

Auf eigene Faust? Roboter auf dem Weg in die Mitte der Gesellschaft.

## Auf eigene Faust?

# Roboter auf dem Weg in die Mitte der Gesellschaft.

Ein Beitrag zum Studienpreis der Körberstiftung von Tillmann Pross.

Für uns Menschen ist es eine Selbstverständlichkeit. Wenn wir *denken*, dass wir ein Objekt brauchen, das für uns unerreichbar auf der anderen Seite des Tisches liegt, *sagen* wir zu unserem Gegenüber, von dem wir *denken*, dass er uns das Objekt *geben* könnte „Kannst du mir bitte das da geben!“ und *zeigen* auf das Objekt. Woraufhin unser Gegenüber „Ja“ sagt und uns das Objekt *gibt*. In unserem Alltag lassen sich zahllose Beispiele finden, in denen wir dieses Zusammenspiel zwischen Sprache, Gedanken und Handlungen einsetzen, um unsere Absichten unter Mitwirkung anderer Personen zu realisieren. *Nur* unter uns Menschen ist es aber eine Selbstverständlichkeit, dass dieses Zusammenspiel zwischen Sprache, Gedanken und Handlungen in der beschriebenen Art und Weise funktioniert - das weiß jeder, der sich schon einmal an einer computerbasierten Telefonhotline versucht hat, die das eigene Anliegen partout nicht versteht, bis man schließlich nach dem zehnten Versuch entnervt aufgibt.

Dieser Aufsatz stellt Grundlagenforschung zur Zusammenarbeit von Mensch und Maschine im Kontext ihrer Auswirkung auf und Relevanz für unsere gegenwärtige Gesellschaft und ihrer zukünftigen Entwicklungsoptionen dar, in der Maschinen eine über den reinen Dingstatus hinausgehende Rolle als „auf eigene Faust“ agierende Entitäten einnehmen werden. Dabei geht es um die Entwicklung von Robotern, die die Rolle einer der beiden Personen im oben skizzierten Beispiel einnehmen können. Im Unterschied zur Telefonhotline ist dabei die Überwindung der bestehenden Beschränkung von Interaktionsformen und Interaktionsmitteln zwischen Mensch und Maschine auf den virtuellen Raum von zentraler Bedeutung; durch den Einsatz von Robotern, die in unser reales Lebensumfeld aktiv eingreifen können. Wenn sich aber Mensch und Maschine in einer geteilten Realität in Bezug auf ihre Einwirkungsmöglichkeiten ebenbürtig gegenüberstehen, nehmen intelligente Maschinen eine über den Blechkasten unter dem Schreibtisch hinausgehende Rolle in unserer Gesellschaft ein, deren Ausgestaltung sorgfältig erwogen werden sollte.

Dass es sich bei meiner Arbeit um Grundlagenforschung handelt, kann nicht als Ausrede für die Ausklammerung von gesellschaftsbezogenen Fragestellungen herangezogen werden. Zwar geht der Zeithorizont der gesellschaftlichen Auswirkungen von Grundlagenforschung der Gegenwart um einige Jahre voraus, nichtsdestotrotz erlaubt es diese Entfernung der Grundlagenforschung von der Gegenwart aber nicht - unter Berufung auf nur mögliche Konsequenzen von Forschung - jegliche Reflektion zukünftiger Auswirkungen auszublenden. Die Geschichte ist reich an solchen Fehleinschätzungen über das Potenzial neuer Technologien, stellvertretend sei hier an den IBM-Vorsitzenden Thomas Watson erinnert, der 1943 konstatierte, dass es einen Weltmarkt für möglicherweise 5 Computer gibt oder Kenneth Olsen, dem Gründer von DEC, der 1977 behauptete, dass es keinen Grund gibt, warum jemand einen Computer zu Hause haben wollte.

Die Gefahr solcher Fehleinschätzungen ist, dass erst über die Bewältigung der gesellschaftlichen Konsequenzen neuer Technologien diskutiert wird, wenn der „Point of no Return“ schon erreicht ist und auf die betreffende Technologie nicht mehr

Auf eigene Faust? Roboter auf dem Weg in die Mitte der Gesellschaft.

verzichtet werden kann. Dies halte ich im Falle des hier vorgestellten neuen Typs von selbstständig in unserer Realität operierender Maschine für fahrlässig, wenn nicht gefährlich.

Trotzdem soll natürlich nicht dafür argumentiert werden, die Ergebnisse meiner Dissertation nie in die Realität umzusetzen und auch sonst Forschungen in diesem Bereich zu unterbinden, sondern vielmehr, dass der Einsatz autonomer, auf eigene Faust agierender Roboter in der Mitte unserer Gesellschaft nicht ohne einen gesellschaftlichen Prozess der Reflektion und Diskussion stattfinden darf. Das mag sich dramatisch und gleichzeitig weit weg anhören, aber ich sehe einen enormen Bedarf an genau einer solchen Diskussion über die gesellschaftlichen Auswirkungen des weiteren Fortschritts in der Schaffung autonomer und intelligenter Maschinen. Um diese Diskussion zu (be)fördern, habe ich diesen Aufsatz geschrieben. Zum einen soll er das Mindestmaß an technischen Einblicken vermitteln, das zur Partizipation in solch einer Diskussion vonnöten ist, zum anderen soll der Kontext gesellschaftlicher Relevanz meiner Forschungen aufgefaltet werden, vor dessen Hintergrund eine solche Diskussion stattfinden kann.

Der bescheidene Status Quo der tatsächlich realisierten Integration von Maschinen in die für Menschen charakteristischen Formen von absichtsgeleiteter Interaktion und Kommunikation steht in einem eigentümlichen Kontrast zu der bis in die Antike zurückverfolgbaren Vision der intelligenten Maschine, die uns tatkräftig als Freund und Helfer zur Seite steht. Grundsätzlich lassen sich zwei treibende Kräfte hinter dieser Vision ausmachen.

Erstens – in theoretischer Hinsicht - wird die Maschine als Projektionsfläche betrachtet, anhand derer man spezifisch menschliche Fähigkeiten exemplifizieren kann; im Sinne einer Modellierung der menschlichen Verfasstheit, die im Übertreffen der göttlichen Schöpferkraft den zweiten, besseren, weil unsterblichen und rein rationalen Maschinenmenschen schafft.

Zweitens fordert - in praktischer Hinsicht - die Schaffung von menschenähnlichen Maschinen – Robotern – die Fingerfertigkeit des Ingenieurs heraus, der komplexe Problemstellungen mithilfe technischer Mittel beherrschbar macht – im hier zur Debatte stehenden Fall die Modellierung menschlicher Interaktions- und Kommunikationsfähigkeit. Die Realisierung technischer Zielsetzungen unter Einsatz von geisteswissenschaftlich motivierten Methoden fordert eine eigentümliche Art der Ingenieurstätigkeit, einen „Ingenieur menschlicher Sprach- Denk- und Handlungsfähigkeit“.

Ausgangspunkt meiner Dissertation ist ein der Robotik entnommenes Anwendungsszenario, in dem es um die gemeinsame Bearbeitung von Aufgaben durch Mensch und Roboter geht. Ein einfaches Beispiel für ein derartiges Szenario mit Menschen wurde bereits zu Beginn vorgestellt. Wie kann man nun an die Aufgabe herangehen, eine der beiden Personen im Ausgangsbeispiel durch einen Roboter zu ersetzen, ohne dass der Roboter zu einer Telefonhotline mit Gesicht wird? Eine entscheidende Rolle scheint wohl die Verbindung von Sprache, Denken und Handeln zu spielen. Die Ausarbeitung dieser Verbindung ist die Hauptaufgabe der im folgenden vorgestellten Forschungen. Ich will die Problemlage anhand des Ausgangsbeispiels in einer detaillierten Darstellung - als „Drehbuch“ - wie folgt rekapitulieren.

Auf eigene Faust? Roboter auf dem Weg in die Mitte der Gesellschaft.

Ich sitze an einem Tisch. Mir gegenüber sitzt eine weitere Person. Auf dem Tisch liegt ein Objekt. Ich will das Objekt auf dem Tisch haben. Ich forme die Absicht, mir dieses Objekt zu beschaffen. Das Objekt ist außerhalb meiner Reichweite, also kann ich das Objekt nicht aus eigenen Kräften bekommen. Aber mein Gegenüber ist in Reichweite des Objekts. Ich nehme an, dass mein Gegenüber kooperativ ist. Ich nehme an, dass er mir das Objekt geben kann. Ich nehme an, dass er mir das Objekt geben wird, wenn ich ihn darum bitte. Ich versprachliche diesen Gedanken als eine Bitte, die meine Absicht, mir das Objekt zu beschaffen, mit meinem Gedanken, dass mir mein Gegenüber das Objekt geben kann verbindet. Ich sage: „Kannst du mir bitte das da geben“? Ich zeige auf das Objekt auf dem Tisch. Ich höre, dass mein Gegenüber „Ja“ sagt, und denke mir, dass dies bedeutet, dass er mich verstanden hat, kooperativ ist, mir das Objekt geben kann und mir dieses Objekt geben wird. Ich sehe, dass mein Gegenüber das Objekt in seine Hand nimmt, Ich denke, dass er es mir gerade geben will. Ich sehe: mein Gegenüber hält das Objekt in meiner Reichweite. Durch Greifen des Objekts kann ich meine ursprüngliche Absicht, das Objekt zu haben, realisieren. Ich greife das Objekt. Ich habe das Objekt, also ist meine ursprüngliche Absicht, das Objekt zu haben, realisiert. Ich bin zufrieden.

Ja, diese Schilderung ist langwierig. Das ist sie aber nur, weil uns die Ausbreitung der involvierten Einzelschritte so trivial und überflüssig vorkommt. Tatsächlich wären aber noch eine ganze Menge mehr Einzelschritte und Details nötig, um das Beispiel tatsächlich so zu explizieren, dass alle notwendigen Details derart präzise erfasst sind, dass sie als Grundlage einer Anweisungssequenz für einen Roboter dienen könnten, die diesen in die Lage versetzen würden, eine der beiden Personen am Tisch zu ersetzen.

Ein Roboter hat nichts als eine leere Festplatte und selbst die kleinste Nebensächlichkeit muss bis ins Detail ausbuchstabiert werden. Muss sie? Es wäre ein wahnwitziges Unterfangen, tatsächlich anzustreben, bis ins letzte Detail vorzuschreiben wie eine solche Interaktion ablaufen *muss*. Stattdessen wird es mir im Folgenden um die *Bedingungen der Möglichkeit* gehen, die solchen Interaktionen zugrunde liegen: was es im Kontext solcher Interaktionen bedeutet, dass man etwas greifen kann, dass man etwas sehen kann, dass man einen Gedanken versprachlichen kann.

Durch die Angabe solcher Bedingungen der Möglichkeit soll der Roboter in die Lage versetzt werden, „auf eigene Faust“ an Interaktionen teilzunehmen, ohne dass er einem im Voraus festgelegten Schema folgen muss, sondern vielmehr weiß, wie eine Interaktion ablaufen *kann*. Dass auch dies kein triviales Unterfangen ist, verdeutlichen die vielfältigen Prozesse die in solch einer Interaktion gleichzeitig und in gegenseitiger Abhängigkeit ablaufen. Handlungen, Sprache, Gesten, Wahrnehmungen und Denkprozesse sind so eng miteinander verwoben, bedingen und beeinflussen sich gegenseitig, dass eine isolierte Betrachtung der ineinander verzahnten Phänomene kaum Aussichten auf Erfolg hat. An diesem Punkt setzen die Arbeiten meiner Dissertation an.

Für die Modellierung *einzelner* Bereiche des vorgestellten Beispiels gibt es in der Robotik, den Kognitionswissenschaften, der Computerlinguistik, der Philosophie, Psychologie der Informatik bereits ausgearbeitete und erfolgreiche Modellierungsstrategien. Dies betrifft etwa die Objekterkennung (Informatik), die Analyse von Sätzen in Bezug auf ihre grammatische Struktur (Computerlinguistik), das Greifen von Objekten (Robotik), das Planen von Handlungen

Auf eigene Faust? Roboter auf dem Weg in die Mitte der Gesellschaft.

(Kognitionswissenschaften), das bewusste Reflektieren von Situationen (Philosophie) oder die Konzeptualisierung von Wahrnehmungen zeitlicher Veränderung (Psychologie). Bis jetzt sind diese Bausteine aber unverbundene Fragmente einer Theorie der freien Mensch-Maschine Interaktion. Die Aufgabenstellung meiner Dissertation ist die Entwicklung einer Schnittstelle, die es erlaubt die genannten Ansätze in einem interdisziplinären Verfahren anzudocken.

Ein derartiger Forschungsansatz bringt eigene Probleme mit sich. Dass zum Beispiel zwischen einzelnen Wissenschaften unterschiedliche Auffassungen über die Modellierung gleicher Problemlagen herrschen, dürfte nicht verwundern, dass aber selbst innerhalb einzelner Wissenschaftsdisziplinen scheinbar unüberbrückbare Differenzen existieren schon.

Eine solche Differenz stellt das zentrale Problem meiner Dissertation dar. In der Bedeutungstheorie natürlicher Sprache, also der Disziplin, die sich mit der Bedeutung von linguistischen Entitäten (Wörtern, Sätzen, Satzfolgen) beschäftigt – einer Schnittmenge von Philosophie und Linguistik - existieren zwei sich grundlegend unterscheidende und miteinander konkurrierende Ansätze zur Modellierung von natürlichsprachiger Bedeutung – Semantik und Pragmatik.

Der semantische Zugang zur Bedeutung natürlicher Sprache orientiert sich bei der Untersuchung der Bedeutung von Sprache an der Idealvorstellung einer Mathematik der Sprache, mithilfe derer sich die Bedeutung beliebiger Ausdrücke berechnen lässt, deshalb wird dieser Ansatz auch als *formale Semantik* bezeichnet. Das zentrale Paradigma der formalen Semantik formuliert der Vater der modernen formalen Semantik, Richard Montague, folgendermassen: "there is in my opinion no important theoretical difference between natural languages and the artificial languages of logicians".

Der Brückenschlag zwischen formaler und natürlicher Sprache wird durch die auf den deutschen Mathematiker Gottlob Frege zurückgehenden Auffassung ermöglicht, dass die Bedeutung eines Satzes seine Wahrheit oder Falschheit ist, und jeder Bestandteil eines Satzes seinen Beitrag zur Bedeutung des ganzen Satzes leistet. In anderen Worten: „Einen Satz verstehen, heißt, wissen was der Fall ist, wenn er wahr ist“, wie es der Philosoph Ludwig Wittgenstein formuliert.

Den entscheidenden technischen Durchbruch zur Entwicklung einer formalen Semantik natürlicher Sprache verdanken wir dabei dem Mathematiker Alfred Tarski, der als erster eine formale Definition von Wahrheit vorgelegt hat, die als Grundlage einer auf Wahrheitsbedingungen basierenden Bedeutungstheorie dienen kann. Die grundlegende Idee Tarskis ist, dass sich Wahrheit als regelgeleitete Beziehung zwischen zwei unterschiedlichen Sprachebenen definieren lässt – im Falle der formalen Semantik einer formalen Logiksprache, in der die Wahrheitsbedingungen natürlichsprachiger Sätze abgebildet werden können und einer mengentheoretischen Modellierung derjenigen Aspekte der Realität, die zur Auswertung der logischen Wahrheitsbedingungen des natürlichsprachigen Satzes nötig sind.

In solch einem mengentheoretischen Modell sind die Objekte (also etwa Personen, Tische) und die Relationen in denen diese Objekte zueinander stehen enthalten. Die logische Form des Satzes „Tillmann sitzt am Tisch“ wäre also wahr genau dann wenn im

Auf eigene Faust? Roboter auf dem Weg in die Mitte der Gesellschaft.

Modell der Realität auf die sich dieser Satz bezieht, die Tillmann korrespondierende modelltheoretische Entität in einer Relation ‚sitzt-am‘ zu einer dem Tisch korrespondierenden modelltheoretischen Entität steht. Auf diese Art und Weise lassen sich auch nicht-triviale Fragmente natürlicher Sprache analysieren.

Der zweite Ansatz zur Bedeutung natürlicher Sprache - die *Pragmatik* - nimmt einen grundsätzlich anderen Weg als die formale Semantik. Ausgangspunkt für die Bedeutung ist hier nicht die mathematische Präzision von Logiksprachen, sondern die Alltagssprache. Das zentrale Kennzeichen unserer Alltagssprache ist dabei die Art und Weise, wie wir Sprache einsetzen, um etwas zu *tun*. Wir heiraten, *indem* wir „Ja“ sagen, wir überzeugen jemanden *dadurch*, dass wir etwas sagen.

In diesem Kontext des alltäglichen Gebrauchs von Sprache spielt Wahrheit –wenn überhaupt - nur eine untergeordnete Rolle. Stattdessen kann der Gebrauch von Sprache (das *indem* oder *dadurch*) gelingen oder misslingen – so wie das für alle anderen Arten von Handlungen auch der Fall ist. Die Bedeutung einer Bitte oder die Bedeutung des Ja-Sagens im Rahmen einer Trauung besteht folglich nicht aus Wahrheitsbedingungen sondern aus Bedingungen, die das Gelingen der Bitte oder des Vollzugs der Trauung spezifizieren – das etwa die Trauung von einer dafür autorisierten Person durchgeführt wird.

Dieser unterschiedliche Zugang zum Phänomen Bedeutung zeigt sich auch im Vokabular von Semantik und Pragmatik, das zur begrifflichen Definition von Bedeutung herangezogen wird. In der formalen Semantik wird Bedeutung ohne Rückgriff auf psychologische Zustände wie z.B. Absichten modelliert. So wie psychologische Zustände keinen Einfluss auf das Ergebnis einer mathematischen Rechnung haben sollen, sollen psychologische Zustände auch keinerlei Einfluss auf die Bedeutung eines Satzes haben. Im alltäglichen Gebrauch der Sprache spielen psychologische Entitäten aber eine zentrale Rolle, sind Handlungen doch durch ihren absichtlichen Ausführung definiert und motiviert und nur mithilfe der Unterstellung einer Absicht als Bitte oder Überzeugungsversuch verständlich.

Die Überwindung dieser holzschnittartig skizzierten Differenz zwischen Semantik und Pragmatik durch die Entwicklung einer Semantik-Pragmatik Schnittstelle ist das zentrale Anliegen meiner Dissertation. Um aber die Kluft zwischen Semantik und Pragmatik zu überwinden, muss man sich einem weiteren Problem stellen: dass man sich *zwischen* zwei tradierten Argumentationslinien befindet und es somit eigentlich keiner Seite recht machen kann, da jeder der konkurrierenden Ansätze Deutungshoheit einfordert. Diese Problematik von Schnittstellenforschung zeigt sich auch in Bezug auf die Behauptungsansprüche meiner Arbeit, die sich *zwischen* Geisteswissenschaften und Ingenieurwissenschaften wiederfinden. In der Einleitung hatte ich meine Forschung als Ingenieurstätigkeit mit geisteswissenschaftlichen Methoden bezeichnet. Die Frage ist dann, ob meine Arbeit eher eine Form von „cognitive engineering“ ohne weitergehende Ansprüche in Bezug auf die Natur des Menschen darstellt oder eine Modellierung kognitiver Fähigkeiten mit dem Anspruch auf Adäquatheit gegenüber der tatsächlichen menschlichen Kognitionsfähigkeit („cognitive modelling“) ist .

Die Antwort auf diese Frage - ob mit meiner Arbeit Implikationen philosophischer oder psychologischer Natur über den menschlichen Kognitionsapparat einhergehen - beruht zuvorderst auf der Akzeptanz der von mir eingesetzten Theoriefragmente aus den

Auf eigene Faust? Roboter auf dem Weg in die Mitte der Gesellschaft.

Einzelwissenschaften. Akzeptiert man diese, dann lässt sich meine Arbeit auch als Bild menschlicher Kognitionsfähigkeit lesen, akzeptiert man sie nicht, ist meine Arbeit die Realisierung einer technischen Aufgabestellung. Damit sitzt meine Arbeit gewissermaßen zwischen allen Stühlen – disziplinar, methodisch und von Ihrem Anspruch. Das lässt sich auch daran festmachen, dass es mir schwer gefallen ist, die Arbeit in die Fachklassifizierung des Studienpreises einzuordnen.

Zum einen handelt es sich um eine Arbeit, die an einer geisteswissenschaftlichen Fakultät akzeptiert wurde – andererseits hat mein Heimatinstitut während meiner Promotionszeit von der philosophisch-historischen Fakultät in die Fakultät für Informatik gewechselt; mein Erstgutachter ist Philosoph, mein Zweitgutachter Robotiker. Letztendlich habe ich mich dafür entschieden, diesen Aufsatz im Bereich Technik einzureichen, da das primäre Anliegen meiner Arbeit die Ausstattung eines Roboters mit einer Semantik-Pragmatik Schnittstelle ist und nicht das Aufstellen von Behauptungen über die menschlichen Semantik-Pragmatik Schnittstelle.

Nach diesen Ausführungen zur allgemeinen Methodik und Problemstellung meiner Arbeit werde ich nun ausführlicher die in meiner Arbeit entwickelte Schnittstelle zwischen Semantik und Pragmatik vorstellen. Dafür gehe ich nochmals etwas detaillierter auf die bereits bekannte Beispielinteraktion ein.

Ein erstes Problem stellt sich bereits, wenn man die Ausgangssituation betrachtet. Uns als Menschen bereitet es keinerlei Probleme, unsere *visuelle* Wahrnehmung, dass wir an einem Tisch sitzen, in eine *semantische Repräsentation* dieser Wahrnehmung umzuwandeln. Eine semantische Repräsentation einer Situation ist dabei als „geistiges Bild“ zu verstehen, als logische Struktur eines Bewusstseinsinhalts, dessen Bestandteile sich auf die Situation *beziehen*, oder, im Fachjargon *referieren*.

Roboter sind nicht per se in der Lage solche Repräsentationen zu formen, sondern müssen vielmehr mit detailliert ausgearbeiteten Mechanismen ausgestattet werden, die es erlauben, visuelle Inputs in semantische Repräsentationen umzuwandeln. Im Beispiel sind aber nicht nur visuelle Eindrücke von Belang, ebenso müssen *Äußerungen*, etwa meine Bitte, mir das Objekt zu geben, unter Einsatz von semantischen Repräsentationen verarbeitet werden. Für Äußerungen lässt sich dabei auf den bereits existierenden Formalismen der formalen Semantik aufbauen, im vorliegenden Fall auf der sogenannten Diskursrepräsentationstheorie (kurz: DRT).

Die grundlegende Motivation für den Gebrauch von semantischen Repräsentationen in der DRT ist dabei, dass semantische Repräsentationen (als Erweiterung klassischer Logiksprachen) die Wahrheitsbedingungen der Äußerung spezifizieren und damit die Bedeutung der Äußerung erfassen. In meiner Arbeit wird die DRT über ihre ursprüngliche Intention als Formalismus zur formalen Analyse natürlichsprachiger Bedeutung hinaus auf die Erfassung nichtsprachlicher Inhalte erweitert, als Grundlage der Verarbeitung von visuellen und akustischen Wahrnehmungen.

Das zweite Problem bezieht sich auf die Referenz von semantischen Repräsentationen. Um auch die Referenz von Repräsentationen zu explizieren, also den Bezug von Bewusstseinsinhalten, setzt man, wie zuvor ausgeführt, in der formalen Semantik noch eine zweite Ebene der Bedeutung an - Modelle. Modelle enthalten Informationen die für die Auswertung des Bezugs gegebener Repräsentation nötig ist. Ein Roboter hat per se

Auf eigene Faust? Roboter auf dem Weg in die Mitte der Gesellschaft.

aber keine Modelle. Woher sollen diese Informationen kommen? Nun, im einfachsten Fall speisen sich die gewünschten Modelle aus direkten Wahrnehmungen des Roboters wie sie etwa seine Objekt – oder Spracherkennung liefert. Der entscheidende Teil der Modelle, der für eine Charakterisierung der Vorgänge in der Beispielinteraktion vonnöten ist, bezieht sich aber auf den nicht direkt wahrnehmbaren Teil der Realität: die Zukunft.

Uns Menschen erscheint die Zukunft vor unserem „inneren Auge“. Wie aber soll man diese Metapher so aufbereiten, dass sie auf einen Roboter übertragbar ist? Ein Anhaltspunkt ist hierbei, dass wir Menschen unsere Zukunft nicht einfach „sehen“, sondern die Zukunft anhand von Vorhersagen über zukünftige Entwicklungsoptionen der gegenwärtigen Situation *entwerfen*. Zentral für den Umgang des Menschen mit seiner Umwelt sind dabei die Entwicklungsoptionen, die er durch seine Einwirkungen auf die Umwelt kontrolliert. Das heißt: der zukünftige Teil von Modellen ist eng verknüpft mit den Begriffen der Planung, der Absicht und der vorhersagenden Erklärung.

Solche Begriffe gehen aber über das Vokabular der formalen Semantik hinaus. Stattdessen sind Planung, Absicht und Vorhersage Elemente der Pragmatik, sie beziehen sich auf Handlungen, nicht auf Wahrheit. Im Hinblick auf die Modelle, mit denen ein Roboter ausgestattet werden sollte, muss dem Roboter also die Möglichkeit gegeben werden, seine Zukunft auf der Basis gegebener Wahrnehmungen anhand von Plänen und Vorhersagen selbst in die Hand zu nehmen. Dies wird in meiner Dissertation mithilfe einer Verbindung von Planungsalgorithmik aus der Informatik und Erklärungsschemata aus der Philosophie geleistet.

Die Ausformulierung der Verbindung und gegenseitigen Abhängigkeit von semantischen Repräsentationen und pragmatisch motivierten Modellen ist das Kernstück der von mir entwickelten Schnittstelle zwischen Semantik und Pragmatik: es ist diese Form der Rückführung der Referenzstrukturen von Repräsentationen auf reale Zustände und zukünftige Pläne, die die Interpretation von semantischen Repräsentationen fundiert (daher der Titel meiner Arbeit 'Grounded' Discourse Representation Theory, kurz: GDRT).

Die GDRT erweitert die pragmatische Aktionsfähigkeit eines Roboters um den Umgang mit komplexen semantischen Repräsentationen bzw. verankert die Forschungsrichtung der formalen Semantik in einem pragmatischen Fundament sensomotorischer Kontrolle. Umgesetzt ist diese Schnittstelle als dynamische Kontrolleinheit, die das Verhalten des Roboters in Echtzeit - als Laufzeitalgorithmus - kontrolliert und koordiniert. Mithilfe dieser Echtzeitkontrolle lässt sich die Verbindung von formaler Semantik und ihrer Wechselwirkungen mit der Pragmatik (z.B. in der Interpretation von Bitten) als dynamische Wechselwirkung zwischen semantischen Repräsentationen und ihren Modellen formulieren.

Technisch setzt die GDRT auf eine neuartige Theorie der Referenzstrukturen – der Bedeutung - semantischer Repräsentationen. Es handelt sich dabei um die Idee, dass sich Referenzstrukturen mithilfe von sogenannten „Ankern“ modellieren lassen. Die Beziehung zwischen einem Referenten - einem Ausdruck der ein bestimmtes Objekt bezeichnet - und seiner Referenz – etwa das Verhältnis zwischen dem Referenten 'Tisch' und dem Objekt Tisch – hat in der von mir vorgeschlagenen Analyse bildlich gesprochen

Auf eigene Faust? Roboter auf dem Weg in die Mitte der Gesellschaft.

folgende Struktur. Die semantische Repräsentation stellt das Schiff dar, das im Meeresgrund, dem Modell der Realität verankert werden muss. Die Semantik-Pragmatik Schnittstelle spezifiziert das *zielgerichtete* Werfen von Ankern für die einzelnen Bestandteile einer Repräsentation in das Modell. Der Technik des richtigen Ankerwerfens liegt hierbei die philosophisch motivierte Annahme zugrunde, dass die Identifikation von Sachverhalten und Entitäten im Modell (in die verankert werden soll) durch *Erklärungen* solcher Sachverhalte und Entitäten ermöglicht wird. Eine Erklärung wird dabei so verstanden, dass sie Informationen liefert, welche spezifische Qualität Sachverhalte und Entitäten so auszeichnet, dass auf sie referiert werden kann, das man sie als Ankergrund einsetzen kann.

Für die Analyse von zeitlicher Referenz, etwa von Verben wie „geben“ sind in Analogie zum Anker Netze nötig, mit denen man aus einer Menge von zeitlich aufeinander folgenden Modellschnappschüssen diejenigen Informationen herausfischt, auf die sich repräsentierte Ereignisse und Zustände beziehen. Dabei wird auf den psychologisch fundierten Vorschlag zurückgegriffen, dass die Segmentierung kontinuierlicher temporaler Prozesse aus unsegmentierten Wahrnehmungen (den Modellzuständen) als auf Kausal- und Planungsstrukturen basierende Erklärung verstanden wird.

Ein beobachteter Prozess kann daher durch die Annahme einer zugrunde liegenden Erklärungsstruktur (etwa einer Kausalerklärung oder einer auf Absichten basierenden Erklärung) als eine temporale Entität identifiziert werden (etwa das Greifen eines Objekts), auf die dann mit einer Repräsentation der temporalen Entität referiert werden kann.

Diese Auffassung von Referenz als *Handlung* des Ankerwerfens wiederum ermöglicht es, sowohl der Wahrheitsbedingensemantik gerecht zu werden (dann geht es darum, die Wahrheitsbedingungen von Repräsentationen in Modelle zu verankern) als auch der Gelingensbedingungenpragmatik (wenn Modelle so verändert werden müssen, dass Verankerungen möglich werden – etwa durch das Ausführen von realen Handlungen, deren zukünftige Effekte sich in den Modellen wiederfinden werden).

Im Beispiel wäre es also sowohl von Bedeutung, dass der Roboter weiß, unter welchen Umständen es wahr ist, dass ich das Objekt auf dem Tisch haben will und dass der Roboter ein Verfahren besitzt, mit dem er die nötigen Operationen bestimmen kann um angemessen auf diese Bitte zu reagieren. Bildlich gesprochen muss der Roboter sein Gedankenschiff der semantischen Repräsentation der Äußerung so durch die Zukunft steuern, dass ich schließlich das Objekt in Händen halte.

Der Punkt, auf den ich mit der Vorstellung dieser Kernthesen der GDRT hinaus will, ist dass die mit der GDRT entwickelte Form von selbstständiger Maschine weit über eine in ihren Möglichkeiten im voraus festgelegte Telefonhotline hinausgeht.

Mit der eben erörterten Kombination von semantischen und pragmatischen Ansätzen zur Verarbeitung von Mensch-Maschine Interaktionen eröffnen sich neue Perspektiven der natürlichen und intuitiven Interaktion zwischen Mensch und Maschine, in der die Maschine nicht länger ein passives Mittel menschlicher Tätigkeit ist sondern *aktiv selbst* ins Geschehen eingreift. So weit so gut. Aber: Warum treibt man diese Forschungen überhaupt voran? Die Antwort auf diese Frage führt uns in die eingangs angerissene Frage nach der gesellschaftlichen Relevanz meiner Forschungen, die ich im Folgenden



Auf eigene Faust? Roboter auf dem Weg in die Mitte der Gesellschaft.

näher ausführen will.

Die Vision vom Roboter, der zweiten Schöpfung des besseren, weil rein rationalen, ausschließlich zweckorientierten und mit übermenschlichen Kräften ausgestatteten Maschinenmenschen ist keine Erfindung der Neuzeit: schon die antiken Griechen träumten von Androiden, das Mittelalter von Golem und Homunculus und spätestens mit der durch den Rationalismus eingeläuteten Wende zur Moderne hat diese Vision eine Leitfunktion als Metapher des menschlichen Strebens nach göttlicher Perfektion übernommen.

Begleitet wurden solche Mythen immer aber auch von Kontrollverlusten über das geschaffene Werk, wobei das bekannteste Beispiel wohl der Shelleysche Frankenstein ist. Grundsätzlich ist die methodologische Strenge der Mathematik und des Maschinenbaus nicht der Verursachung solcher Horrorszenarien verdächtig. Es ist vielmehr die Kombination der mathematischen Methode der Berechnung mit der „animalischen“ Maschine, die der reinen Rationalität zusetzt, sie dunklen Trieben unterwirft und zur „Entfesselung“ des Hybridwesens führt.

Diese Janusköpfigkeit der Menschmaschine, der Ermöglichung von grenzüberschreitender Erweiterung menschlicher Fähigkeiten einerseits und der Wendung der Maschine gegen den Menschen als „Strafe“ für die Grenzüberschreitung andererseits erfordert eine reflektive Begleitung des Einsatzes von auf eigene Faust agierenden Robotern in der Mitte unserer Gesellschaft.

Die Chancen, die der Einsatz von Robotern für aktuelle Problem und zukünftige Entwicklungsoptionen unserer Gesellschaft bietet sind enorm. Roboter sollen dort eingesetzt werden, wo Arbeitskraft fehlt, wo es für Menschen zu gefährlich ist, wo der Mensch nicht präzise genug ist, wo der Mensch nicht billig genug ist, wo der Mensch nicht willig genug ist, wo der Mensch nicht ausdauernd genug ist, wo der Mensch nicht schnell genug ist.

Das anvisierte Potenzial des Einsatzes von Robotern betrifft alle Bereiche menschlichen Lebens. Im privaten Haushalt als Serviceroboter für kranke und alte Menschen, als Unterhaltungsroboter oder als Lebensgefährte. Im Dienstleistungssektor im Bereich der Reinigung, Überwachung, Transport und Logistik oder für Tätigkeiten in Verkauf und Beratung, in der Industrie, in der Medizin, und natürlich, als treibende Kraft von besonderer Brisanz, im Militär.

Dass diese nach Science-Fiction anmutenden Visionen nicht länger nur Visionen bleiben werden, kann die Parallele zwischen der Situation der Robotik heute und der Computer in den 70er Jahren des letzten Jahrhunderts verdeutlichen. Roboter heute, das sind wenige teure Prototypen in Forschungslaboratorien und Industrie, deren Einsatz, Wartung und Bedienung Fachpersonal vorbehalten ist.

Der Durchbruch des Computers als Massenprodukt ab den 1980er Jahren war vor allem durch einfache, mehr oder weniger intuitive Bedienungsmöglichkeiten bedingt. Man musste kein Wissenschaftler mehr sein, um einen Text am Computer zu schreiben oder eine Berechnung durchzuführen. Das entscheidende Kriterium für die gesellschaftliche Akzeptanz von Computern war die Möglichkeit, dass auch Laien durch graphische Oberflächen, Tastatur und Maus ohne größeres Vorwissen mit Computern umgehen

Auf eigene Faust? Roboter auf dem Weg in die Mitte der Gesellschaft.

konnten.

Mit der Möglichkeit zur einfachen und intuitiven Interaktion (etwa durch simple Computerspiele) war dann auch die ökonomische Voraussetzung geschaffen, billige Computer für den Massenmarkt zu entwickeln. Am weiteren Beispiel des Internets schließlich lässt sich sehr schön nachvollziehen, in welcher rasanter Geschwindigkeit sich technische Neuerungen, so sie denn erstmal massenkompatibel geworden sind, bis in die letzte Nische menschlichen Lebens ausbreiten und alle Aspekte gesellschaftlichen Zusammenlebens prägen. Im Hinblick auf die von mir skizzierte Form von Robotik – als auf eigene Faust agierende Maschine die sprechen, denken und handeln kann – und ihres Einsatzes in einer mit Menschen geteilten Realität potenzieren sich sowohl mögliche Chancen als auch die Risiken.

Die Chancen sehen den Roboter als Mitspieler, als Partner - als *Gegenüber* - in den vielfältigen Rollen die wir Menschen in unserer Gesellschaft spielen können: des Verkäufers, des Altenpflegers, des Soldaten, des Butlers, des Freundes.

Die Risiken sind in düsteren Farben in uns allen bekannten Science-Fiction Szenarien ausgemalt worden, in der außer Kontrolle geratene Maschinen die Fundamente unserer Gesellschaft untergraben und letztendlich zerstören. Im Folgenden will ich ein aus dem Einsatz von selbständig agierenden Maschinen resultierendes Risiko der Unterminierung und Zerstörung gesellschaftlicher Fundamente etwas genauer erörtern.

Nicht nur die hier vorgestellten Maschinen, auch wir Menschen handeln auf eigene Faust, trotzdem sind wir Menschen in der relativ komfortablen Lage, uns (noch) nicht in solchen Endzeitszenarien zu bewegen, wie sie für den Einsatz auf eigener Faust handelnder Maschinen entworfen werden. Warum?

Aus moralischer Perspektive kennzeichnet es uns Menschen, dass wir uns als *autonome* Wesen verstehen, die deshalb für ihre Taten *verantwortlich* gemacht werden können, weil *wir* es sind die absichtlich gehandelt haben. Die Absicht einer Handlung liefert die Begründung dafür, inwiefern wir Verantwortung für unsere Taten übernehmen müssen. Auf dieser Tatsache beruht nicht nur unsere institutionalisierte Moral in Form eines Rechtssystems, sondern auch unser alltäglicher Umgang mit anderen Menschen. Werden nun Roboter als autonom und selbstständig konstruiert und mit den technischen Voraussetzungen ausgestattet, Absichten zu fassen und diese auch zu realisieren, müssen wir klären, wie es dann um die Verantwortung für von Robotern ausgeführte Taten steht. Für den Fall von nicht selbständig handelnden Maschinen – etwa einer Saftpresse – ist der Fall klar: es ist entweder Sache des Konstrukteurs, Verantwortung für Verfehlungen die Maschine zu übernehmen oder des Maschinisten.

Für einen Roboter, der ohne Zutun von externen Einflüssen Entscheidungen trifft, sieht die Sache anders aus. Hier ist der Roboter nicht mehr der verlängerte Arm des Konstrukteurs und einen Maschinisten gibt es nicht. Wenn keine Verbindung zwischen dem Konstrukteur und selbständigen Entscheidungen des Roboters existiert, kann der Konstrukteur nicht zur Verantwortung gezogen werden, ohne den gängigen Verantwortungsbegriff ad absurdum zu führen. Wir können aber auch nicht einfach sagen; dieser Roboter ist für sein Handeln verantwortlich. Autonomie und Verantwortung können wir nur *jemandem* – einer Person - zuschreiben, aber nicht einem Ding.

Auf eigene Faust? Roboter auf dem Weg in die Mitte der Gesellschaft.

Wie soll man dann mit einem Fall umgehen, in dem ein Roboter z.B. einem Menschen Schaden zufügt? Die sicherlich bekannteste Idee zur Lösung solcher Probleme stammt von Isaac Asimov, der die folgenden Gebote einer Roboterethik postuliert:

1. Ein Roboter darf kein menschliches Wesen verletzen oder zulassen, dass einem menschlichen Wesen Schaden zugefügt wird.
2. Ein Roboter muss dem Befehl gehorchen, der ihm vom Menschen gegeben wurde, es sei denn, der Befehl kollidiert mit Regel eins.
3. Ein Roboter muss seine Existenz schützen, solange er dadurch nicht Regel eins oder zwei verletzt.

Nun führt die Schlichtheit dieser Regeln aber unweigerlich in Entscheidungsdilemmata, etwa im Falle einer Abwägung verschiedener Handlungsoptionen unter Einsatz der ersten Regel. Soll man zehn Menschen beschützen wenn man dafür drei Menschen schadet? Um solchen Dilemmata zu entgehen, fügt Asimov dann auch später folgende Regel einer „allgemeinen Sichtweise“ hinzu, die aus solchen Situationen heraushelfen soll:

4. Ein Roboter darf der Menschheit keinen Schaden zufügen oder durch seine Untätigkeit gestatten, dass die Menschheit zu Schaden kommt.

Diese Regel bezieht zwar die Abwägung zwischen verschiedenen Handlungsoptionen unter dem allgemeinen Gesichtspunkt des „Schadens“ explizit mit ein, gleichzeitig wird an dieser Regel ein grundlegendes moralisches Problem deutlich. Was als Schaden für „die Menschheit“ aufzufassen ist, kann unter unterschiedlichen Gesichtspunkten bewertet werden. Eine eindeutige, empirischen oder logischen Argumentationen folgende Charakterisierung von „Schaden für die Menschheit“ gibt es nicht. Damit verschließt sich die vierte Regel aber auch einer Form, die sich einfach so auf Robotern implementieren ließe.

Wie vereinbart aber ein Mensch die Vielzahl einzelner Ansichten von etwa „Schaden“ mit einer allgemeinen Sichtweise? Es ist gerade der Versuch einer widerspruchsfreien Vereinbarung des individuellen Handelns mit einer allgemeinen Sichtweise, die den Menschen solche Fragen als *moralisch* betrachten lässt.

Nun ist es aber keinesfalls so, das der Mensch nur aus Prinzip moralisch handelt, die bewusste Selbstbeschränkung des Menschen zugunsten einer dritten Person (einem allgemeinen Gesichtspunkt) kann man auch als einem bestimmten Zweck dienend betrachten: als Konstitution einer Gesamtheit von *Normen*, die ein gesellschaftliches *System* unter dem Gesichtspunkt bestimmter funktionaler Erfordernisse zusammenhält.

So betrachtet ist Moral die Gesamtheit funktionsrelevanter Vollzugsmuster der einzelnen Subsysteme der Gesellschaft. Im Verhältnis der Sichtweisen von Subsystem und System zeigt sich das Muster moralischen Handelns. Die innere Sicht des Subsystems besteht aus idealen Richtlinien des Handelns, Vorgaben bzw. Normen, die in einer bestimmten Situation erprobt werden können. Verantwortung lässt sich dann als Rückkopplungsmedium des Systems verstehen, das die Bewertung von Übertretungen bzw. Erprobungen der Moral ermöglicht.

Auf eigene Faust? Roboter auf dem Weg in die Mitte der Gesellschaft.

Wird z.B. die Norm der körperlichen Unversehrtheit übertreten, wird dies als unmoralisch geahndet, um das System in einem stabilen Zustand zu erhalten, in dem die Funktionsleistung der körperlichen Unversehrtheit und freien Entfaltung gewährleistet ist. Verantwortung lässt sich in diesem Sinne als Regelungsinstrument einer Gesellschaft auffassen. Bedingungen der Möglichkeit von Verantwortlichkeit sind damit die Fähigkeit zur Rückkopplung mit dem gesellschaftlichen Gesamtsystem und die Fähigkeit, Operationen danach auszurichten. Ermöglicht wird diese Fähigkeit durch institutionalisierte Rahmen, in dem verantwortungsbezogene Rückkopplungen vollzogen werden können.

Dieser Rahmen ist - für uns Menschen - der Personenstatus. Und für Maschinen, die auf eigene Faust agieren? Wollen, sollen, müssen wir Roboter als Personen betrachten? Wie kann oder muss der Rückkopplungsrahmen für Maschinen beschaffen sein?

In Bezug auf letztere Frage liefert meine Dissertation zumindest die technische Grundlagen der Rückkopplung in *unseren* gesellschaftlichen Zusammenhängen durch den Einsatz des Zusammenspiels von Sprache, Denken und Handeln in der Explikation interner Beweggründe. Bezüglich der Frage nach dem Status von Robotern in unserer zukünftigen Gesellschaft sollten wir uns über die zentrale Rolle des Personenbegriffs für unsere Gesellschaft im Klaren sein: der Personenbegriff ist die Mitte unserer Gesellschaft. Wir sollten uns gut überlegen, ob wir zulassen wollen, dass Roboter den Weg in diese Mitte auf eigene Faust gehen.

Die Erörterung dieser Problemlage ist kein wissenschaftliches Problem, sondern erfordert einen gesellschaftlichen Diskurs darüber, wie die Interessen und Bedürfnisse von Wirtschaft, Politik, Wissenschaft und Gesellschaft im Hinblick auf die Nutzung von Chancen und der Beherrschbarmachung von Risiken des den Einsatzes autonomer Maschinen auszutarieren sind. Die Befreiung der Maschine aus dem engen Skelett klassischer Blechkasten-Algorithmik wird eine neue Spezies von handelndem Subjekt hervorbringen, auf die wir gut vorbereitet sein sollten.