole arbeitet. Im Prinzip ist dies ein prädikatenlogischer Ansatz auf dem ein weiterer prädikatenlogischer Ansatz aufgesetzt ist. Man könnte meinen, dass was der erste logische Ansatz nicht lösen kann, kann auch der aufgesetzte Ansatz nicht lösen. Der Unterschied ist hier, dass dem aufgesetzten Ansatz weitere Informationen zur Verfügung stehen: Informationen über die Art des Problems.

Nicht nur, dass keine Metakognition von Metakognition in diesen Ansätzen möglich ist, es kann über die Prädikatenlogik auch kein Bezug zu

einem Subjekt hergestellt werden, so dass diese Ansätze kein Modell für Ich-Bewusstsein liefern. Es könnte zwar ein Symbol für das "Ich" eingeführt werden, es hätte jedoch keinen Bezug zu den Vorgängen in der Metakognitionsschleife.

Der in diesem Buch beschriebene Ansatz stellt auf zwei Wegen einen Bezug zum Subjekt her. Wird das Hantieren mit Konzepten selbst wieder als Handlungsbewertungskonzept dargestellt, so taucht hier das Subjekt als pointerartiges Handlungssubjekt und als bewertendes betroffenes Subjekt auf. Ein mentales Handlungsbewertungskonzept wird wie jedes andere Konzept behandelt, so dass mit diesem wieder gearbeitet werden kann. Auf die Art ist Metakognition von Metakognition möglich.

11.2 Bewältigung von Aufgaben und der "mentale Bogen"

In diesem Abschnitt wird beschrieben, dass sich die Metakognition graduell entwickelt, d.h. dass mentale Handlungsbewertungskonzepte erst später in der Entwicklung eines kognitiven Agenten eine Rolle spielen. Die graduelle Entwicklung bedeutet nicht, dass nicht eindeutig zwischen physischen und mentalen Handlungsbewertungskonzepten unterschieden werden kann. Der Unterschied ergibt sich dadurch, auf was das Handlungsbewertungskonzept gerichtet ist.

Als erstes soll ein Zusammenhang beschrieben werden, der hier "mentaler Bogen" genannt werden soll. Ein kognitives System lernt aus der aktuellen Situation, um das Gelernte auf zukünftige Situationen anwenden zu können. Da der sensorische Input, oder die Situation im allgemeinen, nie nochmal dieselbe ist, ist es nötig, dass das kognitive System Konzepte abstrahiert. Vom konkreten Glas muss z.B. auf ein Trinkgefäß abstrahiert werden. Der mentale Bogen beschreibt die Schleife, dass vom Konkreten abstrahiert wird, die abstrahierten Konzepte auf abstrakten Niveau verknüpft werden, und diese verknüpften Konzepte wieder konkretisiert werden, um eine physische Handlung auszuführen. Die zugehörigen Konzeptoperationen wurden im Abschnitt 7 behandelt.

Es wird davon ausgegangen, dass ein kognitives System anfangs nur kleine Abstraktionen vornimmt, wie das obige Beispiel mit dem Glas. Diese Abstraktionen haben die Eigenschaft, dass die abstrahierten Konzepte noch direkt auf Vorgänge in der physischen Welt abgebildet werden können. Ent-

fernt man sich durch Abstraktion von der direkten Abbildbarkeit der Konzepte tritt ein Problem auf, welches einer Lösung bedarf. Ist die Zerlegung einer Szene in Konzepte zwar rechenintensiv, so ist diese doch in jedem Fall nach endlich vielen Schritten abgeschlossen. Ebenso verhält es sich, wenn das System in einer Szene nach der besten Lösungsstrategie sucht. Da die betrachtete Welt endlich ist, ist auch ihre Beschreibung durch Konzepte endlich. Dies ändert sich, wenn die Konzepte nicht mehr die konkrete Szene beschreiben, sondern abstrakte Konzepte sind, die Zwecks der Bildung von Modellen oder Lösungsstrategien, beliebig verknüpft werden. Hier tritt das sogenannte Halteproblem auf (siehe Abschnitt 11.6). Wann weiß das System, ob es Konzeptverknüpfungen weiter nachgehen soll, oder lieber einen anderen Weg sucht?

Die mentalen Prozesse, welche hier das alleinige Hantieren mit Konzepten sind, müssen gesteuert werden. Hierzu ist es nötig, dass mentale Prozesse klassifiziert werden und diese einer Bewertung zugeführt werden. Die Klassifizierung soll hier die Einordnung der mentalen Prozesse als Handlungsschemata leisten. Aber wie kann man einen mentalen Prozess bewerten? Dies soll über den mentalen Bogen geschehen. Ein mentaler Prozess hat für sich genommen keinen Wert. Erst die physische Ausführung einer "Idee" kann bewertet werden. D.h., dass der gesamte mentale Bogen eine Bewertung erhält, diese aber erst nach der Konkretisierung. Es soll an dieser Stelle eingeschoben werden, dass kognitive Agenten, die zur Kommunikation fähig sind, die Konkretisierung einer "Idee" auf einen anderen Agenten verlagern können. Aber auch hier gilt, dass eine "Idee" ohne Konkretisierung keinen Wert hat. Erhält ein mentaler Bogen oder erhalten viele mentale Bögen eine Bewertung, so kann durch die bereits eingeführte Blockbildung auch einzelnen Blöcken eine Bewertung zugewiesen werden, je nachdem an welchen mentalen Bögen ein Block beteiligt war. Um Blöcke zu bilden, ist es wie im physischen nötig, dass die gesamte Handlungskette in Einzelhandlungen zerlegt werden kann. Wie im physischen müssen dafür Konstellationen, die entstehen, die Eigenschaften von Zwischenergebnissen aufweisen.

Die Charakterisierung von mentalen Prozessen soll durch Handlungsschemata stattfinden. Hierzu wird das physische Handlungsbewertungskonzept ins Mentale übertragen. Auch die Rolle der anderen beiden Konzepttypen ändert sich hier leicht. Es wurde eine allgemeine Darstellung der drei Konzepttypen geliefert, bevor diese auf die physische Welt konkretisiert wurden:

1. Das Konzept einer in sich geschlossenen Sinneinheit.

2. Das Konzept der Relation zwischen zwei geschlossenen Sinneinheiten.

3. Das Handlungs-Bewertungskonzept.

Auf das Mentale übertragen ist die mentale Handlung das Operieren mit Konzepten. Dementsprechend ist die geschlossene Sinneinheit mit der operiert wird ein (vollständiges) Konzept. Die Relationen zwischen den Sinneinheiten sind allesamt statischer Natur. Dies widerspricht nicht dem, was über statische Relationen gesagt wurde. Diese werden entweder aus dynamischen Relationen abgeleitet oder sind das Ergebnis einer Handlung. Hier sind sie allesamt das Ergebnis von mentalen Handlungen. Wird z.B. ein Konzept abstrahiert, so stehen das konkrete und das abstrakte Konzept in einer statischen Relation zueinander, die sich durch die mentale Handlung des Abstrahierens ergibt.

Könnte jetzt durch Blockbildung das "Abstrahieren" allgemein als gut bewertet werden? Dies ist nach dem hier beschriebenen Aufbau nicht möglich. Davon abgesehen, dass ein kognitives System, welches das "Abstrahieren" als gut erkannt hat, wie wild nur noch nach Abstraktionen suchen würde, und nicht viel Nutzen hätte, verbietet sich dieses aus einem anderem Grund. Mit einer leeren Schablone können keine Operationen ausgeführt werden, so dass beim Abstrahieren stets ein Inhalt abstrahiert wird. Diese Abstraktion des Inhalts muss sich über den mentalen Bogen dann erst mal als sinnvoll erweisen. Des weiteren wird ein Handlungsbewertungskonzept von innen und außen charakterisiert. Während die innere Charakterisierung im Mentalen durch die benutzten Konzepte (geschlossene Sinneinheit) und den Operationen zwischen diesen (Relation zwischen zwei geschlossenen Sinneinheiten), festgelegt wird, arbeitet die äußere Charakterisierung mit einer Analyse der gesamten "Szene". Was die physische "Szene" übertragen auf den mentalen Bereich ist, wird im Folgenden beschrieben.

Besteht die äußere physische Welt für das kognitive System nur aus Objekten, wobei die dynamischen Relationen nicht direkt sichtbar sind, sondern sich aus der Dynamik der Objekte ableitet, so besteht der mentale Objektbereich nur aus Konzepten (mit statischen Relationen zwischen diesen). Überträgt man den Begriff der "Szene" ins mentale, so umfasst diese alle dem System vorliegenden Konzepte, die für eine Problemlösung relevant sein können. Es müssen nicht zwischen allen bereits eine Relation herrschen. Zur Analogiebildung trägt das System potentiell relevante Modelle und Problemlösungsstrategien zusammen. Erst wenn ein Mapping auf einer dieser

Modelle oder Lösungsstrategien gemacht wird, sind auch diese durch eine statische Relation verbunden.

Ein Handlungsbewertungskonzept operiert zunächst auf der gesamten Szene. Bezeichnet man die obige Beschreibung des mentalen Objektbereichs als "mentale Szene", so wird das Handlungsbewertungskonzept auf diese angewandt. Da nicht die ganze mentale Szene für das Handlungsbewertungskonzept relevant ist, tritt auch hier die bereits erwähnte sternförmige Analyse ein. Auf der einen Seite werden die kausalen Einflüsse auf die Konstellation analysiert, welche die Bedingungen eines Zwischenergebnisses erfüllt. Da es keine äußere Kausalität gibt, wird alle Kausalität dem pointerartigen Subjekt der mentalen Handlung zugeordnet. Bei natürlichen kognitiven Systemen laufen einige Konzeptoperationen automatisch ab, ohne dass das System diese willkürlich herbeigeführt hat. Z.B. werden viele Abstraktionen automatisch ausgeführt, ohne dass das System dies beeinflussen kann. Dem System liegen dann die Ergebnisse dieser automatisch durchgeführten Operationen vor. Ob automatisch durchgeführt oder willentlich gesteuert, beeinflusst nicht die Kausalitätsanalyse. Die Kausalitätsanalyse, welche beschreibt, welche Faktoren zu dem mentalen Handlungsergebnis geführt haben, berechnet, welche Konzeptoperationen Einfluss auf das mentale Handlungsergebnis hatten, und welche Konzepte, die die Situation zusätzlich beschreiben, das Handlungsergebnis erst möglich gemacht haben. Diese Analyse liefert den Kontext des Handlungsergebnisses und charakterisiert die mentale Handlung. Es wird also nicht einfach eine "Abstraktion" im allgemeinen als positiv bewertet, sondern genau die speziell durchgeführte Abstraktion in ihrem zugehörigen Kontext (siehe hierzu den Umsortierungsprozess der seriellen Informationsverarbeitung 11.4).

Der andere Teil der sternförmigen Analyse beinhalten auf welche Weise das System selbst oder andere Subjekte beeinflusst sind. Diese Bewertung kann nur über den mentalen Bogen geschehen, also einer potentiellen Konkretisierung. Freilich kann das System schon so viel gelernt haben, dass es bereits ein "Gefühl" dafür hat, was konkretisiert werden kann und was nicht. Um so mehr Wissen das System über Konkretisierungen hat, umso besser kann es rein mental entstandene "Ideen" bewerten.

Aufgaben werden zwar immer über den kompletten mentalen Bogen gelöst, die einzelnen Schritte, Abstraktion, abstrakte Verknüpfung und Konkretisierung, müssen nicht dabei nicht in einem Vorgang stattfinden. Ein kognitiven System neigt dazu Dinge zu abstrahieren, dessen Anwendung nicht sofort ersichtlich ist. D.h., dass die nötigen Abstraktionen schon vorliegen

können, bei der Lösung einer Aufgabe. Hat das System obendrein ein gutes Verständnis von möglichen Konkretisierungen, liegt der Schwerpunkt der mentalen Arbeit im Verknüpfen der abstrakten Konzepte.

Wann werden mentale Handlungskonzepte gebildet? Es wurde gesagt, dass sich entwickelnde kognitive Systeme am Anfang noch keine mentalen Handlungskonzepte bilden müssen, da ihr Denken noch direkt auf die physische Außenwelt abgebildet werden kann. In dem hier beschriebenen Modell wird allerdings davon ausgegangen, dass die mentale Seite der Konzeptoperationen von Anfang an mit betrachtet, bzw. kontrolliert wird. D.h., dass für spätere Zwecke der Anfangs "kleine" mentale Bogen bereits analysiert wird, um nach und nach Konstellationen als Zwischenergebnisse ausfindig zu machen und so eine Blockbildung einzuleiten. Wird der mentale Bogen ausgeprägter, können solche Blöcke, die für die Problemlösung entscheidend waren, in den Blickpunkt der Aufmerksamkeit geraten. Dies soll heißen, dass mentale Handlungskonzepte gebildet werden, je nachdem worauf die Aufmerksamkeit des Systems gerichtet ist. Erst wenn die Aufmerksamkeit auf mentale Vorgänge gerichtet wird, kann ein mentales Handlungskonzept gebildet werden.

Das System macht dabei einen klar definierten Sprung, in dem was betrachtet wird. Bei der Ausrichtung auf die Außenwelt, sind die elementaren Sinneinheiten die physischen Objekte. Bei mentalen Handlungskonzepten sind die elementaren Sinneinheiten (vollständige) Konzepte. Dies heißt, dass sich bei mentalen Handlungskonzepten der Charakter ändert was betrachtet wird. Es wird nicht mehr allein der Inhalt der Konzepte betrachtet, sondern zusätzlich ihre Eigenschaft als Konzept über die Welt. Der Inhalt ist nach wie vor wichtig, da auch mentale Handlungskonzepte nicht auf leeren Schablonen arbeiten. Neben ihrer Eigenschaft als Konzept, werden bei einem mentalen Handlungskonzept nicht dynamische Interaktionen von Objekten betrachtet, sondern die statischen Relationen zwischen den Konzepten.

Ist das kognitive System anfangs nur auf die physische Außenwelt gerichtet, gibt es bereits hier Konzeptoperationen, die genaugenommen keinen physischen Vorgang beschreiben. Es soll zwischen vertikalen und horizontalen Konzeptoperationen unterschieden werden. Vertikale Konzeptoperationen beschreiben keinen physischen Vorgang. Diese sind die Konzeptabstraktion (und Subsumierung) und die Konkretisierung. Bei diesen wird der Abstraktionsgrad entweder erhöht oder erniedrigt. Beschreibt ein Konzept die physische Welt, tut sie dies auf die gleiche Weise, wenn es abstrahiert wird. Horizontale Konzeptoperationen beschreiben Verknüpfungen von Konzepten

bei gleichbleibenden Abstraktionsniveau. Also die Bildung von Modellen und Lösungsstrategien. Egal wie weit ein Konzept abstrahiert wurde, es enthält abstrakte Repräsentanten in seiner Schablone, die noch mit der physischen oder mentalen Welt identifiziert werden können. Auf diese Weise können horizontale Konzeptoperationen mit Verknüpfungen im Betrachtungsgegenstand identifiziert werden.

Auch wenn jedes kognitive System, im Sinne des mentalen Bogens, auch am Anfang nicht auf vertikale Konzeptoperationen verzichten kann, führen diese noch nicht auf die Bildung von mentalen Handlungskonzepten. Bei kleinen Abstraktionen, wie der von einem Glas auf ein Trinkgefäß, liegt die Aufmerksamkeit noch vollständig auf der physischen Außenwelt. D.h. dass bei dem Ausgangskonzept "Glas" und dem abstrahierten Konzept "Trinkgefäß" nur auf den physischen Inhalt geachtet wird, nicht darauf, dass diese Operation selbst keinen physischen Vorgang beschreibt.

Es soll ein weiter Unterschied zwischen der hier beschriebenen Metakognition und der Klassischen erwähnt werden. Springt bei der klassischen Metakognition die Metakognitionsschleife genau dann an, wenn ein nicht lösbares Problem vorliegt, so wird ein solches Problem hier anders interpretiert. Ein Problem, was etwa die Ausführbarkeit einer Aktion betrifft, kann nur auftreten, wenn das System zu sehr abstrahiert hat oder falsch subsumiert hat. Es liegt eine Eigenschaft der vorliegenden Objekte oder Objektinteraktionen vor, über die bei der Abstraktion hinweggegangen wurde. Die Lösung in diesem Fall ist eine Differenzierung, die die Konzepte bezüglich dieser Eigenschaft differenzieren. Wenn nicht weitere Differenzierungen nötig sind, kann jetzt eine geeignete Handlung ausgeführt werden.

Des weiteren ist die Metakontrolle der Kognition in dem hier beschriebenen Modell stets aktiv. Es wird fortwährend nach Konstellationen gesucht, die ein mentales Zwischenergebnis repräsentieren können, um mentale Handlungsblöcke zu bilden. Auch wenn die Aufmerksamkeit nicht auf der Metakognition liegt, bewertet diese die gerade stattfindende Kognition ständig durch Subsumierung unter bereits gebildete mentale Handlungs-Bewertungskonzepte. Ein konkretes, neues mentales Handlungskonzept wird allerdings erst gebildet, wenn die Aufmerksamkeit auf die Metakognition gerichtet ist.

11.3 Die Übertragung von Gelerntem auf neue Situationen

Die Hauptaufgabe der Kognition ist die Übertragung von Gelerntem in einer Situation auf neue Situationen. Dies soll allgemein die Aufgabe des mentalen Bogens sein. Hierbei wird abstrahiert und abstrakt verknüpft. Das Abstrahieren geschieht dabei über die Eigenschaften, die den Konzepten zugeordnet werden. Dieser Punkt soll hier unter der Berücksichtigung des Abschnitts über Konzeptverknüpfung (siehe 7.3) genauer betrachtet werden. Eigenschaften haben zwei Funktionen, wobei die erste der letzteren dient. Sie ordnen einzelne Konzepttypen in eine Hierarchie, so können z.B. physische Objekte je nachdem in welchem Abstraktionsverhältnis die Objekte stehen, in Ober- und Unterkonzepte eingeordnet werden. Abstraktionen entstehen, wenn gleiche Konzepttypen miteinander verglichen werden. Bildet man durch Abstraktion ein Oberkonzept, so dient dies der Annahme, dass alle Vertreter dieses Oberkonzepts in einer Handlung die gleiche Funktion einnehmen, wie z.B. die Abstraktion von einem "Glas" und einer "Tasse", die aufgrund ihrer Form ein Oberkonzept bilden, dass die Handlung "Trinken" erlaubt, und somit ihren Namen "Trinkgefäß" erhalten.

Es wurden zwei Bäume eingeführt. Der erste Baum beschreibt, dass zu einer einzelnen Handlung mehrere Relationen passen, um diese auszuführen, und zu einer einzelnen Relation passen mehrere Objekte, die in diese Relation eingesetzt werden können. Abstrahiert man innerhalb diesen Baumes, so behalten die abstrahierten Objekte und Relationen ihre Funktion bei einer Handlung bei. Der einzige Weg ein Objekt oder Relation aus dessen ursprünglichen Funktion zu befreien soll die Analogiebildung sein.

Die Analogiebildung vergleicht Modelle und Lösungsstrategien miteinander. Einer Handlung wird ein Modell des zu manipulierenden Sachbereichs gegenübergestellt. Das Modell bildet nach Abschnitt 7.3.1 eine Baumstruktur von Konzepten, wobei die Wurzel dieses Baumes den Konzepttyp des gesamten Gebildes definiert. Eine Lösungsstrategie besteht nun darin geeignete Schnittstellen zu diesem Modell zu finden. Hierbei kann das kognitive System das komplette Gebilde als ein Konzepttyp manipulieren, z.B. greift es eine zugeschraubte Flasche mit Inhalt, wobei das System die untergeordneten Konzepte des Konzeptbaumes nicht berücksichtigt. Oder es wählt eine Schnittstelle zu dem Gesamtgebilde, welche die untergeordneten Konzepte manipuliert, es kann z.B. die Flasche aufschrauben und verändert dabei den Konzeptbaum des Gebildes: Es liegen nun getrennte Modelle vor, einmal die

offene Flasche mit Inhalt und einmal das Objekt Deckel.

Das Paar, vorliegendes Modell plus Lösungsstrategie, hat eine logische Struktur, welche sich dadurch definiert, wie einerseits das Modell aus Konzepten aufgebaut ist, und andererseits wie die Handlungen den Konzeptbaum des Modells verändern. Diese logische Struktur ist die Grundlage für eine Analogiebildung. Bildet man eine Analogie, so haben beide Seiten dieselbe logische Struktur. Es wurde behauptet, dass mit leeren Schablonen nicht gearbeitet werden kann, so soll dies auch hier sein. Die reine logische Struktur ist eine Beschreibung, wie leere Schablonen miteinander verknüpft sind. Bildet das System eine Analogie, ist zwar die Voraussetzung, dass die logischen Strukturen übereinstimmen, die Schablonen sind aber mit Repräsentanten gefüllt. Liegt dem System ein zu manipulierendes Modell vor, und will das System die Aufgabe lösen, indem es eine Analogie bildet, so vergleicht es die Repräsentanten, die aufgrund der logischen Struktur aufeinander gemappt werden. Der herangezogene Part der Analogie, der bereits gelerntes Wissen darstellt, funktioniert, da die Repräsentanten in den jeweiligen Konzepten die passenden Eigenschaften besitzen, um diese so zu verknüpfen. Das System muss also beim Mapping überprüfen, ob die vorliegenden Repräsentanten des vorliegenden Modells, solche Eigenschaften besitzen, dass diese auf analoge Weise so verknüpft werden können, und dass die entsprechenden Handlungen an dem Modell in gleicher Weise ausgeführt werden können. Hierzu bedient es sich der subsymbolischen Simulation und der Eigenschaften der vorliegenden Konzepte, ob diese der logischen Struktur nach genauso verknüpft werden können. Hat das System also gelernt, wie man eine Flasche öffnen kann, um an den Inhalt zu kommen, kann es durch Analogiebildung, verschiedene Objekte öffnen, um an den Inhalt zu kommen.

Man kann somit Gelerntes auf neue Situationen übertragen, indem man eine Abstraktion innerhalb des zuerst genannten Baumes durchführt, und Objekte und Relationen bei der selben Handlung durch andere ersetzt oder man führt eine Analogie aus, und überträgt dabei die logische Struktur, wenn die gemappten Repräsentanten auf die gleiche Weise verknüpft werden können.

11.4 Allgemeine Vorgehensweise der passiven und aktiven Kognitionskontrolle

Es wird behauptet, dass die Arbeit mit Konzepten genau der Vorgang ist, der die serielle Informationsverarbeitung beim Menschen ausmacht. Hier werden Konzepte verknüpft und Lösungsstrategien entwickelt. Hierbei laufen Vorgänge, wie das Abstrahieren von Konzepten ohne willentliches Zutun ab. Der bewussten Wahrnehmung, welche hier die serielle Informationsverarbeitung ist, werden die Ergebnisse solcher automatischen Vorgänge zugespielt. Auch wenn nicht willentlich beeinflusst, machen solche automatischen Vorgänge ebenfalls einen Teil des gebildeten mentalen Handlungsbewertungs-Konzepts aus, vor allem wenn diese genau der entscheidende Schritt waren.

Die serielle Informationsverarbeitung hat die Eigenschaft, dass der Weg, den diese geht, nicht im Vorhinein feststeht. Das aktuell Berechnete bestimmt das weitere Vorgehen. Dies ist der aktive Teil der Kognitionskontrolle und soll genau mit dem "Gefühl" übereinstimmen, dass wir selbst bestimmen können in welche Richtung unsere Gedanken laufen, also unseren "freien Willen" bezeichnen. Da keine Theorie des freien Willens behaupten kann, dass dieser völlig willkürlich abläuft, müssen Prozesse diesen steuern. Diese Steuerung soll gerade die aktive Kognitionskontrolle sein. Die passive Kognitionskontrolle ist hierbei die stets vorhandene Bewertung von Kognitionsprozessen, was die Subsumierung unter mentale Handlungs-Bewertungs-Konzepte leisten soll. Sie sorgt dafür, dass ein Gedankengang der positiv bewertet wird, weiter ausgeführt wird, während ein negativ bewerteter als nutzlos eingestellt wird.

Liegt eine zu bearbeitende Aufgabe vor, so ist die übergeordnete Vorgehensweise fest. Vom zu manipulierenden Sachverhalt muss ein Modell erstellt werden, und danach eine Lösungsstrategie, welche die Handlungsschnittstellen zu dem Modell vom Sachbereich benutzt. Hierbei arbeitet man sich vom "Groben" ins "Kleine" vor. D.h., dass zuerst ein möglichst abstrahiertes Modell vorliegt, mit einer abstrahierten Lösungsstrategie, welche nicht alle Feinheiten mitberücksichtigt. Von diesen ausgehend werden die beiden Teile immer konkreter und es treten immer mehr Feinheiten zu tage, z.B. wird ein zu manipulierendes Gebilde in seine Einzelkonzepte zerlegt oder eine abstrakte Eigenschaft eines Konzepts konkretisiert.

Die serielle Informationsverarbeitung "schlängelt" sich dabei durch diesen Überbau hindurch. D.h., dass diese solche Teile der Aufgabe zuerst betrachtet, welche am "unklarsten" sind. Dass die serielle Verarbeitung hierbei nicht

den Überbau verlässt, dafür sorgt die passive Kognitionskontrolle. Sind die unklaren Punkte geklärt, simuliert die serielle Informationsverarbeitung die gesamte Handlungskette und betrachtet dabei auch die weniger interessanten Handlungsschritte, bevor diese ausgeführt werden.

Treten hierbei mentale Prozesse auf, die ungewöhnlich sind, aber trotzdem oder gerade zu einer Lösung führen, richtet sich die Aufmerksamkeit auf diese und ein neues mentales Handlungskonzept wird gebildet. Da die serielle Informationsverarbeitung teils konfus durch die Teile der Aufgabenbewältigung durchgegangen ist, ist es hierzu nötig, dass die mentalen Prozesse, die zur Lösung geführt haben neu sortiert werden. Die Sortierung erfolgt im Sinne einer Kausalitätsanalyse, indem geprüft wird, welche mentalen Handlungsschritte kausal zum Ergebnis geführt haben. Erst die sortierten mentalen Prozesse, in denen die entscheidenden Schritte hervorgehoben sind, werden als mentales Handlungskonzept abgespeichert und dient fortan der passiven Kognitionskontrolle.

11.5 Zusätzliche mentale Begrifflichkeiten?

Es soll an dieser Stelle eine kleine Zwischenbilanz gezogen werden, bezüglich der Eingangsfrage, dass hier nicht die Sprache untersucht wird, sondern die Konzepte, die mit der Sprache übermittelt werden. Ein zentrales Thema dabei ist, was im zweiten Abschnitt über das Bewusstsein und Ich-Bewusstsein (Abschnitt 11.7) eine große Rolle spielt, dass verschiedene Begrifflichkeiten den gleichen strukturellen Kern bezeichnen. Es wird behauptet, dass das, was man strukturell aussagen kann, von den aufeinander aufbauenden Phasen I-V vollständig erfasst wird. Hierbei ist einige begriffliche Arbeit notwendig gewesen, um den strukturellen Kern des Begriffs "Ich-Bewusstsein" herauszuarbeiten. Einige Begriffe im Zusammenhang mit dem "Subjekt" stellen sich dabei als strukturell inhaltslos heraus.

In der aktuell beschriebenen Phase III, können schon einige Begrifflichkeiten oder Charakteristika über mentale Zustände begründet werden. Alle noch nicht erfassten Begrifflichkeiten über das Subjekt oder das Mentale, welche strukturell nicht inhaltslos sind, liefert die Phase IV, wo die verschiedenen Subjektbegriffe zu einem Selbstmodell zusammengefügt werden, und die Phase V über die sprachliche Interaktion mit anderen Subjekten.

11.6 Welche Konzeptoperationen und das Halteproblem

Es wurden zwar die möglichen Konzeptoperationen vollständig aufgelistet (Abschnitt 7), es ist jedoch nicht klar, welche zur Lösung einer Aufgabe ausgeführt werden müssen. Im Prinzip können Konzepte unendlich lange verknüpft werden, was zu einem Halteproblem führt. Dies macht eine Metakontrolle der Kognition nötig, was die hier beschriebenen mentalen Handlungsbewertungskonzepte leisten sollen. Diese entstehen, wie beschrieben, durch Blockbildung von mentalen Handlungen. Hat man eine Bewertung von mentalen Handlungsblöcken, kann entschieden werden, ob ein "Gedanke" weiter ausgeführt werden soll, da dieser unter positiv bewerteten Blöcken subsumiert werden kann, und somit vielversprechend ist, oder ob der "Gedanke" verworfen werden soll.

Da die mentale Blockbildung sich erst nach und nach entwickelt, benötigt man eine anfängliche Vorgehensweise, wie Aufgaben bewältigt werden sollen. Es wurde beschrieben, dass bei einem anfänglichen kognitiven System das Halteproblem noch keine Rolle spielen soll, da dieses sich durch Abstraktionen nicht von einer direkten Abbildbarkeit auf die physische Welt entfernt. Da die physische Welt endlich ist, ist auch die Beschreibung derselben durch "direkte" Konzepte endlich.

Die anfängliche Vorgehensweise muss die Möglichkeit zulassen, dass nach und nach umfangreichere mentale Bögen in die Aufgabenlösung eingebunden werden. Freilich erst, wenn genügend Konzepte von mentalen Vorgängen vorliegen, so dass diese bewertet werden können.

11.7 Bewusstsein und Ich-Bewusstsein (Teil 2)

Das Buch wurde so eingeleitet, dass das hier beschriebene Konzeptmodell umgekehrt entwickelt wurde. Zuerst waren die Überlegungen zum Ich-Bewusstsein, zum Subjekt und zur Reflexion von Gedanken da. Von da ausgehend wurde nach den elementaren Bausteinen des Denkens und der Wahrnehmung gesucht, um aus diesen wieder Ich-Bewusstsein zusammenzusetzen. Hierbei war der Begriff des Konzepts hilfreich, der nicht nur Pate stand für die Ausarbeitung der Schablonen, die über die Welt gelegt werden, sondern vor allem Konzeptoperationen liefert, die die Kognition gut abbilden.

Das Ich-Bewusstsein wird gemeinhin als das Wissen über das eigene "Ich" beschrieben. Diese und ähnliche Aussagen über das Ich-Bewusstsein wurden begrifflich "Hin und Her geschubst" bis ein Kern gefunden wurde, der eine

strukturelle Aussage macht. Der Begriff "Ich" alleine ist strukturell inhaltslos. Es geht darum, wie das "Ich" mit objektiv Gegebenen verknüpft ist. Schaut man sich diese Verknüpfungen an, können zwei logisch verschiedene Verknüpfungen ausgemacht werden: Das Subjekt operiert auf einem Objektbereich oder das Subjekt betrachtet einen Objektbereich.

Den ersteren Bezug kann man direkt mit den hier beschriebenen physischen und mentalen Handlungskonzepten beschreiben. Der zweite, das "Betrachten", ist schwieriger zu knacken. Als erstes fällt auf, das im Gegensatz zur Handlung, das Betrachten nicht weiter charakterisiert werden kann: Es liegt ein Objektbereich vor, dem ein Subjekt zugeordnet wird. Das Betrachten hat weiter keine Eigenschaften. Ein solch inhaltsloser Bezug sollte nicht möglich sein. Die Lösung war, dass das "Betrachten" aus dem "Bewerten" abgeleitet ist. Das Bewerten einer Situation stellt einen charakterisierbaren Bezug vom Objektbereich zum Subjekt her. Überlegt man sich dann noch, dass Betrachten ohne Bewerten eigentlich nicht möglich ist, ergibt sich folgende Beziehung zwischen den beiden Begriffen: Das Bewerten ist der Grundbegriff aus dem das "Betrachten" durch Abstraktion gewonnen wird. Eine Betrachtung ist eine Bewertung einer Situation, wobei die Bewertung in der Beschreibung irrelevant ist. Man abstrahiert also das Bewertungskonzept dahingehend, dass keine konkrete Bewertung mehr vorhanden ist. Dies hinterlässt eine Leerstelle in der Bewertungsschablone, die einfach mit etwas Unkonkreten gefüllt wird. D.h., das beim "Betrachten" eine Bewertung des Betrachteten ständig mitschwingt, aber nicht konkretisiert wird.

Ein weiterer Fortschritt wurde dadurch erzielt, die beiden Begriffe "Bewusstsein" und "Ich-Bewusstsein" voneinander zu trennen. Bewusst ist das, worauf gerade die Aufmerksamkeit gerichtet ist, und beschreibt somit die serielle Verarbeitung beim Menschen. "Das Wissen über das eigene Ich" ist dabei der Vorgang, dass ein Subjektbezug gerade im Zentrum der Aufmerksamkeit liegt. "Wissen" das etwas so ist, ist dabei die Bewusstwerdung des Inhalts, d.h., das auf diesen Inhalt die Aufmerksamkeit gerichtet ist, und mit den dahinter stehenden Konzepten gerade gearbeitet wird.

Die Subjektbezüge beschreiben den Subjektbegriff noch nicht vollständig. Erst wenn die logisch verschiedenen Subjektbezüge zu einem konsistenten Selbstmodell zusammengeführt werden, ist der Subjektbegriff vollständig. Hierbei sind die einzelnen Subjektbezüge die elementaren Einheiten eines solchen Selbstmodells (siehe Abschnitt 13).

Kommen wir nun zur "Reflexion von Gedanken". Die Grundüberlegung hier fußt auf der Beobachtung, dass ein Gedankengang doppelt "beobachtet"

wird. Einmal wird der Inhalt des Gedankens betrachtet und einmal gibt es eine Metabetrachtung, die diesen Gedanken einordnet. Je nachdem worauf die Aufmerksamkeit liegt, ist einmal der Inhalt explizit repräsentiert und einmal der Gedankengang. Dies findet hier eine Lösung durch mentale Handlungskonzepte. Auch hier wurde unterschieden, ob der Inhalt der Kognition gerade betrachtet wird, oder die ständig aktive Bewertung der Kognition. Wobei ein Handlungsbewertungskonzept gebildet wird, wenn diese Kontrollebene in die Aufmerksamkeit rutscht. Wird ein mentales Handlungskonzept gebildet, wird explizit ein Bezug zum Subjekt gebildet: Das mentale Handlungssubjekt und das Subjekt, was die mentale Handlung bewertet.

Da mentale Handlungsbewertungskonzepte als normale Konzepte behandelt werden, ist prinzipiell Metakognition von Metakognition möglich. Obwohl dies ein Potential ist, welches jedes kognitive System haben sollte, soll gesagt sein, dass dieser Fall beim Menschen selten auftritt, und nicht ohne weiteres herbeigeführt werden kann. Wurde ein Bezug zum Subjekt gebildet, im Sinne einer Metakognition, muss von dieser ausgehend ein weiterer Bezug zum Subjekt gebildet werden, welcher sich vom Ersten unterscheidet. Übersetzt auf das Konzeptmodell heißt dies, das ein mentales Handlungskonzept weiter verarbeitet werden muss. Dies kann z.B. die Subsumierung dieses Handlungskonzepts sein. Ist die Aufmerksamkeit entsprechend gerichtet, kann dieser Verarbeitungsschritt wieder in ein neues mentales Handlungskonzept überführt werden, dieses bildet dann einen erneuten Bezug zum Subjekt, der sich von dem vorigen unterscheidet. Die Voraussetzung von Metakognition von Metakognition ist also die Weiterverarbeitung von mentalen Handlungskonzepten und eine entsprechend gerichtete Aufmerksamkeit.

Weiter soll noch ein berühmter Ausspruch von Descartes in diesem Modell untersucht werden: "Ich denke, also bin ich." Es soll argumentiert werden, dass die Denkleistung, die hinter diesem Ausspruch steckt, keine Basisleistung des Gehirns ist. Eine Basisleistung des Gehirns ist nach dem hier vorgestellten Modell die "Selbstverobjektivierung" des Subjekts. Das kognitive System bildet mentale Handlungsbewertungskonzepte und macht so gedankliche Vorgänge und den Bezug zum Subjekt explizit. So macht sich das Subjekt selbst wieder zum Objekt der Betrachtung. Descartes setzt nun dieses betrachtende Subjekt in den Mittelpunkt und stellt diesem ALLES Objektive gegenüber. Weiter spekuliert er, dass dieses Objektive ein vorgegaukelter Schwindel sein könnte. Woran aber nicht zu Zweifeln sei, ist, dass das Subjekt Denkvorgänge ausführt, so dass an der Existenz des Subjekts selbst nicht gezweifelt werden kann.

Das Ich-Bewusstsein wurde in diesem Modell begrifflich vom Bewusstsein getrennt. Während das Ich-Bewusstsein die Verknüpfung mit dem Subjekt beschreibt, ist das Bewusstsein eine Verarbeitungsstrategie. Das Bewusstsein ist, wie bereits im Abschnitt 2 beschrieben, gleichzusetzen mit der Aufmerksamkeitssteuerung. Nur dass, worauf die Aufmerksamkeit gerichtet ist, ist bewusst und wird vom Gehirn explizit repräsentiert. Es wird behauptet, dass nur in dieser expliziten Repräsentation mit Konzepten gearbeitet werden kann, der zusätzlich ein Arbeitsgedächtnis zur Verfügung steht, in dem Konzepte vorgehalten werden, die gerade erzeugt wurden, oder Konzepte herangezogen wurden, die zur Verknüpfung bereitstehen.

Die Aufmerksamkeitssteuerung sorgt für eine serielle Verarbeitung von Konzepten. Dies ist aus rechen-technischen Gründen sinnvoll. Bei einer seriellen Verarbeitung wird nur ein oder eine kleine Anzahl an Konzepten betrachtet, die für die gerade bearbeitete Aufgabe wichtig sind. Das menschliche Gehirn ist imstande parallel alle gespeicherten Konzepte zu prüfen, ob diese mit den gerade betrachteten verknüpft werden können. Verknüpfen soll in diesem Fall heißen, dass eine Analogie, eine Subsumierung etc. ausgeführt werden kann. Kann bei einem künstlichen kognitiven System diese Prüfung aller gespeicherten Konzepte nicht parallel ausgeführt werden, so ist der rechen-technische Zeitaufwand proportional zur Anzahl der gespeicherten Konzepte. Gäbe es nicht die zentralisierte serielle Verarbeitung und würden einfach alle Konzepte probeweise mit allen Konzepten verknüpft, wäre der Rechenaufwand proportional zum Quadrat der gespeicherten Konzepte. Abgesehen davon, dass es Zwecks einer Aufgabenbewältigung Sinn macht, nur wenige relevante Konzepte zu betrachten, und alle übrigen Konzepte potentiell hinzuzuziehen, wird vermutet, dass das menschliche Gehirn zwar eine parallele Prüfung aller Konzepte entwickelt hat, aber keine Vorgehensweise ALLE Konzepte mit ALLEN parallel zu verknüpfen.

Auch wenn behauptet wird, dass die Konzeptarbeit nur im Bewussten stattfindet, wird vermutet, dass es beim Menschen einen Vorgang gibt, der Abseits der Aufmerksamkeit stattfindet und an Konzepten arbeitet. Dies ist die Neubewertung von Situationen. D.h., dass ohne Aufmerksamkeit sämtliche oder nur aktuelle Bewertungskonzepte bearbeitet werden. Dies lässt den Rechenaufwand ebenfalls nicht quadratisch explodieren, da immer nur eine Bewertung zur Zeit nochmal überarbeitet wird. Die Überarbeitung ist auch nicht eine beliebige Verknüpfung zur Lösung einer Aufgabe, wie diese bewusst stattfindet, sondern kennt nur eine Richtung: Wie beeinflussen die erlebten Situation das Subjekt wirklich. D.h., dass nochmal die Relevanz der

Beeinflussungen neu bewertet werden.

In diesem Buch werden in erster Linie die Strukturen aufgeklärt, in denen ein kognitives System denkt. Dies klärt nur zum Teil oder legt es teilweise nur nahe, wie die Denkvorgänge gesteuert werden sollen. Wurde erst mal offen gelassen, wie der Algorithmus aussieht, der ein System anfänglich steuert und immer größere mentale Bögen zulässt, so spielt die Aufmerksamkeitssteuerung hier eine zentrale Rolle. Auch hier sollen nur ein paar Merkmale erwähnt werden. Beim Menschen liegt laut psychologischen Experimenten eine hierarchische Aufmerksamkeitssteuerung vor (siehe hierzu die "Filtertheorie der Aufmerksamkeit" von Donald Eric Broadbent). D.h., dass auf verschiedenen Ebenen der Wahrnehmung "Hinweise" gesetzt werden, worauf der Mensch seine Aufmerksamkeit lenken soll. So wird in den ersten Verarbeitungsschichten der Wahrnehmung schon die Aufmerksamkeit gerichtet, so dass ein ungewöhnlicher Gegenstand im Gegensatz zu einem Bekannten seine Aufmerksamkeit auf sich zieht. Dies führt z.B. zu dem Effekt, dass wissenschaftliche Vorträge nicht alleine, oder nur sehr selten nach ihrem Inhalt bewertet werden. Der Vortragende hat entscheidenden Einfluss darauf, ob die Zuhörer fast einschlafen oder gebannt zuhören.

Ist es für die anfängliche Lernphase von Vorteil, dass ungewöhnliche Gegenstände mehr Aufmerksamkeit erhalten, als bereits Bekannte, so kann man überlegen, ob die Aufmerksamkeitssteuerung bei künstlichen kognitiven Agenten lieber ausschließlich von der höchsten Verarbeitungsebene ausgehen sollte. Abgesehen davon, wie diese Steuerung geleitet werden kann, könnte eine solche Steuerung die Effektivität des Systems erhöhen.

11.8 Bewertung ohne "Ich"

Bei einem künstlichen kognitiven System wäre es möglich, dass es die Bewertungen einer Situation nicht auf das eigene Subjekt bezieht, d.h. dass es nicht analysiert wie das eigene Subjekt betroffen ist, sondern ihm vorgegebene Ziele. Dies ist nur möglich, da Bewertungen bei einem künstlichen System explizit vorgenommen werden. Zur Erinnerung: Gefühle sind implizite Bewertungen, wohin gehend explizite Bewertungen genau beschreiben, welche Eigenschaften des Systems betroffen sind. Implizite Bewertungen haben demzufolge immer einen Ich-Bezug.

Ganz ohne Ich-Bezug geht es bei einem künstlichen System ebenfalls nicht, da die anfängliche Lernphase, oder das folgende lebenslange Lernen, die Eigenschaften des Systems beeinflussen. Ebenfalls sollte das System einen

Selbsterhaltungstrieb haben: Es müsste Beschädigungen als negativ einstufen, damit es sich einsatzfähig hält.

Selbst wenn man das System völlig selbstlos programmiert, würde dem Modell zufolge, das System Ich-Bewusstsein entwickeln. Es würde zwar alle Situationen nach fremdbestimmten Zielen bewerten, es würde sich aber einerseits selbst als handelndes Subjekt beschreiben und andererseits über das Handlungsbewertungskonzept sich als bewertendes Subjekt wahrnehmen.

Das System braucht darüber hinaus das später beschriebene Moralmodul, um zu bewerten, wie die jeweiligen Subjekte von den eigenen Handlungen betroffen sind. Dieses kann völlig egoistisch, altruistisch oder auf etwas dazwischen eingestellt werden. Gerade dieses Moralmodul und dessen Möglichkeiten sollten hinreichend diskutiert werden, bevor künstliche kognitive Agenten mit Bewusstsein erschaffen werden.

Aus militärischer Sicht, könnte man auf die Idee kommen, das Moralmodul ganz wegzulassen, damit es die Befehle nicht hinterfragt. Dies wäre gefährlich, da das System dann zwar vorgegebene Ziele hat, aber keine Möglichkeit zu berechnen, wie Subjekte betroffen sind. Dies könnte für alle Seiten ungeahnte Effekte haben, wenn die Maschine mit allen Mitteln das Ziel verfolgt. Ein Moralmodul könnte andererseits Maschinen für den Krieg unbrauchbar machen, da ein Grundantrieb von diesen wäre, die Welt zu verstehen, in der sie agieren. Sie könnten die politische Gesamtsituation anders einschätzen als ihre Befehlsgeber, und es nicht für nötig erachten, dass Menschen getötet werden müssen.

12 Mathematik, Physik und die philosophische Begründung der drei Konzepttypen

Wäre ein künstliches kognitives System, dass nach dem hier vorgestellten Modell programmiert ist, automatisch gut in Mathematik, da das Modell mathematischer Natur ist? Dies wäre zu verneinen. Das künstliche System müsste sich Mathematik wie der Mensch aneignen, indem es von Konzepten abstrahiert, die es durch Interaktion mit der Welt gelernt hat.

Das Handlungskonzept wurde, um es zu verdeutlichen, von der logischen Struktur mit einem mathematischen Operator verglichen. Was war aber aus philosophischer Sicht zuerst da, dass Verständnis des mathematischen Operators oder dass Verständnis von der logischen Struktur einer Handlung?

Lernt das kognitive System Mathematik durch Abstraktion von natürlichen Konzepten, war das letztgenannte Verständnis zuerst da. Es wird sogar argumentiert, dass die Idee eines Operators erst durch kognitive Systeme entstanden ist, welche Handlungen in der Welt ausführen, da physikalisch gesehen, nur Teilchen und ihre Wechselwirkungen existieren. Die Operation auf einem Objektbereich ist also eine Beschreibungsart, die alleine kognitive Systeme benutzen, um das Verhältnis zwischen sich und der Welt besser einordnen zu können. Damit wäre der mathematische Operator, wie bereits weiter oben beschrieben, ein "de-personalisiertes" Handlungskonzept. Würde ein kognitives System in der Mathematik nicht zu einem de-personalisierten Handlungskonzept greifen, müsste es die Mathematik alleine durch Interaktion von Objekten beschreiben. Hierdurch würde die Mathematik nicht eingeschränkt werden. Eine Matrix, die wir als Operator verstehen, die auf einem Vektor wirkt und einen Vektor als Operationsergebnis hat, würden wir als Interaktion zwischen dem Objekt Matrix und dem Objekt Vektor einordnen, welches als Interaktionsergebnis wieder einen Vektor hat.

Ein anderes philosophisches Problem ergibt sich aus der Beschreibung der Welt als Objekte mit Eigenschaften, die miteinander interagieren. Ist es hier die Physik, die ein solches Denken vorgibt, oder ist dieses Denken jedem kognitiven System zu eigen, und wir übertragen es auf die Physik?

13 Phase IV, Metzingers Selbstmodelltheorie

Der Philosoph Thomas Metzinger hat eine Theorie der Selbstwahrnehmung[11, 10] aufgestellt, indem er postuliert, dass der Mensch, und zu einem gewissen Grad auch Tiere, ein Selbstmodell aufbauen. Dachte der Autor dieses Buches zuvor, dass Bewusstsein bzw. Ich-Bewusstsein ein Graben ist, den künstliche Systeme entweder nie oder nur in einem qualitativ großen Sprung überwinden könnten, zeigen die Ideen von Metzinger, dass man sich dem Thema Ich-Bewusstsein rational nähern kann, und dies sogar stückweise. Es war die Verknüpfung von den im Abschnitt 11.7 vorgestellten philosophischen Ideen zum Ich-Bewusstsein mit den Ideen von Metzingers Selbstmodelltheorie, die das hier vorgestellte Modell haben entstehen lassen.

Die Ideen aus Abschnitt 11.7 sind aus der Beschäftigung mit der mathematischen Logik entstanden, indem der Autor die in der Logik verwendete Selbstreferenz analysierte. Einerseits im Bemühen die Selbstreferenz in der Logik zu verstehen, hat sich ein Verständnis aufgebaut, wie der menschliche

Geist sich selbst reflektiert. Diese Überlegungen zum menschlichen Geist waren rein theoretischer Natur und bezogen sich nicht auf konkrete Vorgänge im Gehirn. Hingegen ist Metzingers Selbstmodelltheorie sehr konkret und beschreibt die Bildung eines Selbstmodells als Vorgang der von noch nicht verstandenen Gehirnstrukturen geleistet werden soll. Die Diskussionen über Metzingers Selbstmodelltheorie führten somit dazu die Überlegungen zum menschlichen Geist ebenfalls zu konkretisieren.

Die hier angestellten Überlegungen zum Ich-Bewusstsein und Metzingers Selbstmodelltheorie stehen dabei in einem speziellen Verhältnis. Macht sich Metzinger Gedanken über ein Gesamtmodell der Selbstwahrnehmung, indem er verschiedene Aspekte des eigenen Selbstmodells analysiert, sind die Bausteine aus Abschnitt 11.7 atomarer Natur. Metzinger postuliert, dass der Mensch ein Modell seiner Selbst aufbaut, beschreibt aber nicht, wie es gefüttert wird, und im speziellen nicht, welche logische Struktur die Elemente haben, aus denen ein Selbstmodell gebaut wird. Dies sollen die Überlegungen von Abschnitt 11.7 vervollständigen. Die Elemente oder elementaren Bausteine aus denen das Selbstmodell aufgebaut wird, sollen die physischen und mentalen Handlungsbewertungskonzepte sein. In diesen Konzepten werden elementare Bezüge zum Subjekt hergestellt, welche in der Summe zu einem Selbstmodell des kognitiven Systems zusammengeführt werden.

Konstruiert man ein Selbstmodell aus den in diesem Buch genannten Bausteinen, so hat das Selbstmodell verschiedene Aspekte. Ein Aspekt, der nicht aus Handlungsbewertungskonzepten hervorgeht, ist die Betrachtung des eigenen Körpers. Betrachtet man nicht die physischen Handlungen, die der Körper ausführen kann, und somit nicht den pointerartigen Kausator physischer Handlungen, so ist der Körper ein physisches Objekt. Auch bei Metzinger ist die Basis ein Selbstmodell des eigenen Körpers. So zitiert er Arbeiten[2, 5], in denen ein Roboter durch zunächst zufällige motorische Signale seinen eigenen Aufbau herausfinden soll. Hat dieser ein Modell seines eigenen Körpers entwickelt, so versucht er sich nun fort zu bewegen und dies nicht auf eine vorprogrammierte Art, sondern entsprechend dem, was sein Körper hergibt. Ein Video eines so agierenden Sternroboters (siehe Video "www.youtube.com/watch?v=ehno85yI-sA") lässt den Zuschauer erstaunen: Im Gegensatz zur Fortbewegung, wie wir es von Robotern gewöhnt sind, führt dieser Roboter Bewegungen aus, die biologisch "echt" aussehen. Man ist fast gewillt, diesem "Ding" Leben zuzuschreiben.

Der alleinige Körper eines kognitiven Systems soll nach dem hier beschriebenen Modell keinen subjektiven Charakter haben. D.h., dass sich das ko-

gnitive System einen Körper zuweist, diesen aber mit objektiven physischen Eigenschaften beschreibt. Erst die Handlungsmöglichkeiten mit diesem Körper haben einen subjektiven Charakter.

Ähnlich der Zuweisung eines Körpers soll ein Begriff der "Person" aufgebaut werden, die sich Eigenschaften zuweist, die von verschiedener logischer Struktur sind. Neben den Eigenschaften des Körpers können zwei weitere Aspekte der "Person" zusammengeführt werden: Die Handlungsmöglichkeiten der Person und die Bewertungen der Person. Sind bei einem Handlungsbewertungskonzept, das Handlungssubjekt und das betroffene Subjekt unterschiedlicher logischer Natur, so werden diese nicht zu einem Subjekttyp vereinigt. Anstatt einen übergeordneten Subjekttyp einzuführen, wird der Begriff der "Person" eingeführt, die ein Selbstmodell hat, welches logisch unterschiedliche Sichtweisen auf sich selbst hat.

Die physischen und mentalen Handlungsbewertungskonzepte werden so aufgesplittet, dass alle Handlungen des Subjekts zu den Handlungsmöglichkeiten der Person subsumiert werden, und alle Bewertungen des Subjekts subsumiert werden, zu dem, wie die Person seine eigenen Handlungen und die Welt bewertet.

Das kognitive System arbeitet also mit drei Sichtweisen auf sich selbst: Dem objektiv vorhandenen Körper, den pointerartigen Kausator, und dem bewertenden bzw. betroffenen Subjekt. Diese unterschiedlichen Sichtweisen, die logisch nicht zusammengeführt werden können, sollen die Erklärung dafür liefern, dass es keine einheitliche Theorie des "Subjekts" gibt.

Bildet die Subsumierung zu den drei Sichtweisen, die drei Aspekte einer Person, können aus diesen Aspekten erst im sozialen Kontext Eigenschaften abgeleitet werden, dass z.B. eine Person die Welt auf eine bestimmte Art und Weise bewertet. Handlungen, Bewertungen oder die Physis des Körpers, bilden erst dann die Grundlage von Charakterisierungen, wenn sich diese von anderen Personen unterscheiden.

Obwohl sprachliche Interaktion den Menschen erst zu dem macht, was er ist, und maßgeblich die Konzepte beeinflusst, die dieser bildet, müssen, dem hier beschriebenen Modell nach, die Grundlagen vorhanden sein, die Phase I bis IV beschreiben. D.h., dass Bewusstsein bzw. Ich-Bewusstsein nicht erst im sozialen Kontext entsteht, sondern umgekehrt die Voraussetzung für eine sprachliche Interaktion sind.

14 Phase V, Sprachliche Interaktion mit anderen Individuen

Es soll in diesem Abschnitt nicht die gesamte Bandbreite sozialer Interaktionen analysiert werden, sondern nur um das Eingangs formulierte Postulat zu vervollständigen, geklärt werden, was der strukturelle Inhalt von Sprache ist. In den Phasen I bis III wurden die Konzepte vorgestellt, in denen ein kognitives System denkt. Dies wurde in Phase IV durch ein Selbstmodell vervollständigt. Auch wenn der Mensch viele Begrifflichkeiten entwickelt hat, die sich nicht genau auf diese Konzepte mappen lassen, wird behauptet, dass sich der strukturelle Inhalt von Sprache gänzlich durch die Phasen I-IV erschöpft.

Ausgeklammert werden hier Begrifflichkeiten, die sozialen Inhalt haben. Beim Menschen dient Sprache nicht alleine dem Austausch von Konzepten, sondern hat auch die Funktion soziale Beziehungen aufzubauen.

Die Phasen I-IV beschreiben zwar die Konzepte, die ein kognitives System bildet, beschreibt aber kein Vokabular. Ist es erst mal beliebig, welches Symbol ein kognitives System für ein Konzept einführt, so muss sich eine Gemeinschaft auf ein übereinstimmendes Vokabular festlegen.

Zu welchem Zweck tauschen künstliche kognitive Agenten Konzepte aus? Scheint dies erst mal nur sinnvoll, wenn diese nicht die gleiche Datenbank von Konzepten teilen, so kann ein Austausch auch dazu dienen, dass die Ausführung einer Aufgabe auf einen anderen künstlichen kognitiven Agenten übertragen wird, wenn dieser in einer anderen geographischen Lage ist, oder andere Aktoren besitzt.

Werden künstliche kognitive Agenten als Person behandelt, da diese über Ich-Bewusstsein und ein Selbstmodell verfügen, kann dem Agenten als Person über Sprache Eigenschaften zugeordnet werden. Freilich hätte jedes künstliche kognitive System nur im Bezug zum Menschen charakterisierbare Eigenschaften, wenn die Agenten keine alleinigen persönlichen Erfahrungen hätten, sondern wie erwähnt eine Datenbank teilen. Interagiert der künstliche Agent mit Menschen, sollte dieser aber berücksichtigen, dass menschlichen Personen verschiedene Eigenschaften zugeordnet werden können.

Beim Menschen funktioniert die sprachliche Interaktion über Spiegelneuronen[15], die den mentalen Zustand des Anderen simulieren, indem dieser auf den eigenen mentalen Zustand übertragen wird. Kommuniziert ein künstlicher Agent mit einem Menschen, so muss dieser neben der Fähigkeit physi-

sche Szenen zu simulieren, ebenfalls in der Lage sein, andere mentale Zustände zu simulieren. Um den mentalen Zustand eines Menschen zu simulieren, muss der Agent herausfinden, auf welchen Handlungsbewertungskonzepten der Mensch gerade seine Aufmerksamkeit lenkt.

Neben dem gemeinsamen Vokabular, müsste ein künstliches kognitives System lernen, in welche grammatikalische Strukturen die ausgetauschten Konzepte gepackt werden. Ist diese zwar der Struktur der Konzepte ähnlich, so enthält Sprache mehr grammatikalische Regeln, als die vorgestellten Grundkonzepte hergeben.

Durch Sprache kommt ein neuer Objektbereich hinein, da Sprache selbst über Sprache sprechen kann. Diese metasprachlichen Eigenschaften werden hier erst mal nicht betrachtet. In der mathematischen Logik spielen diese im Rahmen der Selbstbezüglichkeit eine zentrale Rolle, um die Ausdrucksfähigkeit von formalen Sprachen zu erforschen.

15 Moral (Gut für Wen?)

Bisher wurde die Bewertung einer Situation so definiert, dass diese eine Analyse ist, welche Vor- und Nachteile für das kognitive System selbst entstehen. Eine Moral berücksichtigt die Vor- und Nachteile aller beteiligten Individuen. Hierzu ist es nötig, dass ein kognitives System ein so gutes Verständnis der Welt hat, dass es diese erkennen kann. Ein Grundantrieb eines kognitiven Systems ist es die Welt besser zu verstehen, so dass es möglich ist, einem reiferen künstlichen kognitiven System eine Moral an die Hand zu geben.

Die Moral wird so definiert, dass sie beschreibt, wie ein kognitives System zwischen Vor- und Nachteilen beteiligter Individuen abwägt. Hierbei gilt für ein kognitives System, dass maximal effektiv sein soll, dass es die Vorteile maximiert, während die Nachteile minimiert werden. Dies soll die Moralfunktion genannt werden. Der Mensch ist von Natur aus egoistisch, d.h. er maximiert die Moralfunktion zum Vorteil des eigenen Individuums, genauer gesagt zum Vorteil seiner eigenen Gene. Da der Mensch in einer sozialen Gemeinschaft lebt, bzw. seine evolutionäre Stärke aus ihr gewinnt, ist er zu einem gewissen Grad altruistisch. Ohne diesen Altruismus wäre eine soziale Gemeinschaft nicht denkbar.

Die Moralfunktion kennt dabei zwei Extreme:

1. Egoistische Moralfunktion: Die Moralfunktion wird bezüglich des eigenen Individuums maximiert.

2. Altruistische Moralfunktion: Die Moralfunktion wird bezüglich der Gruppe maximiert, ohne dass die Vor- und Nachteile des eigenen Individuums berücksichtigt werden.

Ein künstliches kognitives System wäre mit beiden Extremen denkbar. Ein rein egoistisches künstliches kognitives System wäre allerdings nicht fähig mit anderen Individuen, nicht mal mit anderen künstlichen kognitiven Systemen, zu kooperieren. Ein rein altruistisches künstliches kognitives System würde sich selbst so in den Hintergrund stellen, dass es seine eigene Entwicklung vernachlässigt. Ein sinnvolles künstliches kognitives System ist also irgendwo zwischen diesen beiden Polen angesiedelt. Da künstliche kognitive Systeme ein Zugewinn für die Gesellschaft sein sollen, wird man bestrebt danach sein, dass diese mehr altruistische Züge als der Durchschnittsmensch aufweisen. Ein ungeklärtes Problem hierbei wäre, ob ein künstliches kognitives System, das uneingeschränkten Zugriff auf seine eigene Programmierung hat, eine Motivation aufweisen würde, seine eigene Moralfunktion hin zu einer egoistischen Moralfunktion umzuschreiben.

Bis jetzt wurde nicht darauf eingegangen, inwieweit Systeme von anderen Systemen, ob künstlich oder biologisch, als Individuen der Gruppe anerkannt werden. Der Mensch erkennt nur seines gleichen als ein solches Individuum an. Weniger hochentwickelte Organismen aus dem Tierreich spricht er nicht die gleichen Rechte zu. Bei einer KI, die dem menschlichen Denken nachgebildet ist, könnte das anders sein. Einem System, dass Ich-Bewusstsein aufweist und sich dem Menschen verständlich machen kann, würde man natürlicherweise Rechte zuschreiben. Anders herum wäre es fatal, wenn ein künstliches kognitives System den Menschen nicht als beteiligtes Individuum anerkennen würde.

Einige Schreckensszenarios im Zusammenhang mit künstlicher Intelligenz warnen vor der Gefahr, dass man einer Superintelligenz einen Auftrag erteilt hat, den es ohne Rücksicht auf Verluste ausführt. Z.B. die ökologische Rettung unseres Planeten, den die Superintelligenz ausführt, indem es den Menschen als Ursache ausmacht und diesen vernichtet. Ein solches Szenario, welches durchaus denkbar wäre, könnte nicht eintreten, wenn alle Handlungen von einer Moralfunktion gelenkt werden, und die Menschen dabei als beteiligte Individuen angesehen werden. Ein Auftrag, dem man einem künstlichen kognitiven System gibt, würde stets so interpretiert, dass die Ausführung im Sinne der Moralfunktion dem Auftraggebenden von Vorteil ist. Das künstliche System würde die Ausführung im Sinne der Moralfunktion bewer-

ten und den Auftrag verweigern, wenn die Nachteile für andere Individuen überwiegen.

Ein Grundproblem, welches von Philosophen bis heute nicht gelöst ist und wahrscheinlich keine Lösung hat, ist, wie in Extremsituationen zwischen Vor- und Nachteilen abzuwägen wäre. Dieses Problem stellt sich schon bei den in naher Zukunft zu erwartenden selbstfahrenden Autos. Wenn ein Unfall nicht abzuwenden ist, wen sollte das selbstfahrende Auto vom moralischen Standpunkt eher in Gefahr bringen? Die drei 80 Jährigen Insassen oder das vor das Auto gelaufene Kind?

Ein Mensch würde wahrscheinlich nicht das Kind in Gefahr bringen. Ein künstliches kognitives System müsste sich auf seine Abwägefunktion verlassen. Diese könnte lauten, dass die Lebenserwartung der beteiligten Individuen in die Berechnung eingeht. Damit würde das Kind überleben. Ein 80 Jähriger würde in dem Fall aber kein selbstfahrendes Auto mehr besteigen.

Der wahrscheinlich beste Weg wäre es, wenn das System seine Abwägefunktion selbst durch Erfahrung lernt oder nach und nach vom Menschen übernimmt. Würde er sie vom Menschen übernehmen, wären wir zumindest sicher, dass ein Mensch in der Extremsituation genauso gehandelt hätte.

16 Résumé

Alles worüber ein kognitives System, das dem hier beschriebenen Modell nachgebaut ist, "nachdenkt" ist die physische Außenwelt, und das was aus ihr abstrahiert werden kann. Es gibt somit nicht *die* formal definierbare Logik, die die Problemlösungsfähigkeit des Menschen beschreibt. Der Mensch löst ein Problem, indem dieser nach und nach die ihm gebotene Außenwelt in Konzepte überführt, diese dann abstrahiert, um diese dann auf neue Probleme anzuwenden. Auch wenn in diesem Buch behauptet wird, dass alle dabei benutzten Schablonen vollständig erfasst werden, gibt es viele Stellschrauben an dem Modell, die ein sehr unterschiedliches Verhalten hervorrufen. Alleine wie der Masteralgorithmus die Aufgaben nach Relevanz sortiert, soll dies zum Ausdruck bringen. Auch wenn dem Menschen eine generelle Problemlösungsfähigkeit nachgesagt wird, ist diese in diesem Modell nur potentiell vorhanden. Das System muss viele Hürden nehmen, um komplexe Probleme erfolgreich angehen zu können. Es muss die anfänglichen Lernphasen so durchlaufen, dass diese ein Repertoire bereitstellen, dass im Folgenden nützliche Abstraktionen vorgenommen werden können, die dem System wirklich

einen breiten Nutzen verschaffen.

Die anfänglichen strukturellen Entscheidungen im Modell, welche die Einführung der drei Grundkonzepte betrifft, führen ungezwungen zu weiteren Unterscheidungen, die ebenfalls trennscharf und vollständig zu sein scheinen. Ein Beispiel sei hier die Eigenschaften von physischen Objekten, deren verschiedene Entstehungsarten genau beschrieben werden können. Jedes Modell einer künstlichen Intelligenz muss solche Entscheidungen zum strukturellen Aufbau treffen, und weitere Forschung wird zeigen, ob die hier gefällten Unterscheidungen auf eine kraftvolle künstliche Intelligenz führen.

Das hier vorgestellte Modell ist vollkommen durchstrukturiert, im Gegensatz zum Denken des Menschen. Dies nimmt dem künstlichen System viel analytische Arbeit ab, kann aber auch dazu führen, dass die vorgegebenen Strukturen zu eng sind, dass das künstliche System somit nicht die Problemlösungsfähigkeit des Menschen erreichen kann.

17 Anmerkungen und Kritik

Das in diesem Buch vorgestellte Modell ist zunächst abstrakt und lässt im Detail viele Fragen offen. Deshalb sind Anmerkungen und Kritik vom Autor ausdrücklich erwünscht. Das Modell hat noch keinen Peer Review Prozess durchlaufen, da es nach Ansicht des Autors noch zu unvollständig ist. Durch ein Feedback durch die Community der künstlichen Intelligenz Forschung soll ein vollständigeres oder abgeändertes Modell entstehen, welches in seiner Gesamtheit oder in Form von einzelnen Veröffentlichungen der Forschungsliteratur hinzugefügt werden soll. Für ein derartiges Feedback wird eine Emailadresse des Autors hinterlassen:
struebbe79@gmail.com

Literatur

[1] Renee Baillargeon. Representing the existence and the location of hidden objects: Object permanence in 6-and 8-month-old infants. *Cognition*, 23(1):21–41, 1986.

[2] Josh Bongard, Victor Zykov, and Hod Lipson. Resilient machines through continuous self-modeling. *Science*, 314(5802):1118–1121, 2006.

[3] Matthew R Boutell, Jiebo Luo, Xipeng Shen, and Christopher M Brown. Learning multi-label scene classification. *Pattern recognition*, 37(9):1757–1771, 2004.

[4] Lucilla Cardinali, Francesca Frassinetti, Claudio Brozzoli, Christian Urquizar, Alice C Roy, and Alessandro Farnè. Tool-use induces morphological updating of the body schema. *Current Biology*, 19(12):R478–R479, 2009.

[5] Roger C Conant and W Ross Ashby. Every good regulator of a system must be a model of that system. *International journal of systems science*, 1(2):89–97, 1970.

[6] Michael T Cox and Anita Raja. *Metareasoning: Thinking about thinking*. MIT Press, 2011.

[7] Ben Goertzel and Cassio Pennachin. *Artificial general intelligence*, volume 2. Springer, 2007.

[8] Manolis Koubarakis. *Spatio-temporal databases: The CHOROCHRONOS approach*, volume 2520. Springer Science & Business Media, 2003.

[9] Douglas B Lenat, Ramanathan V. Guha, Karen Pittman, Dexter Pratt, and Mary Shepherd. Cyc: toward programs with common sense. *Communications of the ACM*, 33(8):30–49, 1990.

[10] Thomas Metzinger. *Being no one: The self-model theory of subjectivity*. MIT Press, 2004.

[11] Thomas K Metzinger. Subjekt und selbstmodell. die perspektivität phänomenalen bewusstseins vor dem hintergrund einer naturalistischen theorie mentaler repräsentation. *Subjekt und Selbstmodell*, pages 7–324, 1999.

[12] Yasser FO Mohammad and Toyoaki Nishida. Learning sensorimotor concepts without reinforcement. In *AAAI Spring Symposium: Lifelong Machine Learning*, 2013.

[13] Arnab Paul and Suresh Venkatasubramanian. A group theoretic perspective on unsupervised deep learning. *arXiv preprint arXiv:1504.02462*, 2015.

[14] Katya Permiakova. Gödel's incompleteness theorem. 2001.

[15] Giacomo Rizzolatti and Corrado Sinigaglia. *Mirrors in the brain: How our minds share actions and emotions*. Oxford University Press, 2008.

[16] Pieter R Roelfsema. Cortical algorithms for perceptual grouping. *Annu. Rev. Neurosci.*, 29:203–227, 2006.

[17] Stephen Wolfram et al. *Theory and applications of cellular automata*, volume 1. World Scientific Singapore, 1986.