Klaus Scheuermann

Menschliche und technische ‚Agency': Soziologische Einschätzungen der Möglichkeiten und Grenzen künstlicher Intelligenz im Bereich der Multiagentensysteme

Technical University Technology Studies
Working Papers
TUTS-WP-2-2000
Institute for Social Sciences
Technische Universität Berlin

Berlin 2000

Herausgeber:

Fachgebiet Techniksoziologie
Prof. Dr. Werner Rammert

Technische Universität Berlin
Institut für Sozialwissenschaften
Franklinstraße 28/29
10587 Berlin

Sekretariat Rosemarie Walter

E-Mail: rosemarie.walter@tu-berlin.de

Gliederung

1. Einleitung

1.1. Das Projekt der künstlichen Intelligenz

Seit Beginn der Menschheitsgeschichte lassen wir uns von der Möglichkeit einer technischen Herstellbarkeit von Intelligenz faszinieren. Science-Fiction-Figuren - Roboter, Humanoide, Androide, Cyborgs - stellen den kulturellen Hintergrund dar, vor dem wir heute die Frage ‘künstlicher Intelligenz’, d.h. die Möglichkeiten und Grenzen intelligenter Softwareprogramme, Softwareagenten und Roboter diskutieren. Hierbei tritt das Projekt der Erforschung ‚künstlicher Intelligenz‘ (KI) in Abgrenzung von anderen technischen Innovationen wie z.B. dem Rad oder dem künstlichen Licht mit dem Anspruch der Nachbildung bzw. Imitation eines natürlichen Vorbildes, nämlich des Menschen als einem biologischen, geistigen und/oder sozialen Wesen, auf. Allerdings können in diesem Zusammenhang drei unterschiedliche Anspruchsniveaus der KI-Forschung unterschieden werden (Gold 1998, S. 51ff.). Im Zuge ontologischer Ansprüche wird auf die Konstruktion von Maschinen bzw. Softwareprogrammen abgezielt, deren intelligente Ausdrucksformen und Leistungen sich nicht wesentlich bzw. ‘essentiell’ von der natürlichen Intelligenz des Menschen unterscheiden sollen. Anders fokussieren epistemologische Ansprüche auf die Konstruktion von Maschinen, die zwar nicht im einem ontologischen Sinne intelligent sind, aber das intelligente Verhalten der Menschen nachbilden, imitieren bzw. simulieren können. Der Augenmerk dieser beiden Ansätze gilt entsprechend nicht nur der Konstruktion funktionierender Technologien, sondern auch und vor allem - im Sinne kognitionwissenschaftlicher Fragestellungen - Einsichten in die Strukturen und Prozesse menschlichen Denkens und Handelns. Demgegenüber wendet die KI-Forschung als ein Zweig der Ingenieur- bzw. Informatikwissenschaften ihre Methoden und Konzepte auf geeignete Problemstellungen an. Hier wird unter einer anwendungsorientierten Perspektive - im Sinne einer Logik des technisch Machbaren - die Gestaltung und Implementation von für die jeweiligen Aufgabenstellungen und Verwendungskontexte geeigneten Softwaresystemen angestrebt.

Während sich hinsichtlich der anwendungsorientierten Perspektive der KI bei der Entwicklung von neuen Programmiertechniken und -sprachen ein mehr oder weniger kontinuierlicher Wissens- und Technologiefortschritt beobachten läßt (vgl. Brenner u.a. 1998), können hinsichtlich der kognitionswisssenschaftlichen Ansprüche der KI drei paradigmatische Phasen voneinander abgegrenzt werden. Der ersten Phase, dem sogenannten ‘symbolischen Paradigma’ der KI, lag ein ‘Rechnermodell’ des menschlichen Geistes (Krämer 1994) in Form von auf formaler Logik und rationaler Entscheidungs-theorie aufbauenden psychologischen Kognitionstheorien zugrunde. In der zweiten Phase, dem ‘subsymbolischen’ bzw. ‘konnektionistischen’ Paradigma der KI, orientierten sich Modellbildungen der KI vorwiegend an einem biologischen Verständnis des menschlichen Gehirns. Demgegenüber gehen in der dritten Phase der KI, im Rahmen der Entwicklung von ‘Verteilter künstliche Intelligenz’ (VKI) und von ‘Multiagentensystemen’ (MAS), auch soziale bzw. soziologische Vorstellungen und Konzepte in die Programmierung von KI-Systemen mit ein. Nicht mehr die quasi-intelligenten Leistungen von einzelnen Softwareentitäten, sondern vielmehr intelligentes Problemlösen als kollektive Leistung mehrerer - sachlich (funktional), räumlich und/oder zeitlich verteilter - Einheiten soll entsprechend sozialen Vorbildern programmiert werden (zusammenfassend Huhns 1987).

Der Modellbildung der VKI und MAS stellt sich folglich die soziologischen Erklärungen sozialer Ordnung analoge Frage, wie ‘verteilte’ sowie mehr oder weniger heterogen bzw. divergierend orientierte Einzelentitäten ihr Operieren koordinieren und dabei zu spezifischen kollektiven Problemlösungen gelangen. Entsprechend kann hier von einer konzeptuellen ‘Überlappung’ der Modellbildungen der VKI bzw. MAS und der Soziologie ausgegangen werden. An dieser Theoriestelle setzt das in Deutschland neu etablierte Forschungsfeld der ‘Sozionik’ mit dem Ziel der ‘Erforschung und Modellierung künstlicher Sozialität’ an. Im Sinne einer "experimental interaction between two different epistemic cultures" (Rammert 1998b, S.13) sollen durch eine konstruktive Zusammenarbeit von Informatikern und Soziologen Modelle des Sozialen für die Informatik nutzbar, d.h. auf informatorische Systeme übertragen werden (zusammenfassend Malsch/Müller (Hg.)1998, Malsch (Hg)1998, Brauer u.a. 1998).

1.2. Sozionik und die Absichten dieses Textes

Innerhalb des Forschungsfeldes ‘Sozionik’ sind zwei Fragestellungen zentral. Erstens wird auf die Analyse der quasi-sozialen Formen und Dynamiken künstlicher Agentensysteme und zweitens auf die Analyse von hybriden Vergemeinschaftsformen zwischen menschlichen Akteuren und künstlichen Agenten (Softwareagenten und Robotern) abgezielt. Hinsichtlich der ersten Fragestellung kann zwischen den Ansprüchen einer ‘starken’ und einer ‘schwachen’ Sozionik unterschieden werden (Malsch/Müller 1998, S. IV). Die sogenannte ‘starke Sozionik’ will unter Rückgriff auf die Methoden und Techniken der VKI bzw. MAS computerförmige Simulationen von sozialen Zusammenhängen und Prozessen entwickeln, aus denen entweder prognostische Verfahren oder auch Evidenzen für soziologische Theoriebildungen bzw. -entscheidungen gewonnen werden können (vgl. Saam 1996, Ahrweiler/Gilbert (Hg.)1998). Während in diesem Fall die Soziologie von den neuen technischen Simulationspotentialen der VKI bzw. MAS profitieren will, soll sie demgegenüber im Falle der ‘schwachen Sozionik’ als eine Grundlagen und/oder Hilfswissenschaft die VKI bzw. MAS bei der Konstruktion und Implementierung neuer Technologien unterstützen. Soziologische Analysen der Formen und Prozesse sozialer Koordination und des Zusammenhangs von lokalem sozialem Handeln und globalen gesellschaftlichen Strukturen (Mikro-Makro-Problem) sollen in der VKI bzw. MAS verwendet, d.h. in ihre jeweiligen Modellbildungen übersetzt und integriert werden. Im Rahmen der zweiten Fragestellung - der Analyse von Hybridgemeinschaften - steht demgegenüber weniger Modellbildung von Agentensystemen als vielmehr die Analyse ihrer gesellschaftlichen Anwendungsbedingungen und Folgen im Vordergrund. ‘Hybridgemeinschaften’ verkörpern neue soziotechnische Ordnungsformen, insofern hier menschliche Anwender bzw. User auf künstliche, eine Vielzahl neue - potentiell dem menschlichen Handeln vergleichbare - Verhaltensweisen umsetzende Agenten treffen und somit hier noch nicht bekannte Formen und Prozesse von Mensch-Maschine-Kooperationen entstehen.

Der vorliegende Texte will im Rahmen der Sozionik sozialtheoretische Begrifflichkeiten entwickeln, die sowohl eine soziologische Einschätzung der Potentiale der Modellierung künstlicher Sozialität in Form von Multiagentensystemen als auch eine Beschreibung der in

hybriden Gemeinschaften neu auftretenden Sozialformen ermöglichen. Hierbei werden Modellierungsversuche von Multiagentensystemen - in Abgrenzung von einem rein anwendungs- bzw. ingenieurwissenschaftlichen Selbstverständnis der VKI (vgl. Burkhard 1993) - hinsichtlich ihrer Ansprüche einer Nachbildung, Simulation und/oder Imitation des Sozialen, d.h. der Konstruktion einer menschlichen (individuellen und kollektiven) Fähigkeiten vergleichbaren künstlichen Intelligenz von Agenten und Agentenkollektiven ernst genommen. Darüber hinaus geht dieser Text von der Annahme aus, daß in den philosophischen und soziologischen Auseinandersetzungen um das symbolische und subsymbolische Paradigma der KI bereits eine Vielzahl von Argumenten entwickelt wurden, die auch für die soziologische Bewertung der Möglichkeiten und Grenzen einer soziologisch inspirierten und/oder fundierten Modellierung von Agenten und Agentensystemen relevant bleiben. Demzufolge beginnt der Text in Kapitel 2 mit einer kurzen Darstellung der Grundgedanken des symbolischen und des subsymbolischen Paradigma der KI (2.1.) und konfrontiert jene mit philosophischen und soziologischen Einschätzungen und Kritiken, wie sie aus der Perspektive der philosophischen Phänomenologie und Sprachphilosophie sowie der soziologischen Handlungstheorie gegenüber jenen ‘traditionellen’ Ansätzen der KI vorgebracht werden (2.2.). In Kapitel 3 werden dann die Grundüberlegungen der VKI bzw. MAS eingeführt und die unterschiedlichsten Konzepte und Ansätze einander gegenübergestellt (3.1.- 3.4.). In den Blickpunkt gerückt wird das Paradigma ‘sozialer Agenten’, im Rahmen dessen insbesondere die Autoren Les Gasser und Rosario Conte/Cristiano Castelfranchi - im Gegensatz zu der Auffassung einer rein ‘inspirierenden’ Rolle von Sozialmodellen bzw. -metaphern innerhalb der MAS (vgl. Malsch 1997) - eine ‘soziologische Fundierung’ der MAS einfordern, um ein adäquates Verständnis sozialer Intelligenz sowie leistungsfähigere Technologien entwickeln zu können (3.5). Hieran anschließend sollen mit Hilfe von Grundüberlegungen der Giddensschen Strukturierungstheorie - aber auch vor dem Hintergrund der in Kapitel 2 entwickelten Argumente - die Möglichkeiten, aber auch die Grenzen einer soziologischen Fundierung der MAS in Form der Übertragung von soziologischen Kategorien auf künstliche Systeme im allgemeinen und des Anspruches einer Nachbildung bzw. Simulation des Sozialen im besonderen aufgezeigt werden (3.6.).

In Kapitel 4 soll die in Kapitel 3 am Beispiel der Schwiergkeiten bzw. Widersprüche einer soziologischen Fundierung der MAS entwickelte Kritik überzogener Ansprüche hinsichtlich der Nachbildung bzw. Simulation von Sozialität konstruktiv gewendet und potentielle zukünftige Fragestellungen innerhalb des Forschungsbereiches Sozionik angerissen werden. Erstens kann die Soziologie versuchen, nicht nur - wie im Rahmen der MAS bereits weitestgehend geschehen - ihre potentiell formalisierbaren und programmtechnisch umsetzbaren Modelle, sondern auch und vor allem ihre theoretischen Einsichten hinsichtlich der grundsätzlichen Differenzen von technischen und sozialen Systemen wie auch der sozialen Verfaßtheit jeweiliger Anwendungskontexte in der Entwicklung der Agentensysteme einzubringen (4.1.) Zweitens erscheint es angesichts von KI-Produkten, deren Operationsweisen sich in Form von mehr oder weniger autonomen Softwareagenten und Robotern zunehmend den Eigenschaften menschlicher Verhaltensformen annähern, sinnvoll, sowohl für die Analyse jener Operationsweisen als auch für die Verhaltensweisen der mit ihnen konfrontierten User neue Begrifflichkeiten bereitzustellen, d.h. im Zuge der Analyse jeweiliger Hybridgemeinschaften eine Theorie

der Mensch-Maschine-Interaktivität zu entwickeln (4.2.). Drittens können Soziologen hierauf aufbauend versuchen, Einschätzungen hinsichtlich der Innovations- und Risikopotentiale der neuen Agententechnologie zu formulieren, d.h. die Vor- und Nachteile der Modellierung von (vermeintlich) menschlichen Verhaltensweisen und Sozialformen analogen (sozio-) technischen Systemen abzuwägen und sinnvolle von nicht sinnvollen Anwendungsbereichen voneinander abzugrenzen (4.3.).

2. Die traditionelle KI und ihre Kritiker

2.1. Turing und seine Nachfolger

2.1.1. Das symbolische Paradigma

Die Möglichkeit der technischen Konstruktion künstlicher Intelligenz eröffnete Alan Turing (1950) durch den Nachweis, daß jede algorithmische Rechenaufgabe durch eine - zu Turings Zeiten papierförmige, lochkartenähnlich aufgebaute - 'Universalmaschine' ausgeführt und folglich jedes regelgeleitetes menschliches Verhalten maschinell umgesetzt werden kann. Turings These der Mechanisierbarkeit regelförmiger menschlicher Verhaltenformen liegen zwei Annahmen zugrunde (Heintz 1993, S.267ff.). Erstens kann - so die Grundannahme des Funktionalismus - Denken als Informationsverarbeitung unabhängig von unterschiedlichen materiellen Substraten realisiert werden. Zweitens kann menschliches Denken und Verhalten immer vollständig diskursiv dargestellt, d.h. von den Programmierern in Form von Daten und Regelsystemen explizit beschrieben werden. Diese beiden Annahmen wurden zum Fundament einer der beiden Hauptstränge der KI-Forschung, dem sogenannten 'symbolischen Paradigmas' der KI, das die Entwicklung intelligenter Computer in den 1960er und 1970er Jahren beherrschte. Das symbolische Paradigma versteht menschliche Intelligenz als rationales Problemlösen in Form eines auf jeweiligen Daten - 'Repräsentationen' bzw. sprachlichen Darstellungen - der jeweiligen (Um-)Welt aufbauenden logischen Schlußfolgerns (Newell/Simon 1992). Ihm liegt demnach die philosophische bzw. psychologische Vorstellung von menschlichem Denken als einem logischen Kalkulations- bzw. Rechenprozeß sowie das entsprechende sozialtheoretische, in der soziologischen Spiel- bzw. Rational-Choice-Theorie dominierende Bild des Menschen als einem rational abwägenden Handlungsträger zugrunde (vgl. Krämer 1994).

Ziel der Entwicklung symbolverarbeitender KI-Systeme ist es, die einem spezifischen Verhalten zugrundeliegenden Realitätsausschnitte und Regeln - z.B. in Form von 'frames' (Minsky) oder 'scripts'(Schank/Abelson) - in Modellen zu explizieren, zu formalisieren und anschließend in Softwareprogramme umzusetzen. Ihren Programmen liegt hierbei das Prinzip der 'symbolischen Klassifikation'(Green et.al 1997, S.6 und ff.) zugrunde. Adäquate Umweltrepräsentationen werden in spezifische Kategorien und Klassen untergliedert, und anschließend werden jeweils neue Informationen anhand spezifischer Entscheidungsregeln wiederum diesen Kategorien zugeordnet. In diesem Sinne können die entsprechenden KI-Systeme neue Situationen erkennen , d.h. ein 'maschinelles Lernen' umsetzen. Als zentrales Problem des symbolischen Paradigma stellte sich die Modellierung

bzw. symbolische Kategorisierung eines menschlichem Handeln zugrundeliegenden und vorwiegend implizit bzw. praktisch konstituierten Alltagswissens dar. So gelang zwar die Konstruktion von Expertensystemen wie z.B. Dendral, das das medizinische Fachwissen eines klar abgrenzbaren Wissensgebiet darstellt. Auch das Sprachanalyseprogramm Eliza, das eine spezifische - nämlich passive und strikt formale - Form der Gesprächsführung eines Psychotherapeuten nachahmt, war erfolgreich. Im Gegensatz zur Darstellung von durch eine hohe, meist theoretisch bedingte Vorstrukturiertheit gekennzeichneten Aufgabenfeldern sahen sich allerdings Versuche, das für die Bewältigung eher unstrukturierter (Alltags-) Situationen notwendige und ganzheitlich-holistische organisierte Hintergrundwissen menschlicher Akteure zu modellieren, großen Schwierigkeiten ausgesetzt. Während Menschen - wie insbesondere die philosophische Phänomenologie oder die Gestaltpsychologie hervorgehoben haben (vgl. unten) - in solchen Kontexten eher 'intuitiv' die Situation erfassen und in einen Gessamtzusammenhang einordnen können, stoßen auf der Basis von Daten und Regelvorgaben operierende Softwareprogramme auf das Problem komplexer, letztendlich unendlicher Rechenleistungen bei der Berechnung der Gesamtheit situativ relevanter Faktoren. Insbesondere im Falle der von menschlichen Akteuren aufgrund ihres praktisch-impliziten Alltags-bzw. Erfahrungswissens geleisteten Antizipation zukünftiger Situationen entsteht bei regelbasierten Computerprogrammen das sogenannte 'frame problem' in Gestalt der 'kombinatorischen Explosion' der Berechnung zukünftiger Systemzustände (Dennett 1984).

2.1.2. Das subsymbolische Paradigma

Als Reaktion auf die internen Schwierigkeiten des symbolischen Paradigmas verlagerte sich seit Anfang der 1980er Jahre der Schwerpunkt der KI-Forschung zu dem sogenannten 'subsymbolischen' bzw. 'konnektionistischen' Paradigma' (Churchland/Churchland 1990, Dreyfus/Dreyfus 1988). Dieses orientiert sich nicht an einem Rechner- bzw. Kalkülmodell des menschlichen Geistes, sondern zielt vielmehr auf den Nachbau der biologischen Funktionsweise und Architektur des menschlichen Gehirns. Sogenannte neuronale Netze setzen sich anders als durch einen zentralen Prozessor und einen zentralen Datenspeicher organisierte Computer aus einer Vielzahl von einzelnen Mikroprozessoren (Neuronen) zusammen. Hierbei werden Informationen bzw. 'Erfahrungen, letztlich das gesamte Systemwissen bzw. -gedächtnis als sich über das gesamte Netz erstreckende Neuronenmuster verstanden, die durch spezifische Umweltreize aktiviert werden können. Die Intelligenz solcher neuronalen Netze stellt sich folglich im Sinne eines 'biologischen Holismus' (Dreyfus/Dreyfus 1988, insbesondere S. 307ff.) als das emergente, d.h. nicht auf die Eigenschaften seiner Einzelbestandteile reduzierbare Produkt der Verbindungen vieler einzelner 'subsymbolisch' operierender und entsprechend 'unintelligenter' Neuronen dar.

Ziel der Konstruktion neuronaler Netze ist nicht die korrekte Repräsentation einer mehr oder weniger als konstant unterstellten Umwelt im Sinne des symbolischen Paradigma, sondern vielmehr die Implementierung fortlaufender flexibler Anpassungsleistungen an eine sich verändernde Umwelt. Der zentrale Unterschied zwischen dem symbolischen und dem subsymbolischen Paradigma und entsprechend zwischen konventionellen Expertensystemen und neuralen Netzen liegt hierbei in der Methode ihrer Programmierung. Während die Programmierung z.B. von Expertensystemen größtenteils auf vorgängigen

Einprogrammierung von Fakten bzw. Kategorien und entsprechenden Entscheidungsregeln beruht, stellt sich die Programmierung von neuronalen Netzen als ein schrittweiser ‘experimenteller’bzw. ‘induktiver’ Lernprozeß dar. So sind Programme des Konnektionismus vermittels dem Verfahren der ‘subsymbolischen Klassifikation’ (Green et.al., S.6 und ff.) zu einem ‘Lernen’ im Systemverlauf bzw. in der Phase der Systemnutzung in der Lage. Diese Vorstellung eines ‘selbständigen’ Lernens trifft insbesondere auch sogenannte "unsupervised neural networks" (ebd.) zu, die im Gegensatz zu "supervised neuronal networks" (ebd.) nicht auf das ‘Training’ vor dem Systemverlauf festgelegter Kategorien angewiesen sind, sondern vielmehr eigenständig neue Kategorien bilden und neue Informationen in diese einordnen. Der Lernprozeß neuronaler Netze ist folglich durch die fortlaufende - durch den Vergleich alter und neuer Informationsmuster gewährleistete - Integration neuer ‘Erfahrungen’ gekennzeichnet. Kriterium des Lernen ist hierbei ausschließlich ihre erfolgreiche Umweltanpassung, die vermittels des Selektionsprinzips von ‘Versuch und Irrtum’ - entweder mit den Programmierern oder ausschließlich der ‘Umwelt’ bzw. den Nutzern als Instruktions- bzw. Korrekturinstanz - die Fähigkeit erwerben, adäquate von nichtadäquaten Erfahrungen bzw. ‘richtige’ von ‘falschen’ Verhaltensweisen zu unterscheiden.

2. 2. Philosophische und soziologische Bewertung der KI

2.2.1. Der Turing-Test: Computer als Imitationsmaschinen

Alan Turing (1950) hat bzw. hatte vorgeschlagen, bei der Bewertung der Leistungsfähigkeiten intelligenter Maschinen die ontologische Frage - ‘Können Computer dem Menschen vergleichbar denken?’- von der epistemologischen Frage - ‘Können Computer das Verhalten des Menschen imitieren’ – abzukoppeln. Programmen soll genau dann Intelligenz zugeschrieben werden, wenn sie erfolgreich menschliches Sprachverhalten nachahmen, d.h. imitieren bzw. simulieren können. Gemäß dem von Turing eingeführten ‘Imitationsspiel’ (dem sogenannten Turing-Test) ist die Zuschreibung von Intelligenz genau dann gerechtfertigt, wenn ein menschlicher Beobachter, der über einen Fernschreiber einem Menschen und einem Computerprogramm Fragen stellt, anhand der jeweiligen Antworten nicht die Maschine von dem Menschen unterscheiden kann. Im Anschluß an Turing lassen sich im Rahmen der philosopischen und soziologischen Kritik der KI zwei Argumentationslinien unterscheiden (Heintz 1993, S. 279ff.). Auf der einen Seite verneint eine entgegen den Turingschen Intentionen ontologisch argumentierende ‘Hollow shell’-Kritik das Intelligent-Sein von menschliches Verhalten imitierenden Computerprogrammen (2.2.2.). Auf der anderen Seite bestreitet eine epistemologisch argumentierende ‘Poor-substitute’- Kritik, daß Computerprogramme überhaupt die Fähigkeit der adäquaten Imitation menschlicher Verhaltensweisen und Leistungen erreichen können (2.2.3.).

2.2.2. Die ‘Hollow-shell’- Kritik der KI: Können Computer Sprache verstehen?

Gemäß der ‘Hollow-shell’-Kritik der KI wird zugestanden, daß Computerprogramme potentiell (heute oder in Zukunft) menschliches Verhalten erfolgreich im Sinne des Turing-

Testes imitieren können, aber abgestritten, daß dies die Zuschreibung von Intellligenz rechtfertigt. Computerprogramme - so der Grundgedanke dieses Argumentes - sind nicht in einem dem Menschen vergleichbaren Sinnne intelligent, insofern ihnen keinerlei kognitive Eigenschaften wie Selbstbewußtsein, Subjektivität und/oder Intentionalität zukommen. So zeigt John Searle (1990, S.20ff., vgl. derselbe 1986, 1997) als der bekannteste Vertreter dieser Position in seinem sprachphilosophischen Gedankenexperiment 'Chinese room', daß Computerprogrammen auch bei der erfolgreichen Imitation menschlichen Sprachverhaltens nicht die Fähigkeit des Sprachverstehens zukommt: Eine in einem Zimmer eingeschlossene Person - so insistiert dieses Gedankenexperiment vor allem gegenüber dem symbolischen Paradigma - kann dergestalt nach vorgebenen Regeln mit Symbolen operieren, daß ein Beobachter außerhalb des Raumes sinnvolle Aussagen interpretieren kann, während die Person im Raum deren Sinn keineswegs versteht. In anderen Worten: Computerprogramme verfügen nur über einen Zugang zur formalen Regelstruktur von Sprache (Synthax), nicht aber zu ihren sinn- bzw. bedeutungstragenden Aspekten (Semantik), so daß von der Sprachperformanz des Computers nicht auf Intelligenz konstituierendes Sprachhandeln oder Sprachverstehen geschlossen werden kann.

Diese Fähigkeit, sich intentional bzw. 'semantisch verstehend' auf jeweilige Sachverhalte zu beziehen und sie hierbei mit Sinn bzw. Bedeutung zu versehen, hat Searle in späteren Schriften (vgl. Searle 1997) insbesondere in Abgrenzung von funktionalistischen Grundannahmen in Richtung eines biologischen Naturalismus spezifiziert. Intentionalität als das spezifische Merkmal menschlichen Wahrnehmens, Denkens und (Sprach-) Handelns soll nach Searle auf die biologische Ausstattung des Menschen, genauer auf die spezifische Struktur des menschlichen Gehirns zurückgeführt werden. Ähnlich argumentieren auch die philosophische Phänomenologie (z.B. Dreyfus/ Dreyfus 1987) und die phänomenologisch ausgerichtete Handlungstheorie (z.B. Joas 1992a, insbesondere S. 233ff.). Jene verstehen allerdings nicht wie Searle ausschließlich das menschliche Gehirn, sondern die Gesamtheit des menschlichen Körpers und dessen Wahrnehmungsapparat als Grundlage eines sinnkonstituierenden Weltbezugs. So führt auch Rammert (1998a) den originären Charakter einer verstehenden Reflexivität menschlichen Handelns auf ein Computerprogrammen - hier: Computeragenten (vgl. unten) - grundsätzlich nicht zugängliches, auf der körperlichen Verfaßtheit des Menschen beruhendes Sinn- bzw. Bedeutungsverstehen zurück: "Die Reflexionsfähigkeit menschlicher Agenten (...) beinhaltet das Verstehen, nicht nur das Handhaben von Bedeutungen. Bezeichnungen und Bezeichnetes werden wie bei technischen Agenten nicht nur nach festen Regeln aufeinander bezogen, sondern erst der Weltbezug, den sie über den Körper erfahren, erschließt ihnen die Bedeutung der Dinge." (ebd., S. 113)

Der Nachweis, daß Computerprogrammen kein Weltbezug im Sinne einer intentionalen bzw. verstehenden Ausrichtung an jeweiligen Sinn- bzw. Bedeutungszusammenhängen zukommt, wird aber auch - näher dem Grundgedanken von Searles ursprünglich sprachphilosophischem Argument (ohne den phänomenologischen Rückgriff auf die fundierende Rolle des menschlichen Körpers) - aus einer kommunikationstheoretisch-soziologischen Perspektive geführt, die an die Tradition der interpretativen Soziologie anschließt. Nach Alan Wolfe (1991, vgl. auch 1993), der im Anschluß an den Symbolischen Interaktionismus und die Ethnomethodologie argumentiert, haben Computer im Gegensatz zu menschlichen Akteuren keinen kognitiv konstituierten 'social mind' und

somit keinen Zugang zu sinnhaft-semantisch vermittelten Kommunikations- und Interaktionbeziehungen. Entsprechend kommt ihnen weder die Fähigkeit zur interpretativen Sinnkonstruktion und -rezeption (‘meaning producing’) noch die anschließenden praktischen Kompetenzen der Konstruktion und Reproduktion von (bedeutungshaften) soziokulturellen Strukturzusammenhängen zu. Ähnlich argumentiert auch Harry Collins (1996, vgl. auch 1990, 1995), der in Auseinandersetzung mit Dreyfus (1996) betont, daß weniger das in der Phänomenologie betonte ‘embodiment’ als vielmehr die ‘social embeddedness’ jeweiliger Kognitions- und Wissensformen sozialer Akteure der - Computerprogrammen grundsätzlich nicht möglichen - Teilhabe an jeweiligen interaktiv konstituierten ‘social worlds’ zugrundeliegt. Auch Terry Winograd/Fernando Flores (1989) insistieren darauf, daß Computerprogramme nicht an sozialen Prozessen, hier vor allem nicht an sozialer Kommunikation partizipieren können. Sie rekurrieren in ihrer Generalabrechnung mit den (vermeintlich) rationalistischen Grundannahmen des symbolischen Paradigma sowohl auf phänomenologische als auch auf kommunikationstheoretische Überlegungen. Insbesondere im Anschluß an sprechakttheoretische Überlegungen von Searle und Habermas betonen sie die Computerprogrammen nicht zugänglichen normativen, soziale Verpflichtungen umfassenden Aspekte von sozialer Kommunikation : "Computer sind ungeeignet, Verpflichtungen einzugehen und können sich daher nicht selbst am Sprachprozeß beteiligen." (ebd., S. 133)

2.2.3. Die ‘Poor-Substitute’-Kritik der KI: Was Computer können und was sie nicht können

Die ‘Poor-Substitute’- Kritik der KI schließt die - z.B. von Searle nicht in Frage gestellte - Möglichkeit von erfolgreichen Imitationen menschlichen Denken und (Sprach-) Handelns durch Computerprogramme aus. Über die These der Unmöglichkeit eines sinnhaften Welt- bzw. Sozialitätsbezugs von Computerprogrammen hinaus wird hier davon ausgegangen, daß die kognitionswissenschaftlichen Modelle der KI nicht mit den faktischen Formen und Prozessen menschlichen Wahrnehmens, Denkens, Sprechen und/oder Handelns übereinstimmen und folglich jene nicht adäquat imitieren oder substituieren können. Bekannt geworden ist diese Position vor allem durch Hubert Dreyfus’ phänomenologisch geführten Nachweis der ‘Grenzen künstlicher Intelligenz’ (Dreyfus/Dreyfus 1987, vgl. auch Dreyfus 1996), der im Anschluß Heidegger und Merleau-Ponty die zentrale Rolle eines körperlich verankerten, holistisch organisierten und sich kontextfreien bzw. objektiven Beschreibungen entziehenden lebensweltlichen Hintergrundwissen aufzeigt. Im Gegensatz zu diesem praktisch-impliziten ‘know how’ kommt dem ‘know that’, d.h. den sprachlich explizierbaren, formalisierbaren und entsprechend programmierbaren Bestandteilen sowohl unseres Alltagswissens als auch unseres Expertenwissens nach Dreyfus eine eher untergeordnete Rolle zu, so daß der Nachahmung des Menschen durch (regelgeleitete) Computerprogramme enge Grenzen gesetzt sind.

Allerdings kann die phänomenologisch motivierte Kritik des KI-Ansatzes nur schwer erklären, warum sich spezifische Programme wie Dendral oder Eliza in spezifischen Handlungskontexten bewährt, d.h. den Turing-Test bestanden und menschliche Akteure erfolgreich imitiert bzw. substituiert haben. Insbesondere der Versuch von Dreyfus (1996), diesem Phänomen mit der Unterscheidung verschiedener, mehr oder weniger regelförmig

strukturierter und entsprechend mehr oder weniger erfolgreich bzw. vollständig programmierbarer Wissensbereiche beizukommen, erweist sich als widersprüchlich, insofern gemäß phänomenologischen Grundannahmen jegliche Wissensformen in einem nichtformalisierbaren lebensweltlichen Kontext verankert sind (so Heintz 1993, S.282 und ff.). Demgegenüber schlägt Bettina Heintz (ebd.) eine spezifisch soziologische Analyse der Imitationsfähigkeiten intelligenter Maschinen vor, die nicht auf die Differenzierung von Wissensformen, sondern auf die Bestimmung und Abgrenzung von Handlungstypen abstellt und somit die (vermeintliche) Äquivalenz von Mensch und Maschine von zwei Seiten beleuchten kann: "Der Computer in Turings Imitationsspiel hat immer zwei Möglichkeiten, das Spiel zu gewinnen. Entweder er verhält sich wie ein Mensch - oder aber der Mensch wie eine Maschine." (ebd., S. 292) Hieran anschließend vertritt Heintz die These, daß in der Moderne in Folge kultureller und gesellschaftlicher Rationalisierungsprozesse (Webers Bürokratiethese, Taylorismus, etc.) von einer sukzessiven Annäherung menschlichen Denkens und Handelns an berechen- und formalisierbare Prozesse ausgegangen werden kann, die als ‘maschinenähnliche’ Handlungs- bzw. Verfahrensweisen mehr oder weniger erfolgreich von regelbestimmten Maschinen imitiert und substituiert werden können: "Die Computerisierung ist nicht der Anfang, sondern der vorläufige Endpunkt einer Entwicklung, die sehr viel früher, lange vor dem Einsatz der ersten Computer begonnen hat und dazu geführt hat, daß sich Menschen in zunehmendem Maße regelhaft - maschinenähnlich - zu verhalten haben. (...) Ohne die tiefgreifende Umstrukturierung von Handlungsfeldern unter der Maxime der Regelhaftigkeit und Berechenbarkeit wäre nicht ein so breites Spektrum menschlichen Handelns so normiert worden, das eine maschinelle Imitation problemlos ermöglicht."(ebd., S.299)

Mit dieser Vorstellung eines maschinenähnlichen Verhaltens menschlicher Akteure rekurriert Heintz auf den von Collins (1990) eingeführten Begriff der ‘behaviour specific acts’ - an anderer Stelle auch: ‘machine-like acts’ -, mit dem dieser den Erfolg spezifischer KI-Programme erläutert. Die Collinssche Analyse geht von der traditionellen handlungstheoretischen Unterscheidung von Verhalten und Handeln aus. Während Verhalten (z.B. ein Wimpernschlag) auf bloßen Reiz-Reaktions-Abfolgen beruht und somit eindeutig als kausale Folge einer Körperbewegung beschrieben werden kann, stellt sich Handeln als kontingent im Sinne von ‘auch anders möglich’ dar: eine Körperbewegung wird genau dann als Handeln beschrieben, wenn sie aufgrund von - dem Akteur selbst bewußten oder nichtbewußten - Intentionen vollzogen wird und andere Intentionen zu anderen Effekten hätten führen können.[1] Anschließend verweist Collins auf das als soziologisches Problem des Fremdverstehens bekannte und auch - allerdings ausschließlich hinsichtlich von Sprachintentionen und Sprachperformanzen - dem Turing-Text zugrundeliegende Problem der Bestimmung interner Handlungsintentionen durch einen externen Beobachter (entweder einen Interaktionspartner oder auch einen soziologischen Beobachter). Entscheidend für Collins’ Argument sind hierbei solche Handlungsformen, die sich entweder bewußt oder unbewußt an festen Regeln ausrichten und sich demzufolge einem Beobachter als nicht-kontingent, d.h. als Verhalten darstellen: "My argument is that

1. In anderen Worten: Handeln kann im Gegensatz zu Verhalten ‘mißlingen’ , insofern nur hier die Effekte mit den Akteursintentionen abgeglichen, d.h. entweder als intendierte oder im Falle des Scheiterns als nichtintendierte Handlungsfolgen interpretiert werden können.

even thought a defining characteristic of actions is that they may be carried out in many ways, humans somethimes forgo this option (...) Behaviour specific action are acts that humans always try to instantiate with the same behaviour." (ebd. S. 33) Solche regelgelorientierten und somit verhaltens- bzw. maschinenähnlichen Handlungsformen können von einem alter ego nicht von einem bloßen Verhalten unterschieden und folglich von regelgeleiteten Maschinen im allgemeinen und Computerprogrammen im besonderen erfolgreich imitiert werden.

Der Nachweis von 'behaviour specific acts' als solchen Handlungsformen, die durch Computerprogramme z.B. innerhalb des Turing-Testes imitiert und folglich durch diese innerhalb spezifischer Anwendungskontexte substituiert werden können, kann aber in unterschiedliche Richtungen weitergeführt werden. So behaupten z.B. Bamme u.a. (1983) die sukzessive Annäherung maschineller Logiken und menschlicher Handlungsformen als Folge der zunehmenden sozialisatorischen Internalisierung formaler Regelstrukturen in der Moderne, wobei sie auf strukturalistische Grundannahmen sowie auf eine von Lacan vertretene strukturalistisch-psychoanalytische Interpretation der Piagetschen Entwicklungspsychologie rekurrieren. Demgegenüber bezieht sich Collins – ähnlich wie auch Alan Wolfe (vgl. oben) oder auch die KI-Kritikerin Lucy Suchman (1987, 1997) - in seiner weiteren Analyse der 'behaviour specific acts' auf Grundüberlegungen der interpretativen Soziologie, wobei er sich insbesondere auf eine (antistrukturalistische) ethnomethodologische Interpretation des Wittgensteinschen Regelkonzeptes stützt. Hierbei verläßt er die Teilnehmerperspektive des alter ego bzw. des teilnehmenden Beobachters im Turing-Test und verweist aus einer externen soziologischen Beobachterperspektive auf - den Akteuren selbst sowie ihren Handlungspartnern nichtoffensichtliche bzw. nichtbewußte – interpretative Flexibilitäten und praktische Kompetenzen, die deren Handeln nicht nur in eher unstrukturierten Handlungskontexten bzw. -phasen, sondern auch in strukturierten Kontexten und vor allem im Falle der 'behaviour specific acts' zugrundeliegen. Nach Collins unterscheiden sich regelorientierte 'behaviour specific acts' menschlicher Akteure von einem regeldeterminierten Operieren von Computerprogrammen in zweierlei Hinsicht. Auf der einen Seite liegen sowohl ihrem Erwerb als auch ihrem Vollzug spezifische Anstrengungen zugrunde. "It is important to note that machine-like action is not easy for humans. It takes a lot of training and a substantial effort of will for a human to disguise action in this way." (ebd. S. 35) Auf der anderen Seite beruhen 'behaviour-specific acts' wie alle Formen eines Routinehandelns auf einer Vielzahl subtiler, sowohl interpretativer (kognitive) als praktischer Anpassungs- und Reparaturleistungen menschlicher Akteure, wie auch im Falle jeweiliger Mensch-Maschine-'Interaktionen' beobachtet werden kann: "It turns out that most tasks seem to be merely routine in fact depend on the operativ being of an actor rather than a machine, and altering the behaviours to instantiate the action in subtle ways." (ebd.)

Hierauf aufbauend kann eine ethnomethodologische Perspektive auf Mensch-Maschine-'Interaktionen' nach Collins erklären, warum trotz der grundsätzlichen Differenzen von maschinellen Operationsformen und menschlichen Handlungsformen diese von ihren Anwendern als menschlichen Handlungs- bzw. Interaktionspartnern äquivalente 'Gegenüber' anerkannt und in jeweilige Arbeitszusammenhänge integriert werden (ähnlich auch Suchman ebd.). Es sind weniger die faktisch beschränkten Imitations- bzw. Substitutionsfähigkeiten der KI-Systeme als vielmehr die deren 'Schwächen'

kompensierenden nichtbewußten Anpassungs- und Reparaturleistungen der User, die dem ‘Funktionieren’ von Mensch-Maschine-‘Interaktionen’ auf der vermeintlichen Grundlage von Computerprogrammen als dem Mensch vergleichbare Akteure zugrundeliegen. "It is our ready willingness to repair such deficiences that allows current computers to work with us. It is the invisibility of this repair work - because of its pervasiveness even in ordinary speech - that makes usuable to mistake machines of humans."(Collins ebd.,S. 211)

2.2.4. Die Grenzen der KI

Die interpretative Soziologie im allgemeinen und die Ethnomethodologie im besonderen betonen die interpretativen, praktischen und unbewußten Anpassungs- und Reparaturleistungen sozialer Akteure, die letztendlich im Sinne eines ‘Mehr oder weniger’ gleichermaßen die - insbesondere in der pragmatistischen Philosophie unterschiedenen (vgl. Malsch 1998, S.290ff.) - Handlungskontexte bzw. -phasen eines impliziten routinisierten (gewohnheitsmäßigen) Problemlösens und eines neue Problemlösungen hervorbringenden Innovations- bzw. Gestaltungshandeln konstituieren. Den offenen und spontanen Charakter des Wechselspiels von einem soziale Strukturen eher reproduzierenden Routinehandeln und einem Strukturen eher aktiv produzierenden ‘Gestaltungshandeln’ hat Lucy Suchman (ebd.) im Anschluß an die Theorie von George Herbert Mead erklärt. Im Sinne eines quasi-dialogischen (Joas) oder eines quasi-dialektischen (Wagner 1992) Verhältnisses des ‘Me’ als der Gesamtheit aller gesellschaftlichen, potentiell regelorientiert bzw. routinisiert umgesetzter Verhaltenserwartungen und dem ‘I’ als einer je kontextuell bedingten und spontan umgesetzten ‘Kreativitätskomponente’ verweist sie auf ‚Situiertheit‘ als zentrales Merkmal sozialen Handelns." One kind of activity is an essentially situated and ad hoc improvisation - the part of us, so to speak, that actually acts. The other kind of activity is derived from the first, and includes our representations of action in the form of future plans and retrospective accounts." (Suchman ebd, S.51) Hieran anschließend muß von dem reifizierenden Charakter sowohl von Explikationen sozialen Handelns in Form jenes leitender Akteurskonzepte (Intentionen, Pläne, etc. z.B. als Selbstbeschreibungen) als auch von entsprechenden sozialwissenschaftlichen Modellierungs-, Formalisierungs- und Prognoseversuchen ausgegangen werden. "Consequently, our descriptions of our actions come always before and after the fact, in the form of imagined projections and recollected reconstructions." (Suchman ebd., S.51) Auf diesem offenen, eigensinnigen bzw. ‘unscharfen’ Charakter sozialer Handlungsprozesse beruhen, so Suchman, die Grenzen nicht nur von sozialwissenschaftlichen Modellierungsversuchen, sondern auch von jeglichen Versuchen der Imitation und Substitution sozialen Handelns durch regeldeterminierte Computerprogramme.

Das von Collins und Suchman vertretene ethnomethodologische bzw. interpretative Handlungs- und Interaktionsverständnis sowie die aus ihm abgeleitete Bewertung der Möglichkeiten und Grenzen der KI integrieren sowohl Elemente der ‘Hollow-Shell’- als auch der ‘Poor-Substitute’-Kritik. Die mehr oder weniger aktiven und kreativen Anpassungs-, Reparatur- bzw. Gestaltungsleistungen der Akteure rekurrieren auf Computern grundsätzlich nicht zugängliche soziokognitive Sinnzusammenhänge, wie insbesondere Collins (ebd.) mit seiner - sich von der phänomenologischen KI-Kritik abgrenzenden und in diesem Sinne originär soziologischen - Vorstellung einer ‘social

embeddedness' sozialer Wissens- und Handlungsformen betont (vgl. oben). Entsprechend können gemäß dieser Position Computerprogramme weder im Sinne der soziologischen Handlungstheorie handeln noch im Sinne des symbolischen Interaktionismus interagieren, insofern ihnen weder die Fähigkeiten einer (sozio-)kognitiv vermittelten interpretativen Sinnkonstruktion und -rezeption noch die auf jenen aufbauenden handlungsförmigen Kompetenzen einer mehr oder weniger aktiven und kreativen Konstruktion und Reproduktion sozialer Strukturen zukommen. In anderen Worten: Insofern nur menschliche Akteure an symbolisch (semantisch) konstituierten 'social worlds' teilnehmen können, können auch nur sie - unter Rückgriff auf sozialisatorisch erworbene und durch soziale Partizipation kontinuierlich aktualisierte Wissensformen und Praktiken - neue und sozial sinnvolle bzw. 'eingebettete' Problemlösungen hervorbringen. Der auf pragmatistischen Grundannahmen beruhende Hinweis auf vorwiegend nichtbewußt routinisierte, gewohnheitsmäßig und/oder regelorientiert vollzogene Handlungsformen kann zwar erklären, warum Computerprogramme von ihren Anwendern in spezifischen Handlungskontexten als vermeintlich dem Menschen vergleichbare Handlungs- bzw. Interaktionspartner akzeptiert werden (vgl.Rammert 1998, S. 113). Er rechtfertigt allerdings nicht - so schlußfolgert Collins aus seiner Analyse der 'behaviour specific - Erwartungen bzw. Einschätzungen einer vollständigen bzw adäquaten Imitation und Substitution menschlicher Leistungen durch KI-Systeme (dies legt m.E. auch die analoge Analyse eines 'reflexive monitoring of action' bei Giddens nahe, vgl. 3.5.3.3.).

Diese Einwände werden von Collins und Suchman in Auseinandersetzung mit den regeldeterminierten KI-Programmen des symbolischen Paradigmas entwickelt, müssen aber nach Wolfe (ebd.) auch angesichts konnektionistischer KI-Systeme des subsymbolischen Paradigmas aufrechterhalten werden. Zwar gesteht Wolfe den konnektionistischen Systemen im Unterschied zu den Programmen des symbolischen Paradigma zu, daß diesen Lernpotentiale im Sinne der Modifikation ihrer Operationen angesichts uneindeutiger Informationen, d.h. die Fähigkeit der Anpassung an eine sich dynamisch verändernde Umwelt zukommt. "The machine - more accurately (...) a set of parrelell machines - can 'learn', because it can react to ambigious or incomplete instructions." (ebd., S. 1085) Allerdings dürfen diese passiven Anpassungsprozesse nicht mit sozialen Lernprozesssen verwechselt werden. Nach Wolfe ist die sozialisatorische Entwicklung im allgemeinen wie auch soziales Lernen im besonderen durch symbolvermittelte Interaktionsbeziehungen gekennzeichnet, innerhalb derer die Teilnehmner als 'Lernende' nicht nur Verhaltensformen z.B. als Regelvorgaben eines Instrukteurs passiv und adaptiv rezipieren, sondern auch aktiv und kreativ um- bzw. neugestalten. Sowohl von außen 'überwachte' als auch 'unüberwachte' Prozesse maschinellen Lernens konnektionistischer Systeme (vgl. 2.1.2.) unterscheiden sich demnach von sinnhaft konstituierten sowie mehr oder weniger aktiv gestaltend vollzogenen sozialen Lernprozessen. Weder die "rule following programmes" der symbolischen KI noch die "rules excepting rogrammes" der subsymbolischen KI, so Wolfe, können die sozialem Handeln im allgemeinen und sozialem Lernen im besonderen zugrundeliegenden sinnhaft-kognitiven Fähigkeiten eines "meaning producing" und die (jene voraussetzenden) handlungsförmigen Kompetenzen eines innovativen "rule making" erlangen (ebd.).

Der hier vorgestellte sozialtheoretische, gleichermaßen ontologisch und epistemologisch argumentierende Nachweis der Grenzen der Leistungsfähigkeit von KI-Systemen

unterscheidet sich von solchen KI-Reflexionen, die Intelligenz von technischen Systemen im Anschluß an den philosophischen Pragmatismus (Dewey) oder im Rahmen der sozialkonstruktivistischen Wissenschafts- und Techniksoziologie (vgl. Latour 1991, 1998, Pickering 1993) ausschließlich als eine von den jeweiligen Systemnutzern zugeschriebene Qualität bzw. Eigenschaft verstanden wissen wollen (zusammenfassend Rammert 1998a, insb. S. 115ff., vgl. auch 4.2.) Im Rahmen der KI-Debatte wird jene Position insbesondere von Daniel Dennett (1987) vertreten, der in Auseinandersetzung mit der Position von Searle dafür plädiert, Intelligenz nicht als eine essentielle Eigenschaft, sondern vielmehr als eine Zuschreibungskategorie der Anwender zu verstehen. Dennett geht davon aus, daß es sich für menschliche Anwender angesichts von komplexen, nichtdurchschaubaren und/oder nichtprognostizierbaren Systemen - nicht-trivialen Systemen im Sinne von von Foerster - als pragmatisch sinnvoll darstellt, nicht auf ein notwendigerweise unvollständiges Wissen über deren interne Eigenschaften und Prozesse zurückzugreifen, sondern vielmehr einen 'intentionalen Standpunkt' einzunehmen, d.h. ihnen Intentionalität (Absichten und Ziele) sowie potentiell andere kognitive Eigenschaften (Überzeugungen, Wünsche, 'freier Wille', etc) zuzuschreiben. Auch wenn eine solche pragmatistisch inspirierte Perspektive für die empirische Analyse jeweiliger Mensch-Maschine-'Interaktionen' sicher aufschlußreich ist (vgl. 4.2.), bleibt sie m.E. auf die Teilnehmerperspektive der Systemanwender beschränkt. Entsprechend kann sie nicht deren - nur aus der Beobachterperspektive zugänglichen und m.E. für die soziologische KI-Analyse zentralen - Anpassungs-, Reparatur- und Gestaltungsleistungen der User berücksichtigen. Eine solche Vernachlässigung der Kompetenzen und Potentiale menschlicher Akteure, die entsprechende 'Unterstellung' von Intentionalität sowie die anschließende 'Illusion' einer dem Menschen vergleichbaren künstlichen Intelligenz stellen sich zwar nicht notwendigerweise als problematisch dar, wenn sie den Einschätzungen der jeweiligen User, wohl aber, wenn sie den Konzepten der Konstrukteure von KI-Systemen zugrundeliegen. In Auseinandersetzung mit Dennett verweist Bradshaw (1997, S. 5) mit der Unterscheidung von "ascription" (Zuschreibung) und "description" (Beschreibung bzw. Definition) darauf, daß im Rahmen der KI nicht nur User den Systemen Intentionalität zuschreiben, sondern auch (zumindest) eine Reihe von Systemkonstrukteuren bzw. Programmierer(n) bei der Definition, Modellierung und Implementation ihrer KI-Systeme diesen Intentionalität oder andere, dem Menschen vergleichbare Eigenschaften unterstellen. Wenn aber - wie ein sozialkonstruktivistisches Technikverständnis betont (z.B. Akrich 1992) - in die Definition und Konzeptualisierung von Technologien immer soziale Vorstellungen auch hinsichtlich des Verhaltens der zukünftigen Nutzer eingehen und folglich deren Implementationen weniger als die Installation technischer Artefakte als vielmehr die Konstitution soziotechnischer Netzwerke (Latour) verstanden werden müssen, ist ein verkürztes bzw. reduktionistisches Bild menschlicher Nutzer nicht unproblematisch. Vielmehr kann jenes - insbesondere im Zusammenhang der Forderung 'sozialadäquater' bzw. 'sozialverträglicher' Technologien – ein Verfehlen der Bedürfnisse der Nutzer und letztlich auch das Scheitern (die Nichtakzeptanz) der Programme in ihren Anwendungskontexten nach sich ziehen. Ob und inwieweit solche Einwände gegenüber einem Reduktionismus der KI auch angesichts der Ansätze der VKI und MAS, die die heute avanciertesten Formen des KI-Forschung darstellen und im Unterschied zu den Konzepten der symbolischen und subsymbolischen

KI explizit Überlegungen soziologischer Theorie rezipieren, aufrechterhalten werden müssen, soll im folgenden untersucht werden.

3. Verteilte künstliche Intelligenz und Multiagentensysteme

3.1. VKI und MAS: Der soziologische 'turn' der KI

Seit Anfang der 1980er Jahre rückt ein drittes Paradigma der KI-Forschung, die 'verteilte künstliche Intelligenz' (VKI) und deren Unterbereich der 'Multiagentensysteme', in den Blickpunkt (zusammenfasssend Müller, H.J. 1993).[2] Die VKI stellt dem gleichermaßen dem symbolischen und dem subsymbolischen Paradigma der KI zugrundeliegenden 'individualistischen' Fokus auf einzelne und omnipotente Problemlöseeinheiten die Vorstellung von kollektivem Problemlösen als der Leistung mehrerer, je spezifische Teilaufgaben bearbeitender Softwareentitäten gegenüber. Ausgangspunkt der VKI ist hierbei die Frage, wie angesichts mehrerer - funktional, räumlich oder zeitlich - verteilter Einzelkomponenten deren Koordination oder in einem stärkeren Sinne deren Kooperation bei der kollektiven Bearbeitung jeweiliger Aufgaben hergestellt werden kann. Für einen solchen Perspektivenwechsel der KI-Forschung können mehrere Gründe angeführt werden (Nwana 1997, S. 8ff.). Erstens stellt die Kombination vieler einzelner Softwareentitäten Lösungspotentiale bereit, die über die Summe der Möglichkeiten der Einzelentitäten hinausgehen, so daß jetzt auch ausschließlich kollektiv zu bewältigende Aufgaben bearbeitet werden können. Zweitens umgehen verteilte Systeme nicht nur die Problemlösungs- und entsprechenden Ressourcenbeschränkungen, sondern auch die Ausfallrisiken zentral organisierter und folglich auf das Funktionieren spezifischer einzelner Softwareentitäten angewiesener Systeme. Drittens können durch verteilte Systeme einfachere und oft auch schnellere Lösungswege für grundsätzlich auch durch zentral organisierte Systeme lösbare Probleme bereitgestellt werden. Darüber hinaus sind verteilte Systeme in der Lage, bestehende Systeme - z.B. Expertensysteme oder neuronale Netze der traditionellen KI - zu integrieren und somit deren Lösungspotentiale zu synthetisieren.

2. Der Begriff 'Distributed artificial intelligence' bzw. 'Verteilte künstliche Intelligenz' wurde in den USA geprägt, wo sich diese Forschungsrichtung ab Mitte der 1980er Jahre als ein eigenständiges Forschungsgebiet der KI etablierte (vgl. u.a. Huhns (ed.) 1987, Bond/Gasser (ed.) 1988, Avouris/Gasser 1992, Gasser/Huhns (ed.) 1989, O'Hare/Jennings (ed.) 1996). In Europa institutionalisierte sich die VKI in Form der 'Modelling autonomous agents in a multiagent world' (MAAMAW)-Konferenzen (vgl. u.a. Demazeau/Müller (ed.) 1990, Castelfranchi/Werner (ed.) 1994, Castelfranchi/Müller (ed.) 1995). Interessant für diesen Bereich sind auch die internationalen ‚Agent theories, architectures and languages' (ATAL) -Workshops (vgl. u.a. Wooldridge/Müller/Tambe 1995,Wooldridge/Müller/Jennings 1997). Der ersten (und bisher einzige) Sammelband im deutschen Sprachraum ist Müller, H.J. (Hg) 1993. Die Forschung der Multiagentensysteme (MAS) stellt einen Unterbereich der VKI-Forschung dar, der sich in den letzten Jahren durch seinen - durch das begriffliche Leitkonzept 'autonome Agenten' nahegelegten (vgl. 3.2.2.) - Bezug auf soziologischen Theorien oder Sozialmetaphern von der allgemeinen VKI absetzt (so Strübing 1998, S. 59, Fn. 2).

Den Startpunkt der VKI-Forschung markiert das von Hewitt (1977, vgl. 1991) eingeführte ‘Concurrent Actor’- Modell. Dieses versteht einzelne teilproblemlösende Softwareeinheiten als heterogen (parallel und asynchron) ausgerichtete und teilautonome ‘Actoren’, die durch die Übertragung spezifischer Informationen (‘message passing’) kommunizieren und im Zuge eines kontinuierlichen Austausches ihrer Teilergebnisse kollektive Problemlösungen hervorbringen. Dem Actor-Konzept liegt die sogenannte ‘scientific community metaphor’(Kornfeld/Hewitt 1988) zugrunde, insofern hier verteilte Systeme nach dem Vorbild der sozialen Organisationform der Wissenschaft, in der lokal gewonnene (Teil-) Ergebnisse vermittels Publikationen translokal verbreitet und kollektiv bewertet werden, modelliert werden. Es sind folglich weder die Eigenschaften und Leistungen der Einzelactoren noch eine übergeordnete zentrale Kontrollinstanz, sondern vielmehr die durch Kommunikation gewährleisteten Abstimmungen und Synthesen der Teilergebnisse, die die kollektive Problembearbeitung innerhalb von als umweltoffen und dynamisch (entwicklungs- bzw. lernfähig) konzipierten Actoren-Gesamtsystemen garantieren (Hewitt 1991).

Der Erfolg bzw. die kollektive Systemintelligenz solcher verteilten offenen Systeme, für die es soziale Vorbilder über die Organisationsform von ‘scientific communities’ hinaus auch im Bankennetz, im Bibliothekswesen oder in der dezentralen Funktionsweise des ‘World Wide Web’ gibt, kann nicht wie im Turing-Test anhand computationaler Einzelleistungen bewertet werden. Der von Star (1989, vgl. Strübing 1998, S. 79/80) im Rahmen der VKI eingeführte ‘Durkheim-Test’ fokussiert entsprechend auf die systemische Problemlösungskapazität bzw. die ‘kollektive Intelligenz’ von jeweiligen Gesamtsystemen, die im Sinne von ‘Hybridsystemen’ nicht nur eine Vielzahl von computationalen Einheiten, sondern auch eine Vielzahl von menschlichen Nutzern umfassen.[3] Allerdings geht Star davon aus, daß das Gesamtsystem weder anhand der Leistungen seiner computationalen Einzelkomponenten noch - aufgrund der unterschiedlichen und potentiell divergierenden Perspektiven und Interessen der einzelnen Nutzer - hinsichtlich seiner kollektiven Systemintelligenz eindeutig bewertet werden kann. Demgegenüber schlägt sie aus einer pragmatistischen Perspektive vor, empirisch die Bewertungen der Nützlichkeit bzw. der ‘sozialweltlichen Brauchbarkeit’ der Effekte jeweiliger Innovationen zu untersuchen, wie sie aus den unterschiedlichen Perspektiven sowohl der Systementwickler als auch der Systemanwender vorgenommen werden. Jene voraussichtlich divergierenden Bewertungen sollen nach Star dann wiederum im Sinne einer ‚partizipativen Technikentwicklung‘ in - möglichst gemeinsam zu vollziehende - Systemgestaltungs- bzw. Systemverbesserungs-prozesse eingebracht werden.

3.2. Mikro- und Makromodelle in der VKI

3. Malsch (1998, S. 284/285) weist darauf hin, daß hier der Begriff ‘Durkheim-Test’ hinsichtlich der Bewertung von kollektiver Intelligenz in die Irre führt, insofern kollektive Intelligenz bei Star gerade nicht als eine kontextunabhängige Qualität im Sinne Durkheimscher (emergenter) ‘Sozialer Tatsachen‘ verstanden wird, sondern vielmehr aus den unterschiedlichen, je kontextuellen Perspektiven der Nutzer bewertet werden soll.

Der Modellbildung der VKI setzt ähnlich wie soziologische Analysen gesellschaftlicher Zusammenhänge an der Bestimmung der Formen und Prozesse der Koordination bzw. Kooperation ‘individueller’ Einzelentitäten, d.h. insbesondere an der als soziologisches Mikro-Makro-Problem bekannten Frage nach der Hervorbringung globaler Ordnungszusammenhänge aus lokal konstituierten Ereignissen an (vgl. Rammert 1998, S. 110ff.). Analog zum soziologischen Paradigmenstreit zwischen Mikro- und Makrosoziologie hat sich auch die VKI im Zuge der Weiterentwicklung des Hewittschen Actorenmodells in zwei, ihren Gegenstandsbereich entweder makrotheoretisch ‘top down’ oder mikrotheoretisch ‘bottom up’ modellierende Forschungsrichtungen aufgespalten (Brenner u.a. 1998, S. 39ff., vgl.Wooldridge/Jennings 1995b).

3.2.1. Makromodelle in der VKI: ‘Schwarze Bretter’ und ‘Vertragsnetze’

Top-Down-Modellierungen der VKI setzen an den übergreifenden Ordnungsstrukturen des Gesamtsystems an, die gleichermaßen die Aufgabenverteilung an jeweilige - hier oft nicht als Agenten, sondern als Knoten bzw. Module bezeichnete - Einzelkomponenten wie auch die anschließende Synthese der Teillösungen bestimmen (zusammenfassend Kirn 1995, Avouris/Gasser 1992b). Im Gegensatz zu den eher heterogen organisierten Multiagentensystemen (vgl. unten) werden zu Beginn des Systementwurfs alle Teilaufgaben eindeutig defininiert und im Systemverlauf durch spezifische Zentral- bzw. Kontrollinstanzen auf die eine oder andere Weise koordiniert, so daß - im Sinne einer sogenannten ‘benevolent assumption’ - die ‘gutwilligen’ Einzelkomponenten ihr lokales Operieren ausschließlich an der Erreichung des bzw. der globalen Ziele des Gesamtsystems ausrichten. Die einfachste Form der Koordination der Einzelkomponenten ist eine fest vorgebene und nicht veränderbare hierarchische ‘Organisationsstruktur’, die zwischen Aufgaben delegierenden Komponenten - sogenannten ‘master agents’ - und aufgabenausführenden Einheiten - sogenannten ‘slave agents’ - unterscheidet. Ein etwas differenzierteres hierarchisches Organiationsmodell findet sich bei sogenannten ‘Blackboard-Systemen’ (Lesser/Corkhill 1983, Hayes-Roth 1988), in denen mehrere Problemlöser als einzelne, im Anschluß an die symbolische KI als regelbasierte bzw. expertensystemartige modellierte, Wissensquellen (Knowledge Sources) über eine globale bzw. zentrale Datenstruktur - das sogenannte ‘schwarze Brett’ - miteinander verbunden werden. Das schwarze Brett hat einerseits die einem Protokoll bzw. einer wissenschaftlichen Publikation vergleichbare Aufgabe, den bisherigen Verlauf des Problemlösungsprozesses zentral zu dokumentieren. Andererseits kommt dem durch ‘master agents’ verwalteten schwarzen Brett die Funktion zu, angesichts mehrerer geeigneter Problemlöser die Aufgabenverteilung zu kontrollieren, d.h. aufbauend auf dem gesamtsystemischen Wissen die jeweiligen Aufgaben an die jeweils geeignetesten ‘slave agents’ zu delegieren. Schwachpunkt dieses Ansatzes ist allerdings in Analogie zu hierarchischen bzw. zentralistischen Organisationformen der traditionellen KI das Problem insbesondere in den Ordnungs- bzw. Kontrollinstanzen (‘master agents’) auftretenden Ressourcenengpässen oder gar Systemausfällen.

In Kontrakt- bzw. Vertragsnetzsystemen (Davis/Smith 1983, Smith/Davis 1988) organisieren wie innerhalb von Blackboardsystemen zentrale Module - hier: ‘manager agents’ - die Aufgabenverteilung zwischen verschiedenen, hier als ‘contractor agents’

bezeichneten, Teilproblemlösern. Diese Aufgabenverteilung verkörpert einen marktähnlichen Mechanismus, da sie sich an dem Prinzip der Aushandlung des 'besten Angebots' von, um die Bearbeitung von Teilaufgaben konkurrierenden, 'contractor agents' orientiert. Vertragsnetzsysteme unterscheiden sich somit von Blackboardsystemen durch ein 'Mehr' an Flexibilität, das sowohl auf einer effizienteren Nutzung von Ressourcen als auch auf der Minimierung von Ressourcenengpässen und Systemausfällen beruht. Auf der einen Seite stehen für eine Aufgabe stets mehrere Agenten zur Verfügung, so daß stark beanspruchte Agenten sich nicht um die Vergabe eine Aufgabe 'bewerben ' oder nicht notwendigerweise bei der Vergabe berücksichtigt werden müssen. Auf der anderen Seite können im Gegensatz zu der strikteren Hierarchie der Blackboardsysteme im Systemverlauf der Vertragsnetzsysteme 'contractor agents' zu 'manager agents' werden, insofern sie übernommene Aufgaben weiter aufteilen oder als Ganze weiter delegieren. Allerdings steht einer solchen im Unterschied zu den Blackboardsystemen gewonnenen Flexibilität des Gesamtsystems der erhöhte Bedarf an Verhandlungsleistungen, d.h. die Notwendigkeit eines kontinuierlichen - sich auf die benötigten Rechnerkapazitäten niederschlagenden – Informationsaustausch zwischen den verhandelnden Agenten gegenüber.[4]

3.2.2. Das Agentenparadigma der MAS

'Top- down'-Modellbildungen der VKI setzen vorrangig an Einzelentitäten bzw. 'Agenten' als den Trägern globaler Ordnungs- bzw. Strukturvorgaben an, wobei deren effektives bzw. 'harmonisches' Zusammenwirken im Falle der Blackboardsysteme durch einen zentralen Wissensspeicher als Organisationsinstanz und im Fall der Vertragsnetzsysteme durch eine übergreifende quasi-ökonomische (Aushandlungs-)Logik bestimmt wird. In Abgrenzung zu dieser Auffassung von Einzelmodulen als Träger von übergreifenden Systemvorgaben bzw. -zielen gehen Bottom-Up-Modellbildungen im Rahmen des Agentenparadigma (MAS) zunehmend auch von eigenständigen, von den kollektiven Orientierungen abweichenden und potentiell konfligierenden Orientierungen und Verhaltensformen von Einzelmodulen aus (zusammenfassend Moulin/Chaib-Draa 1996, Bradshaw 1997). Im Zuge des Fokus auf einen heterogenen, autonomen und potentiell antagonistischen Charakter von Agenten steht weniger die vorgängige, mehr oder weniger strikte Programmierung einer effektiven Arbeitsteilung sowie entsprechender gesamtsystemischer Ordnungsmuster, sondern vielmehr Möglichkeiten der Bearbeitbarkeit von - zum Zeitpunkt der Systemgestaltung - noch nicht bekannten Problemen sowie die Modellierung entsprechender Kompetenzen der Einzelagenten im Vordergrund. Aber welche Leistungen müssen computationale Systeme -

4. Im Gegensatz zu Blackboard- und Vertragsnetzsystemen gehen Ansätze der Multiagentenplanung ('multi-agent planning') davon aus, daß jeder einzelne Agenten 'alleine' einen Plan zur Bearbeitung der jeweiligen Gesamtaufgabe erstellt. Im Falle eines 'centralised multi-agent planning' werden diese Pläne wiederum einer zentralen Kontrollinstanz vorgelegt, während im Falle eines 'decentralised agent planning' (entsprechend den unten genannten spieltheoretischen Ansätzen) jeder einzelne Agent seine Vorgehensweise mit allen anderen Agenten abgleichen muß (Spieltheorie). Im ersten Fall kann entsprechend wiederum das Problem von Ressourcenengpäßen und Systemausfällen auftreten, während im zweiten Fall sich als Hauptproblem der hohe Aufwand an Informationsaustausch bzw. 'Kommunikation' zwischen den Einzelentitäten darstellt (vgl. von Martial 1993).

vorwiegend Roboter oder Softwareagenten[5] - verkörpern, damit ihnen ein Agentenstatus zugeschrieben werden kann?

Franklin/Grasser (1997) betonen die wahrnehmungsförmigen, effektorischen und umweltsensiblen - gleichermaßen anpassungs- wie auch gestaltungsfähigen - Leistungen von Agenten innerhalb einer sich fortlaufend verändernden und/oder gestaltend veränderten Umwelt: "An agent is a system situated within and part of an environment that senses that environment and acts on it, over time, pursuit of its own agenda and so act to effect what it senses in the future". Demgegenüber betont Bradshaw (1997, S. 7) die in Franklin/Grassers Agentendefinition nicht im Vordergrund stehende Eigenständigkeit bzw. Autonomie der Agenten (vgl. Wooldridge 1996, S. 47ff.). So bestimmt er Agenten als eine "software entitiy which functions continuously and autonomously in a particular environment, often inhabited by other agents and processes." (Bradshaw ebd., S. 7) Die Annahme eines kontinuierlichen und autonomen ‘Funktionierens’ - so erläutert Bradshaw - impliziert erstens die auch von Franklin/Grasser hervorgehobene Flexibilität bzw. Anpassungsfähigkeit der Agenten, insofern diese ihr Operieren sinnvoll an einer sich fortlaufend verändernden Umwelt ausrichten können. Zweitens sind Agenten autonom, insofern sie nicht notwendigerweise auf vorgängige Eingaben der Programmierer oder auf das später korrigierende Eingreifen der Nutzer angewiesen sind. Darüber hinaus kommt nach Bradshaw Agenten die Fähigkeit zu einem maschinellen Lernen (vgl. 3.3.) sowie eine spezifische ‘soziale’ und (potentiell) räumliche Positionierung innerhalb eines Agentensystems und/oder einer jeweiligen Umwelt zu. "An agent that functions continously in an environment over a long period of time would be able to learn form its experience. In addition, we expect an agent that inhabits an environment with other agents and processes to be able to communicate and cooperate with them, and perhaps move from place to place in doing so."(ebd.)

Eine noch differenziertere Position als Bradshaw vertreten Wooldridge/Jennings (1995a, S. 116ff.), die zwischen ‘schwachen’ und ‘starken’ Agentendefinitionen unterscheiden. Gemäß ihrer ‘schwachen’ Agentendefinition kommen Agenten zumindest die Eigenschaften der Autonomie, der Sozialität bzw. Sozialfähigkeit (‘social abiltiy’), der Reaktivität (‘reactivtiy’) und der Pro-Aktivität (‘pro-activeness’) zu. Diese Agenteneigenschaften stimmen auf den ersten Blick weitestgehend mit denen von Bradshaw überein. Allerdings wird hier hinsichtlich der Sozialität von Agenten ihre über eine Agentensprache vermittelte Kommunikationsfähigkeit und hinsichtlich dem effektorischen Operieren ihre Fähigkeit zu einem initiativen, d.h. zu einem aktiv zielorientierten Verhalten und zu entsprechenden Verhaltensselektionen (vor dem Hintergrund mehrerer zur Verfügung stehender Verhaltensoptionen) betont. "Agents do no simply act in response to their environment, they are able to exhibit goal-directed behaviour

5. Einen Versuch deren Klassifikation unternehmen Franklin/Graeser (1997) Sie unterscheiden zwischen a) Biologischen Agenten b) Robotern c) Softwareagenten. Letztere umfassen ‘Artificial Life Simulationen’, ‘task specific agents’,’entertainment agents’ und Computerviren. Im Rahmen der aufgabenspezifischen Agenten stehen heute die autonom und meist mobil recherchierenden Informationsagenten des WWW und die sich jeweiligen Nutzerprofilen anpassenden Interface-Agenten im Blickpunkt. Insbesondere die Entwickung sogenannter ‘Personal digital assistents’ (PDA), die im Zuge der Übernahme des jeweiligen Nutzerprofils eine Vielzahl von Aufgaben (E-Mail-Sortieren, Terminabsprache, etc.) übernehmen sollen, haben für Furore gesorgt.

by taking the initiative."(ebd., S. 116) Während diese ‘schwache’ Agentendefinitionen den Vergleich mit menschlichen Akteuren nicht notwendigerweise implizieren, steht demgegenüber bei ‘starken’ Agentendefinitionen in der Tradition der KI-Forschung die Orientierung an der Modellierung und Implementation menschlicher Eigenschaften und Leistungen im Vordergrund. So schreiben jene Agenten menschlichen Akteuren analoge mentale bzw. kognitive (im Sinne reflexiver Agenten, vgl. 3.4.1.) oder normative (im Sinne sozialer Agenten, vgl. 3.4.2.) Merkmale und Kompetenzen zu. Ein solches ‘starkes’ Agentenverständnis findet sich auch bei der Modellierung von ‘emotionalen Agenten’, deren Operationsweisen über kognitive Prozesse hinaus auch auf ‘menschenähnlichen’ Gefühlzuständen beruhen sollen (vgl. z.B. Bates 1994, Maes 1994b). Darüberhinaus spielen für einige Agentenkonzepte die Modellierung von Charakter- bzw. Verhaltenseigenschaften wie z.B. Vertrauenswürdigkeit (benevolence), Ehrlichkeit (veracity) oder auch Diskretion eine wichtige Rolle.[6]

6. Ein wichtiges Unterscheidungsmerkmal von Agenten ist deren von Bradshaw betonte Mobilität. Mobile Agenten kennzeichnen nicht nur notwendigerweise den Bereich der Robotik, sondern stellen insbesondere im Bereich der Softwaresysteme eine wichtige Innovation dar. Während stationäre Agenten bei der Kooperation mit anderen Agenten auf Versenden von Nachrichten bzw. auf ‘Kommunikation’ angewiesen sind, können sich mobile Agenten innerhalb eines jeweiligen Rechnernnetzwerkes bewegen und ‘vor Ort’ - auf einem anderen Rechner - ihre jeweiligen Aufgaben bearbeiten.Vorteile hierbei ist nicht nur die Minimierung des Koordinations- bzw. Kommunikationsaufwand, sondern auch, daß solche Agenten - im Sinne einer zusätzlichen Spezifikation ihrer Autonomie - im Fall des ‘Offline’ ihres ‘Heimatrechners’ und/oder ihrer Nutzer weiter arbeiten können. Mobile Agenten stellen insbesondere im Rahmen der Informationsagenten (Suchmaschinen wie z.B.’Web Crawler’) eine wichtige Innovation dar.

3.3 Maschinelles Lernen

Ein bzw. das entscheidende Charakteristikum von Agenten gegenüber herkömmlichen Computerprogrammen im allgemeinen und den Produkten der traditionellen KI im besonderen ist ihre erweiterte Fähigkeit, sich an jeweilige Umweltbedingungen anzupassen und in diesem Sinne zu ‘lernen’. So sind insbesondere die sogenannten ‘Interface Agenten’ z.B. in Form sogenannter PDA’s (‘Person Digital Assistents’) in der Lage, sukzessiv die Ziele, Interessen bzw. Präferenzen der User zu übernehmen und somit ihr Verhalten an diese anzupassen (Maes 1994a, Negroponte 1997, Braun 1998). Nach Green et. al. (1997, S.7 und ff.) werden im Rahmen der VKI und MAS über die Verfahren der "symbolischen Klassifikation" regelbasierter KI-Systeme (vgl. 2.1.1.) und der "subsymbolischen Klassifikation" konnektionistischer KI-Systeme (vgl. 2.2.2.) hinaus drei weitere Lernverfahren eingeführt, die ein - bisher nur im Falle ‘unüberwachter’ neuronaler Netze erreichtes - ‘selbständiges’ Lernen während dem Systemverlaufs in Form eines "Learning directly from the environment" ermöglichen.

Im ersten Fall eines "Reinforcement learning" (Verstärkendes bzw. Rekursives Lernen) passen sich die auf einfachen Wahrnehmungs-Aktions-Ketten beruhenden Systemoperationen vermittels eines spezifischen Lernalgorithmus und anschließender Selektionsstrategien einer jeweiligen Umwelt, insbesondere in Form des Verhaltens anderer Agenten oder des Benutzerverhaltens, an.[7] Im zweiten Fall eines "Learning by Observation" wird entweder das Verhalten der Nutzer - Maes spricht hier von einem "looking over the shoulder of the user" (Maes ebd., S. 33) - oder das Verhalten anderer, nicht als Instrukteure, sondern als gleichberechtigte ‘Partner’ verstandener Agenten direkt beobachtet und imitiert. Im dritten Fall eines "Instructional Learning" werden Agenten während des Systemverlaufs explizit durch spezifische kommunikative Anweisungen von ihrem Nutzer oder aber auch von anderen Agenten instruiert. Entscheidend ist, daß diese drei Formen eines direkten und selbständigen Maschinenlernens nicht notwendigerweise ‘höhere’ kognitive Fähigkeiten voraussetzen. Sie können entsprechend auch und vor allem von reaktiven Agenten (vgl. 3.4.2.), die nicht wie - in der Tradition der symbolischen KI stehende - reflexive bzw. deliberative Agenten (vgl. 3.4.1.) auf anspruchsvollen kognitiven Eigenschaften aufgebaut sind, umgesetzt werden. Kognitive Kompetenzen setzt demgegenüber das meist im Zusammenhang reflexiver Agenten genannte ‘fallbasierte Schließen’ ("Case based reasoning") voraus, daß ähnlich wie die symbolische Klassifikation spezifische Klassen unterscheidet, in die z.B. jeweilige Nutzer aufgrund ihrer Verhaltensweisen eingeordnet werden. Sobald ein Nutzer einer Klasse zugeordnet ist, kann sein potentiell veränderndes Verhalten auch die Schemata bzw. Stereotype der Klasse modifizieren, wobei hier im Unterschied zu den Formen eines direkten, d.h. vorrangig kurzfristigen Lernens eher langfristig Nutzerprofile und somit Anpassungen der Programmkategorien an die User erreicht werden können.

7. Maes (1994a, S. 34) spricht im Zusammenhang von Interface-Agenten von einem "Feedback"-Lernen, insofern diese hier implizit - oder auch explizit im Sinne von Anweisungen (vgl. unten) - die Verhaltensformen ihrer Nutzer übernehmen. Dieses Feedback-Lernen kann nach Maes die Form eines Beispiellernes bzw.’Lernens nach Vorbild’ annehmen, wenn Nutzer das angestrebte Verhalten exemplarisch demonstrieren und dieses infolgedessen von den Agenten übernommen wird.

3.4. Agentenarchitekturen

‘Bottom-Up’-Modellbildungen im Bereich der Multiagentensysteme setzen bei der Systemgestaltung an den Eigenschaften und Kompetenzen der Einzelagenten an, wobei sich in Form ‘reflexiver’ und ‘reaktiver’ Agentenarchitekturen zwei Forschungsparadigmen herausgebildet haben (vgl. Müller 1996).

3.4.1. Reflexive Agenten

Im Falle reflexiver bzw. deliberativer Agentenarchitekturen werden Agenten in enger Anlehnung an die Grundannahmen der symbolischen KI und deren Bild von (menschlichen) Akteuren als rationalen, alleswissenden bzw. omnipotenten Entscheidern bzw. Problemlösern mit symbolischen Repräsentationen der (Um-)Welt, Schlußfolgerungskompetenzen und weiteren für rationale Kalkulationsleistungen notwendigen kognitiven Merkmalen ausgestattet. Solche kognitiven Mermale lassen sich nach Wooldridge/Jennings (1995a, S. 120 und ff.) in zwei Kategorien, nämlich in ‘information attitudes’ in Form von Wissensüberzeugungen (‘beliefs’) und handlungsbeeinflussenden intentionalen ‘pro-attitudes’ wie z.B. Wünsche oder Absichten, unterteilen. Im Rahmen des Belief-Desire-Intention- (BDI-) Paradigmas (z.B. Rao 1996, Haddahi/Sundermayer 1996) werden genau drei kognitive Komponenten in Form von Wissensüberzeugungen hinsichtlich der Beschaffenheit der Welt (‘beliefs’), Wünschen als den Wertungen anzustrebender Systemzustände (‘desires’) und ‘intentions’ als aus der Gesamtheit der Wünsche als realistisch erreichbar abgeleiteter Ziele der Agenten voneinander abgegrenzt.[8]

Als zentraler Ausgangspunkt dieser Modellbildung stellt sich das Problem dar, daß intentionale Haltungen wie Überzeugungen, Wünsche oder Absichten - im Gegensatz zu den in der symbolischen KI im Vordergrund stehenden, abbildtheoretisch verstandenen Repräsentationen - grundsätzlich innerhalb einer formallogischen Sprache nicht eindeutig beschrieben werden können (Wooldridge/Jennings 1995a, S. 121ff., vgl. auch Werner 1996). Angesichts dessen greifen die Modellbildungen reflexiver Agenten meist auf das von Hintikka (1962) eingeführte ‘Possible Worlds’- Modell zurück. Dieses beschreibt ausgehend von einem bestimmten Jetzt-Zustand eines Systems und seiner Umwelt alle denkbar möglichen zukünftigen Systemzustände, wobei nur solche Merkmale, die gleichermaßen hinsichtlich aller zukünftigen ‘possible worlds’ angenommen werden können, als relevante kognitive ‘Überzeugungen’ (Wünsche, Absichten, etc.) angesehen und entsprechend formalisiert werden. Als Schwäche des ‘Possible-World’- Ansatzes erweist sich das ihm immanente ‘Logical omniscience problem’, demgemäß logische

8. Neuere Ansätze erweitern diese Bestimung dreier mentaler Zustände um die Komponenten Ziele (goals) und Pläne (plan). Ziele stellen hier die - im Rahmen der traditionellen BDI als Intentionen gekennzeichnete - Untermenge von Wünschen dar, die sich für den Agenten als grundsätzlich bearbeitbar, d.h. als realistisch oder nicht miteinander in Konflikt stehend - darstellen. Intentionen werden als eine Untermenge der Ziele verstanden, die der Agent im Anschluß an eine Priorisierung seiner Ziele und die Reflexion auf die benötigten Ressourcen auswählt. Plane werden entsprechend hier als Handlungsanweisungen in Folge konsistenter Zusammenfassungen der Intentionen verstanden.

Schlußfolgerungen aus Überzeugungen wiederum (nur) Überzeugungen, d.h. keine gesicherten Schlußfolgerungen nach sich ziehen. Dieses Problem und die ihm zugrundeliegende Annahme alleswissender Problemlöser führt - vielmehr analog zum 'frame problem' der symbolischen KI (vgl. 2.1.1.) - bei Systemen mit notwendigerweise endlichen Ressourcen zu unüberwindbaren Schwierigkeiten in Form der Unmöglichkeit der Kalkulation aller zukünftigen Systemzustände.

Dem reflexiven Agentenparadigma liegt ähnlich wie der symbolischen KI die Vorstellung von Agenten als rational kalkulierenden, letztlich allwissenden bzw. omnipotenten Problemlösern zugrunde. Hieran anschließend sind reflexive Agenten nur zu sehr beschränkten Reaktionen auf eine sich dynamisch verändernde Umwelt in der Lage, insofern sie ihre vorgängig programmierten kognitiven Eigenschaften aufbauend auf einem fallbasierten Schließen (vgl.3.3.) nur noch minimal oder nur langfristig aktualisieren können. Zwar sind sie flexibler als die Systeme der symbolischen KI, insofern ihre 'beliefs' partiell modifiziert und insbesondere im Zuge der Abgleichung von relativen stabilen Wünschen und eher modifizierbaren Intentionen (als den realistisch erreichbaren Wünschen) eine gewisse Anpasungsfähigkeit an die Umwelt erreicht wird. Schlußendlich sehen sich aber reflexive Agentensysteme ähnlich wie die 'Problemsolver' der traditionellen symbolischen KI der Schwierigkeit einer weitestmöglichen vorgängigen Explikation der für ihr Operieren und Entscheiden notwendigen Kognitionen und der Unmöglichkeit der Berechnung aller in Zukunft 'möglichen' Systemzustände gegenüber.

3.4.2. Reaktive Agenten

Einen Ausweg aus den Schwierigkeiten des reflexiven Agentenparadigmas bietet die Entwicklung 'reaktiver' Agenten, die auf die Modellierung anspruchsvoller kognitiver Merkmale und entsprechender rationalschlußfolgernder Kompetenzen verzichtet und vielmehr auf die Programmierung einfacher, flexibler und robuster bzw. fehlertoleranter Agenten abzielt (Brooks 1990, Steels 1990, vgl. zusammenfassend Maes 1990). Hierbei werden im Sinne einer 'intelligence without representation' (Brooks 1990) die verhaltensbasierten, auf rezeptorischen Wahrnehmungsfähigkeiten und effektorischem 'Bewirken-Können' beruhenden Anpassungsleistungen der Agenten in einer dynamischen Umwelt in den Vordergrund gerückt. Maes (1990, zitiert nach Moulin/Chaib-Draa 1996, S. 39) unterscheidet zwei Grundannahmen des reaktiven Paradigmas in Form einer 'task level decomposition' hinsichtlich des modulhaften Aufbaus der Einzelagenten und in Form einer 'emergenten Funktionalität' des Ordnungszusammenhangs eines gesamten, möglicherweise nur aus Agent und User bestehenden Agentensystems. Hinsichtlich des Aufbaus von Einzelagenten wird im Sinne einer von Brooks (1991) im Anschluß an Minsky (vgl. 1990) entworfenen 'subsumption architecture' eine - ohne symbolische bzw. kognitive Repräsentationen arbeitende - Schichtenarchitektur einer Vielzahl von autonom je eigene Aufgaben bearbeitenden Modulen angenommen. Diese autonomen Einzelmodule der Agenten - hier: von mobilen Robotern - sind je einzeln zuständig für die unterschiedlichen Teilaufgaben der sensorischen Wahrnehmung, der (allerdings im Gegensatz zu reflexiven Agenten eher begrenzten) internen Modellbildung und Kalkulation sowie der motorischen Kontrolle, während der Informationsaustausch zwischen ihnen auf ein Minimum beschränkt bleibt. Die Einzelagenten besitzen entsprechend weder ein umfassendes Bild

bzw. Modell ihrer Umwelt noch eindeutige, z.B. auf einer hierachischen Ordnung der Einzelmodule beruhende Ziele. Vielmehr entsteht bzw. ‘emergiert’ ihr Verhalten spontan bzw. situativ aus den Wechselwirkungen der ‘autonomen’, d.h. nicht global- bzw. zielorientierten Module: "The global behaviour of the agent is not necessarily a linear composition of the behaviours of its modules, but instead more complex behaviour may emerge by the interaction of behaviours generated by the individual modules." (ebd.)

Aber nicht nur der Modellierung des Aufbaues der Einzelagenten, sondern auch dem Verständnis der Verhaltensweisen und Interaktionen innerhalb eines Gesamtagentensystems liegt die Vorstellung einer Emergenz von nichtgeplanten übergreifenden Ordnungsformen zugrunde. Reaktive Agenten sind auch ohne vorgängige kognitive Kompetenzen zu einem funktionalen Operieren wie auch entsprechenden Selektionen innerhalb einer jeweiligen Umwelt in der Lage, insofern ihre je spezifischen Funktionalitäten implizit überhaupt erst in der Interaktion mit dieser Umwelt - den anderen Agenten, Usern und/oder anderen Umweltbedingungen des Gesamtagentensystems - hervorgebracht werden: "The functionality of an agent is viewed as an emergent property of the intensive interaction of the system with its dynamic environment." (ebd.) Gemäß der Vorstellung einer "emergent functionality" (ebd.) beruht letztlich auch die gesamtsystemische Koordination bzw. die entsprechende gesamtsystemische Intelligenz auf einem aus dem wechselseitigen Wahrnehmen und Agieren der Einzelagenten ‘emergierenden’ und für die Reproduktion des Gesamtsystem ‘funktionalen’ Abstimmungsmuster. Ein solches Konzept kann auf der einen Seite problemlos an einen auch dem Konnektionismus der traditionellen KI zugrundeliegenden ‘biologischen Holismus’ anschließen (vgl. Ferber 1996), der in Bezug auf unterschiedliche biologische Vorbilder (z.B. den Insektenstaat) Koordination bzw. ‘Sozialität’ als das Ergebnis negativer Rückkopplungsmechanismen zwischen subkognitiven (instinktgeleiteten) ‘Agenten’ erklären will. "Emergence of functionality and of stable states are produced by the joined forces of different feedback mechanisms. Positive feedback tends to create diversies among agents, whereas negative feedback regualtes societies, imposing a conservative force upon their social structures." (ebd. 1996, S. 290). Auf der anderen Seite rekurriert Maes im Anschluß an Steels (1990) weniger auf biologische als vielmehr auf quasi-soziale Selbstorganisationsprozesse. Ähnlich wie auch die Soziologin Karin Knorr-Cetina (1990)[9] versteht Maes Handlungsziele und die diesen zugrundeliegenden Handlungsselektionen als in Interaktionen emergierende Phänomene, die weniger auf kognitiven als vielmehr auf rezeptorischen und verhaltensbasierten bzw. praktischen Kompetenzen aufbauen und ‘in interactu’ hervorgebracht werden. Anschließend wird von der Selbstorganisation von Ordnungszusammenhängen in Form von aus den jeweiligen interaktionalen Wechselwirkungen emergierenden höherstufigen organisationalen bzw. globalen Systemmustern ausgegangen. "The organizational level is allowed to emerge from spontaneous self-organizing mechanisms acting among subcognitiv agents." (Conte/Castelfranchi 1996, S. ..) Im Rahmen eines solchen Modells wird - wenn ich Maes und Steels richtig verstehe - hinsichtlich der längerfristigen Selektion und Stabilisierung

9. So schreibt Knorr-Cetina: "Eine Umschreibung des Phänomens Emergenz kann lauten, daß soziale Episoden eine selbstorganisierende Struktur und ihre je eigenen Prozeßcharakteristiken haben. (...) Die theoretisch (und methodisch) relevante Einheit ist nicht die soziale Handlung, die soziale Beziehung oder gar das Individuum als Aktor oder Merkmalsträger." (1990, S. 143)

des Gesamtagentensystems von einer quasi-evolutionären und/oder quasi-funktionalen Selbststabilisierung ausgegangen, wobei gleichzeitig den einzelnen Agenten im Unterschied zu instinktgeleiteten biologischen Agenten Autonomie und Zielorientiertheit zukommt, insofern sie im Anschluß an spezifische Lernprozesse eigene, allerdings 'immer schon' in die Evolution des Gesamtsystems integrierte Ziele generieren können.

Die Beschränkung der Modellierung reaktiver Agenten ist, daß deren Zielgenerierung im Gegensatz zu reflexiven Agenten keinerlei anspruchsvolle kognitive und rationale Kompetenzen in Form von - z.B. für den Systemnutzer rekonstruierbaren - Plänen oder Konzepten zugrundeliegen. Entsprechend impliziert die Modellierung und Implementation reaktiver Agenten notwendigerweise eine unvorhersehbare und potentiell nicht gewünschte Systementwicklung. Trotz (oder: gerade aufgrund) der Gewährleistung spezifischer Lernmöglichkeiten wie z.B. einem 'verstärkendem' bzw. 'rekursiven' Lernen kann das Entstehen erwünschter Effekte nicht garantiert, sondern muß die Emergenz nicht erwünschter oder langfristig nichtfunktionaler Effekte miteinkalkuliert werden. Die Unvorhersehbarkeit der systemischen Entwicklung sowohl aus der Perspektive der Systemprogrammierer wie auch der Systemnutzer - die Möglichkeit eines 'out of control' reaktiver Agentensysteme (so Brooks ebd.) - stellt die zentrale Problematik reaktiver Agenten und/oder reaktiver Agentensysteme dar. Dieser Problematik liegt die in der Soziologie vieldiskutierte Schwäche von Vorstellungen einer ‚emergenten Funktionalität' - wie letztlich jeglicher Selbstorganisationskonzepten (vgl. entsprechend zu Luhmann 3.5.3.2.) – zugrunde. Jene können bei der Analyse von emergenten Ordnungen im allgemeinen und von spezifischen kollektiven Koordinations- und Problemlösungsmechanismen im besonderen nicht das konkrete 'Wie' deren Entstehung rekonstruieren, sondern müssen deren Existenz (Funktionalität, evolutionäre Stabilität, etc.) immer schon voraussetzen (so Ellrich/Funken 1998). Ähnlich argumentiert auch Werner (1996), der aus der Perspektive des reflexiven Paradigmas darauf insistiert, daß sich die Vorstellung einer Emergenz von (gewünschter) Koordination, Kooperation und hier vor allem von kollektiven Problemlösungen innerhalb von Agentensystemen ohne eine Vermittlung von kognitiven Komponenten als grundsätzlich fragwürdig darstellt: "Somehow the solution will emerge magically out of their interactions." (ebd., S. 26).

Im Anschluß an die Debatte um die Vor- und Nachteile von reflektiven und reaktiven Agentenarchitekturen werden auch sogenannte 'hybride Agentenarchitekturen' eingeführt (vgl. Müller, J.P. 1996), die Vorteile der beiden Ansätze in Form einer aus reflexiven und reaktiven Komponenten aufgebauten Schichtenarchitektur integrieren. Während die reaktive Schicht die rezeptorische Wahrnehmung und effektive Aktion des Systems vollzieht und dabei schnelle bzw. flexible Bearbeitung von eher einfach strukturierten Rohinformationen garantiert, ist die reflexive Schicht im Zuge ihrer kognitiven und schlußfolgernden Kompetenzen für die langfristige und komplexe Planung sowie Entscheidungsfindung des Gesamtsystems zuständig. Kiss (1992, zitiert nach Rammert 1998, S.108) schlägt im Rahmen hybrider Agentenarchitekturen eine dreischichtige Agentenarchitektur vor, die entsprechend den jeweiligen situativen Anforderungen entweder 'fully deliberative actions', 'complex skilled routines' oder 'simplex reflex actions' ausführt. Ein solches Agentenmodell erweist sich nach Rammert (ebd.) als anschlußfähig zu differenzierteren soziologischen Handlungsmodellen wie vor allem dem von Anthony Giddens im Rahmen seiner Theorie der Strukturierung konzipierten

Handlungsmodell, das zwischen unbewußt vermittelten Handlungsformen, einem praktisch-routinisierten Handeln und einem diskursivem Handeln sozialer Akteure unterscheidet. (vgl. unten 3.5.3.3.)

3.5. Soziologische Fundierungen der MAS

3.5.1. Soziale Agenten

Das Paradigma reflexiver Agenten fokussiert - auch wenn es sich problemlos anschlußfähig an spieltheoretische Ordnungsvorstellungen der soziologischen 'Rational choice theory' darstellt - individualistisch auf omnipotente Einzelkomponenten, während innerhalb des reaktiven Paradigmas biologisch inspirierte und soziologisch meist nicht weiter ausgeführte Vorstellungen systemischer (funktionaler) Selbstorganisation den Modellen von Koordination und Ordnung zugrundeliegen. Ein explizit und anspruchsvoll soziologisches Verständnis des wechselseitigen Charakters sozialer Beziehungen im allgemeinen und von Koordination bzw. Ordnung im besonderen findet sich demgegenüber bei der Konstruktion von 'sozialen' und/oder 'interaktiven' Agenten. So grenzen Moulin/Chaib-Draa (1996) in einem Einleitungsaufsatz zur VKI-und MAS-Forschungslandschaft soziale Agenten von reflexiven wie auch von reaktiven Agenten ab. Hier werden die den reflexiven Agenten zugeschriebenen kognitiven - hier: intentionalen - Kompetenzen als Voraussetzung der Entwicklung sozialer Modelle bzw. Fremdbilder als dem zentralen Charakteristikum sozialer Agenten verstanden. "A reactive agent reacts to changes in its environment or to messages from other agents. (...) An intentional agent is able to reason on its intentions and beliefs, to create plans of actions, and to execute those plans (...) In addition to intentional agent capabilities, a social agent posseses explicit models of other agents." (ebd., S. 8/9) Nach Moulin/Chaib-Draa richtet sich das Verhalten der Einzelagenten im Sinne des von George Herbert Mead entwickelten Modells eines 'taking the role of another' fortlaufend an dem Verhalten der anderen Akteure aus, so daß Koordination bzw. Kooperation auf der Hervorbringung von kognitiven und normativen Fremdbildern (und nach Mead auch: von entsprechenden Selbstbildern) sowie auf deren wechselseitigen Stabilisierung im Sinne einer (mehr oder weniger konsensuellen) Perspektivenverschränkung der Agentenorientierungen beruht.

In der soziologischen Theorielandschaft finden sich eine Vielzahl von Konzepten einer wechselseitigen Entstehung und längerfristigen bzw. globalen Etablierung von Fremd- und Selbstbildern, die sich explizit oder implizit im Rahmen der Modellbildungen der VKI bzw. MAS wiederfinden lassen. In der Tradition der (eher) makrosoziologisch ausgerichteten Ansätze des Strukturfunktionalismus und der Systemtheorie werden Fremd- und Selbstbilder als mehr oder weniger stabile und 'immer schon' vorhandene kognitive und normative Verhaltenserwartungen z.B. im Sinne der Luhmanschen 'Erwartungserwartungen' konzipiert. Hier sind es dem Handeln und Interagieren vorgängige Interaktions- bzw. Kommunikationsmedien wie Geld, Macht, Einfluß und Wertbindung (bei Parsons) - darüber hinaus auch Wahrheit, Liebe, etc. (bei Luhmann) -, die die wechselseitigen Erwartungen der Akteure strukturieren und somit deren Koordination sicherstellen (vgl. Schulz-Schaeffer/Malsch 1998, S. 246ff.). Die Ordnungsvorstellungen der 'Top-Down'-Modellbildungen der VKI (vgl. 3.1.) lassen sich aus dieser soziologischen Theorieperspektive problemlos rekonstruieren. Auf hierarchischen Strukturen beruhende

Modelle - z.B. die Unterscheidung zwischen ‘master agents’ und ‘slave agents’ – rekurrieren auf Macht und Kontraknetzsysteme auf ökonomische Ordnungsprinzipien, d.h. auf Geld als sozialen Koordinationsmechanismus (ebd., S. 246ff.). MAS-Ansätze, die auf Modellierungen von ‘sozialen Gesetzen’ (vgl. Shoham 1997), von sozialen Verpflichtungen in Form unterschiedlicher ‘commitments’ (Jennings 1996) oder von ‘joint commitments and mutual beliefs’ (Cohen/ Levesque 1990) abzielen, können entsprechend den sozialen Koordinationsmedien Einfluß und/oder Wertbindung zugeordnet werden. Auch die Übernahme eher kleinerformatiger organisationssoziologischer Modelle (vgl. Fox 1988, Kirn 1996) fügt sich in dieses Leitkonzept ein, insofern auch hier die entscheidende Rolle übergreifender, das jeweilige Handeln, Interagieren und Entscheiden strukturierender Vorgaben bei der sozialen Koordination im allgemeinen und der kollektiven Aufgabenbearbeitung im besonderen in den Vordergrund gerückt wird.

Allerdings erweisen sich solche ‘makrotheoretisch’ konzpierten Sozialmodelle der MAS als begrenzt, insofern sie Einzelagenten als Träger bzw. Vehikel vorgängiger Strukturen konzipieren und somit das innerhalb des Agentenparadigmas angestrebte Ziel der Konstruktion autonomer, lernfähiger und (im Sinne von Wooldridge/Jennings) zu einem initiativen bzw. aktiv zielorientierten Verhalten fähiger Agenten verfehlen (vgl. 3.2.2.). Wie auch dem Konzept übergreifender Interaktionsmedien (insbesondere in der Parsonianischen Version) in der soziologischen Theoriediskussion aus der Perspektive mikrosoziologischer Ansätze einer interpretativen (symbolisch interaktionistischen, ethnomethodologischen, etc.) Soziologie vorgeworfen wird, ignorieren diese Modelle die Eigenständigkeit und Gestaltungsmächtigkeit kontextuell situierter Akteure bei der Hervorbringung von Sozialität und reichen entsprechend nach Florian (1998) nicht aus, "um eine angemessene Vorstellung des situationsspezifischen Zusammenspiels zwischen der individuellen und der kollektiven Ebene bei unterschiedlichen Formen gemeinsamer Aktivitäten zu erlangen." (ebd. S. 319, vgl. Schulz-Schaeffer/Malsch 1998, S. 246ff.). Dem Anspruch der Modellierung sozialen Akteuren vergleichbarer automer Agenten werden demgegenüber wohl eher solche Ansätze gerecht, die ausgehend von der Vorstellung von "agents situated in a social world" (Conte/Castelfranchi 1996, S. 540) sich sowohl von einem quasi-individualistischen Fokus auf die Gestaltung der Einzelagenten als auch von einer ‘übersozialisierten’ Perspektive auf übergreifende Organisations- bzw. Ordnungmuster abgrenzen und auf die kontextuell sowie individuellhandlungsförmig konstituierten Mechanismen der (wechselseitigen) Produktion und Reproduktion von sozialen Koordinations- bzw. Ordnungsstrukturen abstellen.

In diesem Zusammenhang greifen die Vertreter der MAS vorwiegend auf - aufgrund ihrer problemlosen Formalisierbarkeit naheliegende - Ansätze der soziologischen Spieltheorie (‘Rational-Choice-Theory’) zurück, die spezifische Verfahren sozialer Strategiebildung (‘tit for tat’, etc.) als Grundlage sozialer Kooperation analysieren (Rosenschein/Genesereth 1988.). Allerdings bleiben spieltheoretische MAS-Ansätze eng einem dem reflexiven Agentenparadigma zugrundeliegenden Modell von Agenten als Trägern stabiler Ziele und allumfassender Wissenskompetenzen verhaftet. So laufen auch sie - über das auch hier relevante Problem sehr hoher Rechenkapazitäten hinaus – Gefahr, von einem zentral berechneten ‘one best way’ abweichende Verlaufsformen von Agentensystemen im allgemeinen und Möglichkeiten einer In-Interactu-Hervorbringung von neuen Verhaltensweisen (‘Lernen’) im besonderen auszuschließen. An dieser Stelle

werden dann zwei traditionell innerhalb der Soziologie gegenüber spieltheoretischen Sozialvorstellungen vorgebrachte Argumente relevant. Erstens betont der Symbolische Interaktionismus, daß Merkmale sozialer Agenten weniger Voraussetzungen als vielmehr Folgen sozialer Interaktionen darstellen und folglich die Hervorbringung von sozialer Kooperation weniger auf relativ starren Strategiebildungen, sondern vielmehr auf kontingenten - potentiell neue Verhaltensformen bzw. Problemlösungen generierenden - Aushandlungsleistungen der Teilnehmer beruht (vgl. Schulz-Schaeffer 1998, S. 135ff.). Auf der anderen Seite rücken innerhalb der soziologischen Mikro-Makro-Debatte sowohl Relativierungen einer vollständigen Autonomie bzw. Freiheit der die Sozialität produzierenden Akteure als auch Reflexionen der Formen einer Etablierung bzw. Stabilisierung globaler Ordnungs-zusammenhänge vor dem Hintergrund situational beschränkter Handlungsepisoden in den Blickpunkt (vgl. Rammert 1998, S.110ff.). Wie solche soziologischen Vorstellungen im Zuge des expliziten Anspruches einer sich von spieltheoretischen Konzepten abgrenzenden soziologischen 'Fundierung' der MAS aufgegriffen werden und welche Probleme dabei zu Tage treten, soll im folgenden erstens am Beispiel des auf Grundannahmen des Symbolischen Interaktionismus zurückgreifenden Konzeptes von Les Gasser und zweitens anhand des auf Überlegungen der Strukturierungstheorie von Giddens rekurrierenden Ansatzes von Rosario Conte und Christiano Castelfranchi erläutert werden.

3.5.2. MAS und symbolischer Interaktionismus: Les Gasser

Les Gasser (1991, 1992, Gasser u.a. 1987, vgl. zusammenfassend Malsch 1998a) bezieht sich bei dem Versuch einer soziologischen Fundierung der Gestaltung von Multiagentensystemen auf Grundannahmen des Symbolischen Interaktionismus von Mead. Hierbei geht er davon aus, daß weder eine vorgängige bzw. 'essentialistische' Zuschreibung von Akteurseigenschaften noch die Modellierung übergreifender Ordnungsstrukturen, sondern vielmehr sich interaktiv konstituierende und dynamisch reproduzierende soziale Beziehungen den Ausgangspunkt gleichermaßen von soziologischer Theoriebildung wie auch von der Modellierung von Multiagentensystemen darstellen. "We shall examine the possibilities for building systems as structures of interaction, from which useful entities we can call 'agents' emerge." (Gasser 1992, S. 201) Eigenschaften und Kompetenzen der Agenten sind nicht Voraussetzung, sondern vielmehr Resultat von Interaktionen, wobei im Zuge der Vorstellung einer interaktiv konstituierten 'Emergenz von Sozialität' betont wird, daß "the whole (society) is prior to the part (the individual), (...) and the part is explained in terms of the whole."(Gasser 1991, zitiert bei Conte/Castelfranchi 1996, S. 540). Entsprechend stehen nicht problemlösende Einzelagenten, sondern vielmehr Agentenkollektive im Blickpunkt, die 'in interactu' jeweilige Koordinations- bzw. Kooperationsordnungen aushandeln und im Systemverlauf (potentiell) neue Problemlösungsmuster hervorbringen. Zwar ähnelt Gassers interaktionistisches bzw. 'emergentistisches' Sozialkonzept (Malsch 1998, S. 276) den sozialtheoretischen Vorstellungen von Maes (vgl. 3.4.2.), setzt aber nicht an den Wahrnehmungs- und Verhaltensmerkmalen reaktiver bzw. subkognitiver Agenten, sondern vielmehr - entsprechend dem Verständnis sozialer Agenten bei Moulin/Chaib-Draa - an den Sozialvorstellungen bzw. -modellen der interagierenden Akteure an. Entsprechend

beschreibt Gasser soziale Interaktion bzw. Kooperation im Sinne der Meadschen Vorstellung von wechselseitigen Perspektivenverschränkungen als das Ergebnis eines Abgleiches von aus der Beobachtung der Anderen resultierenden Fremd- und entsprechenden Selbstbildern der Akteure. Seine Modellierung von MAS geht entsprechend von einem kontinuierlichen kommunikativen Austausch von spezifischen, jeweilige Orientierungen bzw. 'Identitäten' der Agenten darstellenden Selbstbeschreibungen aus, die sich im Anschluß an spezifische Verhandlungs- und Abstimmungsleistungen der Akteure sukzessive zu weiterreichenden Zusammenhängen bzw. Netzwerken von sozialen Abhängigkeiten und Verpflichtungen verdichten (Gasser 1991, S.ff.).

Bei der Erklärung der globalen Ausbreitung und Stabilisierung von zuallerst zwischen lokal und situational situierten Anwesenden konstituierten Abhängigkeiten und Verpflichtungen greift Gasser im Anschluß an Latour (1987) auf Konzepte aus dem Bereich der 'konstruktivistischen Wissenschaftssoziologie' zurück (ebd., S.126ff.). Jene erläutern den Aufbau und die Reproduktion makrosozialer Abhängigkeits- bzw. Verpflichtungsnetzwerke anhand von objekthaft-technisch vermittelten Kommunikations- bzw. Ordnungsformen, deren Funktionen am Beispiel von Laboraufzeichnungen (Latour 1987), wissenschaftlichen Publikationen (vgl. Knorr-Cetina 1981, 1990) sowie einer Vielzahl anderer 'boundary objects' (Landkarten, Gebäudeformen, etc., vgl. Star 1989) analysiert werden. Solche unterschiedlichen, von Knorr-Cetina (1981, S. 25ff. und 1990, S. 10ff.) als 'macropresentations' reflektierte Ordnungsformate garantieren die Vernetzung zwischen einzelnen räumlich und zeitlich begrenzten Mikroepisoden, insofern sie in einem spezifischen Interaktionskontext konstruiert, in einen anderen Interaktionskontext 'verschoben' und dort von den Teilnehmern neu angeeignet werden. Diese ‚macrorepresentations' können räumliche und zeitliche Distanzen wie auch potentielle soziokulturelle Barrieren zwischen divergierenden 'social worlds' (Star 1989) überwinden, weil sie ein von seinen jeweiligen Entstehungs- und Aneignungskontexten und deren je spezifischen sozialen Interessen und soziokognitiven Orientierungen abstrahierendes formales Wissen verkörpern, das in Form von mobilen, objekthaft bzw. technisch konstituierten Wissensbehältern zwischen jenen Kontexten transportiert werden kann. Entscheidend für die Analyse von diesen Repräsentationen ist, daß ihnen einerseits meist nicht nur ein abstrahierender, sondern auch 'verkennender' oder gar strategisch inszenierter Charakter zukommt, der andererseits gerade kein soziokognitives Mißverstehen oder soziale Disaggregation nach sich zieht. Nach Knorr-Cetina werden "Ordnungsformate und andere Repräsentationsformen in vielen aufgreifenden Situationen nicht 'at face' value genommen" (1990, S.16), sondern vielmehr von den sozialen Akteuren 'durchschaut' und 'unterlaufen', wobei diese Akteurskompetenzen entweder im Sinne einer "praxeological perspective of culture" (Knorr-Cetina 1996, S. 310) vorrangig auf praktische Fähigkeiten (vgl. Hirschauer 1994) oder vorrangig auf kognitive Fähigkeiten im Sinne expliziter und impliziter Wissensformen (vgl. Colllins 1994) zurückgeführt werden. Die soziologische Pointe des Konzeptes von 'macrorepresentations' ist folglich eine - von den sozialen Akteuren allerdings aufgrund ihrer interpretativen Flexibilität antizipierte sowie erfolgreich ‚bearbeitete' - Differenz bzw. 'Spannung' zwischen kontextfreien und technisch konstituierten Wissensbausteinen auf der einen sowie kontextuell und sinnhaften aktualisierten Wissensformen auf der anderen Seite.

Gassers Konzeption einer symbolisch interaktionistischen Fundierung der Modellierung von Multiagentensystemen liegen m.E. zwei für die soziologische Bewertung von Multiagentensystemen instruktive Schwierigkeiten hinsichtlich seines Verständnisses von sozialer Emergenz und seiner Analyse globaler Ordnungszusammenhänge zugrunde (vgl. Malsch 1998a). Erstens stellt sich sein 'emergentistisches', ausschließlich auf die interaktive Konstruktion von Agentenmerkmalen abstellendes Konzept sowohl aus soziologischer Perspektive als einseitig wie auch aus der Perspektive der Informatik als nicht direkt umsetzbar dar. So läßt sich das von Gasser postulierte strikte Verständnis einer 'in situ' Konstruktion und Aushandlung von Sozialität insbesondere im Anschluß an die sozialpsychologischen Annahmen von Mead nicht durchhalten, denen zufolge Interaktionsprozesse nur vor dem Hintergrund vorgängiger bzw. sozialisatorisch erworbener Identitätsmuster verstanden werden können. Wie Gasser bei seiner Analyse der Konstitution wechselseitiger Selbstbeschreibungen - die die Fähigkeiten der Akteure "to represent and reason about the knowledge, actions, and plans of other agents" (Gasser 1991, S. 109, zitiert nach Malsch 1998) voraussetzen - impliziert anerkennt, müssen spezifische Komponenten sozialer Identitäten - z.B. im Sinne der von Anthony Giddens in seiner soziologischen Strukturierungstheorie betonten 'Dualität von Struktur' (vgl. 3.5.3.3.) – gleichermaßen als Voraussetzung und Folge dynamisch konzipierter Interaktionsprozesse verstanden werden. Entsprechend werden auf der einen Seite gegenüber Gassers emergentinistischem Ansatz 'individualistische' Einwände, die auf den Eigenbetrag sozialer Akteure bei dem Vollzug sozialer Interaktionen abstellen, erhoben (so z.B. Conte/ Castelfranchi 1996, vgl. 3.5.3.). Auf der anderen Seite verstrickt sich Gassers Anatz bei der konkreten Implementation von Agentensystemen in Widersprüche (vgl. Malsch ebd., S.275 ff.), insofern die durch die Meadsche Sozialpsychologie nahegelegte Vorstellung einer sukzessiven Sozialisation der Agenten in einem informatorischen Zusammenhang wenig Sinn macht. Vielmehr müssen Agenten 'ex nihilo', d.h. vor dem ersten Programmlauf des Agentensystems mit spezifischen Eigenschaften und Kompetenzen ausgestattet werden, um wechselseitige Kommunikationen und Aushandlungen überhaupt erst in Gang bringen zu können. "Im Unterschied zu Mead, der die 'Selbstbeschreibung' eines menschlichen Individuums zeitgleich mit der Rollenübernahme im frühkindlichen Sozialisationsprozeß entstehen läßt, bleibt Gasser (...) gar nichts anderes übrig, als seine Agenten, schon bevor sie in den Interaktionsprozeß eintreten, mit einer kooperationsbereiten Identität auszustatten." (ebd., S. 276) Diese Notwendigkeit der Modellierung vorgängiger Agenteneigenschaften liegt auch dem von Gasser mitentwickelten, eine Jagd simulierenden Agentensystem 'MACE' (Gasser u.a. 1987) zugrunde. Dieses stattet im Sinne der Spieltheorie einzelne Agenten bzw. 'Jäger' bei der (vorgängigen) Programmierung mit eindeutigen Zielvorgaben und Verhaltensstrategien aus, woraufhin dann im Systemverlauf nichtvorhersehbare spezifische Agentenkonstellationen und potentiell neue kollektive Problemlösungen emergieren sollen. Ein solches Agentensystem erreicht nicht das anspruchsvolle Ziel einer Umsetzung bzw. Simulation der Theorie des Symbolischen Interaktionismus, sondern stellt offensichtlich vielmehr eine implementationstechnisch notwendige Kombination von Grundannahmen der Spieltheorie und des Symbolischen Interaktionismus dar.[10]

10. Eine solche Kombination von Annahmen der Spieltheorie und des Symbolischen Interaktionismus ist nach Nwana u.a. (1997, S. 49ff.) als 'spieltheoretischer Verhandlungsansatz ' ('game theory-based negotiation')

Allerdings erweist sich nicht nur Gassers Interaktionskonzept, sondern auch seine Vorstellung von ‘macrorepresentations’ als dem konzeptuellen Garant einer ‘action at distance’ bei seiner Übertragung auf Multiagentensysteme als problematisch. Zwar kann dieses Konzept innerhalb der jeweiligen technischen Systeme in Form von die Informationsübertragung gewährleistenden ‘envelopes’ (Hewitt) oder ‘Grenzobjekten’ (Star) als Kommunikationsmittel bzw. -behälter umgesetzt werden. Allerdings geht bei einer solchen Übersetzung die soziologische Pointe des Konzeptes der ‘macrorepresentations’ in Form der - in einer Vielzahl anderer Kommunikationstheorien in begrifflichen Unterscheidungen von Daten, Nachrichten bzw. Informationen auf der einen und Wissen auf der anderen Seite betonten - Differenz zwischen kontextfreien, technisch vermittelten ’Wissensbausteinen’ und kontextuellen, soziokognitiv bzw. sinnhaft konstituierten Wissensformen verloren. "Die metaphysischen Mucken der Multiagentensysteme liegen aus soziologischer Sicht dann aber darin, daß künstliche Agenten gar nicht feststellen können, ob sie es mit Technik oder Sozialität zu tun haben."(Malsch ebd., S. 287) Aber nicht nur die Unterscheidung zwischen ‘technisch’ und ‘sozial’, sondern auch andere sozialtheoretisch relevante begriffliche Differenzen wie beispielsweise zwischen ‘gewohnheitsmäßigen Konventionen’ und ‘heiligen Normen’ oder auch zwischen kognitiven und normativen Regelzusammenhängen können aus der Perspektive von technischen Agenten (-systemen) - die entsprechend der ‘Hollow-Shell’-Kritik der KI grundsätzlich nicht zur interpretativen Deutung von soziokognitiven Sinnzusammenhängen befähigt sind (vgl. 2.2.2.) - nicht rekonstruiert werden. Demzufolge können sozial relevante und sozialtheoretisch gehaltvolle Unterscheidungen nicht nur - wie Schulz-Schaeffer/Malsch (1998, S. 251) schreiben - "nicht ohne Komplexitätsverlust", sondern m.E. darüberhinaus auch ‘nicht ohne Bedeutungsverlust’ auf technische Systeme im allgemeinen und informatorische Agentensysteme im besonderen übertragen werden.

3.5.3. MAS und Strukturierungstheorie

3.5.3.1. Das Konzept von Conte/Castelfranchi

Conte/Castelfranchi (Conte/Castelfranchi 1995a,b 1996, Castelfranchi 1990, 1995, vgl. Florian 1998, S. 315ff.) beanspruchen eine soziologische Fundierung der MAS, die im Anschluß an ‘Lösungsversuche’ der soziologischen Mikro-Makro-Debatte ein angemessenes Verständnis der Wechselwirkungen zwischen von Individuen situativ konstituierten Sozialepisoden und weit in Raum und Zeit ‘ausgedehnten’ sozialen Strukturzusammenhängen liefern soll. Ausgangspunkt ihrer Überlegungen ist eine Kritik der in der VKI und MAS dominierenden Konzepte, denen sie erstens eine explizite oder implizite ‘benevolent assumption’ und folglich eine einseitige Ausrichtung an der Kooperationsorientierung sozialer Agenten sowie die Vernachlässigung von Macht,

---

in der MAS weit verbreitet (vgl. Corry et.al , Müller H.J. 1996). Diese Modelle setzen an der vorgängigen Modellierung und Programmierung von relativ strikten Agentenmerkmalen in Form von spezifischen Nutzenorientierungen und rational kalkülierbaren Verhaltensstrategien der Akteure an, woraufhin jene als Verhandlungpartner - oft ähnlich wie in den Kontraktnetzsystemen durch spezifische ‘manager agents’ als Vermittler bzw. ‘Mediatoren’ organisiert - die jeweilige Aufgabenverteilung bei der kollektiven Problembearbeitung aushandeln.

Herrschaft und/oder Konflikt als konstitutiven Bestandteilen sozialer Beziehungen nachweisen. "The ‘overcooperative’ paradigm is so powerful that conflict is admitted only as a reconcilable and socially useful phenomenon." (1996, S. 530, vgl. Castelfranchi 1990) Zweitens kritisieren sie die Vorstellung hyperkognitiver Agenten, die im Rahmen des reflexiven Agentenparadigma sowie von spieltheoretischen Ansätzen der MAS von einem vollständigen Überblick der Agenten über die jeweiligen Handlungssituationen ausgeht und dementsprechend ausschließlich an ihrer Fähigkeit zur strategischen Zielverfolgung - oder im Sinne spieltheoretischer Verhandlungsansätze (vgl. oben): an ihren Fähigkeiten zu einer strategisch orientierten Aushandlung - ihrer jeweiligen Ziele ansetzt. "Agents are by default omniscient and strategic and likely to take what in a multiagent context is seen as the most rational course of action, namely negotiation and cooperation. More in general, agents are modelled as having in their minds the representation of social links." (ebd.) Demgegenüber betonen Conte/Castelfranchi die Bedeutung vorgängiger und kontextübergreifender globaler Struktur- bzw. Ordnungsmuster z.B. in Form von "preexisting norms, habits and procedures" (ebd.), die sich der kognitiven Wahrnehmung bzw. Reflexion der Agenten entziehen und dementsprechend deren Autonomie bzw. ‘Freiheit’ einschränken. Hierbei zielen Conte/Castelfranchi (Conte/Castelfranch 1995a, S. 1ff., ähnlich auch Gilbert 1995) aber nicht auf ein strukturtheoretisch-deterministisches Ordnungsverständnis im Sinne der ‘Top-Down’ - Modellbildungen der VKI, sondern vielmehr unter Berufung auf die Theorie der Strukturierung von Anthony Giddens (vgl. Giddens 1977, 1984, 1992) auf einen zwischen strukturtheoretischen und handlungstheoretischen bzw. interaktionistischen Sozialvorstellungen vermittelnden Ansatz (vgl. Rammert 1998, S. 108ff.). Entsprechend dem Giddensschen Grundgedanken einer ‘Dualität von Struktur’, demgemäß "gesellschaftliche Strukturen sowohl durch das menschliche Handeln konstituiert, als auch zur gleichen Zeit das Medium dieser Konstitution sind" (Giddens 1992, S.148), sollen soziale Ordnungszusammenhänge gleichermaßen als partiell einschränkende Voraussetzungen als auch als - partiell nicht reflektierte bzw. antizipierte - Folgen sozialen Handelns und Interagierens verstanden werden.

Im Anschluß an diese sozialtheoretischen Grundüberlegungen distanzieren sich Conte/Castelfranchi in zweierlei Hinsicht von Gassers interaktionistischem MAS-Konzept. Auf der einen Seite darf ihres Erachtens im Zuge eines ‚emergentinistischen’ Interaktionsverständnisses nicht das gesellschaftliche ‘Ganze’ gegen seine individuellen ‘Einzelteile’ ausgespielt und somit der Eigenbeitrag deren invididuellen Orientierungen ignoriert werden. "Even to account for collective minds and activities individual motivations ought to be accounted for, since they have an impact on the nature and quality of groups and interaction." (ebd.) Auf der anderen Seite soll im Gegensatz zu Gasser der quasi-objektive Charakter von "uncommited structures" betont (ebd.), d.h. nicht nur im Sinne des Macrorepresentations-Konzepts der Koordination ermöglichende, sondern auch der einschränkende bzw. ‘hinter dem Rücken’ der Akteure wirkende Charakter von globalen Strukturzusammenhängen hervorgehoben werden. "We share Gasser’s view that most social relations preexist to interactions and commitments among the individuals, but social relations and organizations are not held or created by commitments (mutual, social) of the individuals. (...) Autonomous agents in a multiagent world (...) find themselves as ‘socially situated agents’. They find themselves in a network of relations (interference,

dependence, concurrence, power, etc.) that are independent of their awareness and choices." (ebd.)[11]

Hinsichtlich eines Modells, das die Verschränkung zwischen individuellen Orientierungen und globalen, quasi-objektiven Strukturzusammenhängen adäquat beschreibt, haben Conte/Castelfranchi mehrere Anläufe unternommen. So unterscheiden sie (1992, S. 80ff.) zwischen drei Formen bzw. Stufen sozialer Ordnungszusammenhänge in Form von intentionalen, extern gesteuerten und funktionalen Kooperationen.[12] Im Falle der intentionalen Kooperationen ist die Gesamtheit der sozialen Beziehung den beteiligten Akteuren kognitiv präsent, d.h. sie können ihre wechselseitigen Abhängigkeiten sowie ihre jeweiligen, für das Gelingen der Interaktionen im Sinne einer gemeinsamem Zielerreichung notwendigen Verhaltensweisen antizipieren und reflektieren. Im Falle einer sogenannten 'Out-Designed Cooperation' ist demgegenüber nicht mehr das gesamte Handlungswissen den beteiligten Akteuren kognitiv gegenwärtig, vielmehr sind Teile dessen an eine 'externe Struktur' ausgelagert, d.h. an Ordnungsinstanzen wie z.B. einen Mediator oder eine Leitperson ('master'- oder 'manager'-agents) delegiert. Entscheidend für Castelfranchi/Contes' Analyse ist allerdings die dritte Form einer funktionalen Kooperation, die ihre Vorstellung einer von den Akteuren selbst nicht mehr 'durchschauten' Reproduktion von Sozialität im Sinne einer "objective relation of mutual dependence" (ebd., S. 86) erläutert. Sowohl von den Absichten der Akteure als auch von Planungsinstanzen unabhängige Handlungseffekte erweisen sich - so rekurrieren auch Conte/Castelfranchi auf die Vorstellung einer "emergenten Funktionalität" (ebd., S.80) von Sozialität (vgl. 3.4.2.) - als funktional hinsichtlich der Stabilisierung bzw. der (funktional verstandenen) Bestandserhaltung jeweiliger gesellschaftlicher Gesamtzusammenhänge. Allerdings soll, so spezifizieren Conte/Castelfranchi (1996) die Zielrichtung ihres Ansatzes, die 'emergente Funktionalität' jeweiliger Handungseffekte nicht im Sinne des reaktiven Paradigmas als Folge der Orientierungen und Aktionen subkognitiver Agenten, sondern vielmehr als "emergent functionalities of actions intented and planned by cognitive agents" (1996, S.531) verstanden werden.

Die Schwäche dieses Modells einer emergenten Funktionalität ist, daß es - trotz seiner konflikttheoretischen Programmatik - im Zuge eines bestandsorientierten Funktionsbegriffs ausschließlich globale bzw. makrosoziale Funktionalitäten, aber nicht entsprechende Dysfunktionalitäten in den Blickpunkt bekommt (vgl. Ellrich/Funken, 1998, S. 382).[13]

---

11. Zur Gegenüberstellung der Konzepte von Gasser und Conte/Castelfranchi, die diese Kritik an interaktionistischen Ansätzen hervohebt, vgl. Florian (1998, S 309ff.) Eine andere Auffassung vertritt hier Strübing (1998), demgemäß der Symbolische Interaktionismus und insbesondere Gassers Ansatz durchaus die Vorgängigkeit und Bedingtheit sozialer Akteure reflektiert, insofern er sie als Träger übergreifender 'social worlds' konzipiert."Gasser et.al beziehen sich explizit auf den 'negotiated oder approach" und die Theorie sozialer Welten des neueren SI (Symbolischer Interaktionismus, K.S) und damit auf eine sozial konstituierte Handlungskompetenz von Akteuren. Hewitt und andere VKI-Forscher ziehe lediglich Konsequenzen aus der Einsicht, daß soziale Strukturen a) nur interaktiv erfahrbar b) auch nur in sozialem Handeln (und in dessen sukzessiver Aggregation) änderbar sind."(ebd., S. 69)
12. Darüber hinaus identifizieren Conte/Castelfranchi (ebd. S. 82) einen vierten Typ in Form einer 'accidential cooperation', der aber - wie sie anmerken - weniger einen originären Typ sozialer Kooperation, sondern vielmehr einen 'evolutionary forerunner' sozialer Beziehungen darstellt.
13. Vgl. auch Ellrich/Funken (1998), S. 380. Der bestandsorientiertem Funktionsbegriff wird in der Soziologie aus unterschiedlichsten Theorieperspektiven in Frage gestellt. So zum Beispiel aus der

Darüber hinaus stellt sich ein Stufenmodell als nicht geeignet dar, um die konkreten Formen und Prozesse des Übergangs von intentionalen zu nichtintentionalen bzw. funktionalen Kooperationen rekonstruieren zu können. Einen weiteren Versuch, das 'Wie' des Übergangs von intentional und nichtintentional vermittelten Sozialbeziehungen, d.h. auch der 'Verschränkung' von individuellen Kognitionen und übergreifenden Kollektivitäten bzw. Ordnungsformen zu bestimmen, unternimmt Castelfranchi (1995) im Zuge der Kategorisierung unterschiedlicher Formen sozialer Verpflichtungen. So unterscheidet er zwischen einem sogenannten 'I-Commitment', das die internen kognitiven Einstellungen eines individuellen Akteurs zu seinen Handlungsoptionen und -vollzügen bezeichnet, und einem 'C-Commitment', das die gemeinsamen Verpflichtungen innerhalb eines sozialen Kollektivs umfaßt. Demgegenüber stellen sogenannte 'S-Commitments' die Verschränkung zwischen den individuellen und kollektiven Verpflichtungen, d.h. zwischem den I-Commitments und den C-Commitments her. Diese drei Formen sozialer Verpflichtungen reichen allerdings nur für das Verständnis lokal und situational begrenzter Strukturkontexte, nicht aber für die Analyse raum-zeitlich 'ausgedehnter' Sozialbeziehungen wie z.B. die Strukturzusammenhänge 'großer' Organisationen aus, so daß Castelfranchi weitere normative Ordnungstypen - "generic commitment", "generic meta-commitment" und "organisational commitment" (1995a, S. 47) - einführt.

Entscheidend für Castelfranchis Analyse ist allerdings auch hier, daß sich die Gesamtheit gesellschaftlicher Globalzusammenhänge nicht ausschließlich vermittels von normativen Ordnungstypen, sondern im Sinne von "uncomitted structures" (1996, S. 340) auch über eine Vielzahl von nicht normativ verankerten Abhängigkeits- und Machtbeziehungen (politischen und ökonomischen Ressourcenverteilungen im Sinne von Giddens) konstituiert. In diesem Zusammenhang gelingt es m.E. dem von Castelfranchi entwickelten Modell sozialer Verplfichtungen wie auch ähnlichen Überlegungen bei Conte/Castelfranchi (1995a) nicht, eine eindeutige konzeptuelle Bestimmung des Verhältnisses von individuellen Agentenkognitionen zu jenen nicht normativ verankerten Makrostrukturen zu liefern. Vielmehr lassen sie - entsprechend der etwas widersprüchlichen Ankündigung des Ansatzes als einem "dialectic and co-evolutionary approach" (Castelfranchi/Conte 1996, S. 532, zitiert nach Florian, ebd., S. 318) - zwei unterschiedliche Interpretationen zu, die sich aus der Perspektive soziologischer Theoriebildung als konkurrierend bzw. einander ausschließend darstellen. Ein Koevolutionsmodell und ein Dialektikmodell des Verhältnisses von individuellen (Mikro-) Orientierungen und kollektiven Makrostrukturen stimmen m.E. mit den von Ellrich/Funken (1998, S. 358ff.) identifizierten konkurrierenden Perspektiven auf emergente Sozialphänomene in Form von entweder - im Sinne einer 'Emergenz von oben' - an spezifischen sozialen Selbstorganisationsprozessen ansetzenden "Rotationstheorien" oder - im Sinne einer 'Emergenz von unten' - an Sozialität konstituierenden individuellen und kollektiven Handlungseffekten ansetzenden "Transformationstheorien" überein. Insofern sich - soweit ich sehe - zu diesen beiden divergierenden theoretischen Konzepten bei Conte/Castelfranchi keine eindeutigen Hinweise finden lassen, sollen sie im folgenden exemplarisch anhand der von einer Koevolution von psychischen und sozialen Systemen

---

Perspektive der RC-Theorie von Elster (1987), aus der Perspektive der (funktional-strukturellen) Systemtheorie von Luhmann (1970) und aus der Perspektive der Strukturierungstheorie von Giddens (1976, ähnlich auch Joas 1992a).

ausgehenden Sozialtheorie von Niklas Luhmann und der von einem (quasi-) dialektischen Verhältnis von individuellen Handlungsformen und globalen Struktur- bzw. Systemzusammenhängen ausgehenden - und von Conte/Castelfranchi selbst als theoretischem Leitkonzept hervorgehobenen - Theorie der Strukturierung von Giddens eingeführt sowie hinsichtlich ihrer Relevanz für eine soziologische Fundierung der MAS miteinander verglichen werden.

3.5.3.2. Die Luhmannsche Systemtheorie

Die Luhmannsche Theorie sozialer Systeme (1984, 1990) versteht das Verhältnis individueller Akteursorientierungen und globaler Ordnungszusammenhänge als Koevolution von wechselseitig 'geschlossenen' und sich quasi-evolutionär reproduzierenden Systemen. Sowohl über Gedanken bzw. Bewußtsein konstituierte psychische Systeme als auch kommunikativ konstituierte soziale Systeme (Interaktions- ,Organisations- und gesellschaftliche Teilsysteme) werden als 'autopoietisch' beschrieben, insofern sich ihre kontinuierliche 'Selbstherstellung' am je eigenen Sinn- bzw. Ordnungsschemata orientiert und sich hierbei - im Sinne einer Rekursivität bzw. 'Selbstbezüglichkeit' der Systementwicklung - ausschließlich auf vorherige Systemereignisse bezieht. "Die Einheit eines Systems, eines einzelnen Gedankens oder einer einzelnen Kommunikation, kann immer nur im System unter rekursiver Vernetzung mit anderen Elementen desselben Systems erzeugt werden." (Luhmann, 1990, S. 30) Im Anschluß an diese Vorstellung einer selbstbezüglich-rekursiven 'Geschlossenheit' der Systeme wird deren Umweltkontakt, d.h. auch die Verschränkung von psychischen Bewußtseinssystemen und sozialen Kommunikationssystemen als strukturelle Kopplung aufgefaßt. Gemäß diesem Modell beschränkt sich der "systemische Außenkontakt auf eine nur für Beobachter sichtbare struktuelle Kopplung, die im System (und nur dort) Irritationen erzeugen kann, die sich an dessen Strukturen zeigen und zu Neuspezifikationen dieser Strukturen mit Mitteln der systemeigenen Operationen führen können" (ebd.,S. 530).

Luhmanns Autopoiesis-Konzept versteht individuelle Bewußtseinssysteme und kollektiv -soziale Kommunikationssysteme als zwar strukturell gekoppelte, aber im Sinne einer Koevolution ihre je eigene Selbstorganisation vollziehende Prozesse. Es geht entsprechend in der Tradition von Durkheim von der Vorstellung quasi-objektiver 'Sozialer Tatsachen' und in diesem Sinne von einer Emergenz von Sozialität aus, die sich 'immer schon' aktiven und kreativen Konstitutionsleistungen von - als Bewußtseins- und Handlungsträger in der Umwelt sozialer Systeme verorteten - Akteuren entzieht. Hieran anschließend stellt sich bei Luhmann die gesamtgesellschaftliche Entwicklung als ein Evolutonsprozeß dar, der in einem sukzessiven Wechselspiel der Hervorbringung und anschließenden 'Bearbeitung' von sozialer Komplexität kontinuierlich neue, 'angemessene' bzw. 'bessere' Mechanismen zur Wahrnehmung und Lösung von sozialen Problemen, d.h. kollektiven Lernprozessen hervorbringt (vgl. Ellrich/Funken, S. 386ff.). Hinschtlich der Modellierung von MAS legt die Vorstellung einer quasi-eigenständigen sozialen Evolution dann nahe, nicht nur die Autonomie der technischen Agenten gegenüber (quasi-) gesellschaftlichen Vorgaben, sondern gleichzeitig auch die "Autonomie der Agentengesellschaft (...) sowohl gegenüber ihren Agenten als auch gegenüber ihren menschlichen Konstrukteuren" zu betonen (Florian ebd., S. 316).

Zwar kommt Luhmanns Modell von weder durch steuernde Eingriffe noch durch Agentenorientierungen vermittelten evolutionär-selbstorganisatorischen Gesellschaftsprozessen einer Übertragung auf - von Programmierern oder Nutzern unabhängige Entwicklungen bzw. Prozesse vollziehende - Systeme der MAS entgegen, zieht aber auch Probleme nach sich (vgl. Ellrich/Funken 1998, S. 370ff.). Auf der einen Seite sperrt sich das Autopoiesis-Konzept im allgemeinen und die Vorstellung einer strukturellen Kopplung von psychischen und sozialen Systemen im besonderen gegen eine Formalisierung und Algorithmisierung bei der konkreten Gestaltung von Agentensystemen: "Die algorithmische Fassung einer Relation, die Unabhängigkeit und Vorausgesetztheit als Eigenschaften zweiter koevolutionierender Systeme erfaßt, transformiert zwangsläufig die präsumptive ‘strukturelle Kopplung’ zwischen zwei emergenten Bereichen in eine ‘operative Kopplung’." (ebd., S. 384) Auf der anderen Seite liegt nach Ellrich/Funken das Manko des Luhmannschen Ansatzes wie letztlich aller selbstorganisationstheoretischen bzw. ‚rotationstheoretischen’ Analysen von sozialer Emergenz „gerade dort, wo es selbst seine Stärke verortet: im Erfassen problemorientierter Innovations- und Irritationspotentiale." (ebd. S. 386) Insofern Luhmanns Vorstellung von Koevolution die wechselseitige ‘Autonomie’ von Individuum und Gesellschaft hervorhebt, grenzt sie sich zwar von strukturtheoretischen Annahmen hinsichtlich der Determiniertheit sozialen Agierens ab, bekommt aber nicht die in handlungstheoretischen Agency-Konzepten herausgearbeitete und auch in den obigen Agentendefinitionen hervorgehobene aktive und kreative Gestaltungsmächtigkeit sozialer Akteure in den Blick. Darüber hinaus kann Luhmanns Entwicklungsmodell im Zuge der zirkulären Vorstellung einer ‘evolutionär gleichursprünglichen’ Hervorbringung von sozialen Problemen und deren Problemlösungen zwar ‘post hoc’ die Selektion und Stabilisierung spezifischer Problemlösungs- und Lernmuster konstantieren, nicht aber - im Sinne einer ‘Assymetrisierung’ von Problemen und Problemlösungen - das konkrete ‘Wie’ der Hervorbringung von Problemlösungen rekonstruieren. Insbesondere stellt es über eine evolutionär verstandene systemische Anschlußfähigkeit keine Kriterien für die Bewertung jeweiliger sozialer Innovationen bzw. sozialen Lernens bereit. Soziologische Vorstellungen einer Selbstorganisation und Evolution sozialer Ordnungszusammenhänge vernachlässigen - so reformulieren auch Ellrich/Funken gegenüber Luhmann traditionelle handlungs- bzw. akteurstheoretische Einsichten - die konkreten Formen und Prozesse einer aktiven handlungsförmigen Hervorbringung von Sozialität im allgemeinen sowie von in spezifischen Agentenkonstellationen (aktiv und kreativ) erzeugten Problemlösungs- bzw. Lernleistungen im besonderen.

3.5.3.3. Die Giddenssche Theorie der Strukturierung

Die Giddenssche Theorie der Strukturierung (1977, 1984, 1992, 1995, vgl. zusammenfassend Joas 1992a,b sowie Cohen 1989) geht im Unterschied zu Luhmann im Sinne einer ‘Dualität von Struktur’ von einem ‘dialektischem’ Charakter des Verhältnisses von gesamtgesellschaftlichen Struktur- und/oder Systemzusammenhängen und den Handlungsorientierungen und -weisen sozialer (indvidueller) Akteure aus. So bestreitet sie als eine "konstitutionstheoretische" (Joas 1992a, S.336) oder auch "tranformationstheoretische" (Ellrich/Funken ebd.) Alternative zu Systemtheorien die

Annahme jeglicher selbstorganisatorischer Eigendynamiken und Stabilisierungprozesse und will demgegenüber Akteure als die "einzigen treibenden Kräfte in menschlichen Sozialbeziehungen" (Giddens 1992, S. 235) verstehen. Zwar konstantiert auch die Strukturierungstheorie makrosoziale, weit in Raum und Zeit ausgreifende systemische Sozialbeziehungen, die sich im Sinne einer "Entbettung" (1995, S. 33) aus jeweiligen Handlungs- bzw. Interaktionskontexten dem kontextuell begrenzten Handlungswissen (‘knowledgeability’) und der Handlungsmächtigkeit (‘capability’ im Sinne eines Eingreifen- und Gestaltenkönnens) der Akteure entziehen. Gleichzeitig kommt aber diesen globalen Struktur- bzw. Systemzusammenhängen in der Strukturierungstheorie im Sinne von "systems that do not emerge" (Cohen 1989, S. 42) kein eigenständigen Charakter wie in Luhmanns Autopoiesis-Ansatz zu. Vielmehr wird von deren Ein- und Rückbettung in jeweilige Handlungskontexte ausgegangen, d.h. im Sinne eines "realistischen Systemkonzeptes" die "Systemanalyse auf die realen Wechselwirkungen gesellschaftlicher Akteure begrenzt. Ausgehend von der Identifikation unterschiedlicher Grade der ‘Systemhaftigkeit’ sollen dann makrosoziale Ordnungszusammenhänge als Folgen der Wechselwirkungen und Vernetzungen eher intendierter oder eher nichtintendierter akteurieller Handlungsfolgen rekonstruiert werden (Joas 1992, S. 325). Die Kontinuität sozialer Prozesse beruht hier auf einem (quasi)-dialektischen Prozeß, insofern sich raumzeitlich situierte Akteure auf vorgängig etablierte und weit in Raum und Zeit ausgreifende systemische Strukturen - kognitive und normative sowie politische und ökonomische Ressourcenverteilungen - beziehen und diese in ihren meist routinisierten Handlungsvollzügen mehr oder weniger kreativ verändernd reproduzieren.

Für das Giddenssche Handlungsmodell (vgl. Giddens 1992, S. 55ff.), das zwischen unterbewußt gesteuerten sowie durch implizite Wissensformen oder explizite Wissensformen vermittelten Handlungstypen unterscheidet, steht der kontinuierliche und rekursive Charakter von Handlungsprozessen im Vordergrund. In Übereinstimmung mit Grundüberzeugungen des philosophischen Pragmatismus kann soziales Handeln nicht im Sinne intentionaler bzw. teleologischer Handlungstheorien oder im Sinne soziologischer Entscheidungstheorien anhand der (analytischen) Unterscheidung von vorgängigen Handlungsintentionen, bewußten Entscheidungen und nachgängigen Handlungsresultaten interpretiert werden (vgl. Joas 1992a, b). Vielmehr versteht Giddens Intentionen als handlungsinterne Phänomene und fokussiert auf eine auf impliziten Wissens- und Erfahrungsleistungen der Akteure beruhende "reflexive Steuerung des Handelns" (Giddens 1992, S. 55). Ein solches ‘reflexive monitoring of action’ setzt nach Giddens zwar den menschlichen Körper als Grundlage sozialen Handelns voraus, legt aber nicht notwendigerweise die - z.B. in Joas’ (1992a, b) oder in Rammerts (1998) Giddens-Interpretation ausgedrückte - phänomenologische Annahme einer Sozialität fundierenden Rolle des menschlichen Körpers nahe (so Ellrich/Funken 1998, S. 366/ Fn. 42). Vielmehr steht m.E. für Giddens wie auch für die von Collins, Suchman und Wolfe im Zuge ihrer KI-Reflexionen entwickelten Handlungs- und Interaktionsmodelle (vgl. 2.2.2. - 2.2.4.) die durch kognitive bzw. interpretative Leistungen vermittelnde ‘social embeddedness’ jeweiliger Handungsformen, d.h. Sozialität als Produkt von wechselseitigen, sinnhaft vermittelten Kognitionsprozessen auf der Grundlage vorwiegend impliziter Wisssenformen im Vordergrund. Darüber hinaus stimmt das Giddenssche Verständnis routinisierter Handlungsformen mit der Collinsschen Analyse von ‘behaviour specific acts (vgl. 2.2.3.)

überein, insofern er unter Rückgriff auf das Wittgensteinsche Regelkonzept zwar die handlungsorientierende Rolle von "Regeln des gesellschaftlichen Lebens als Techniken oder verallgemeinerbaren Verfahren" (Giddens 1992, S. 73) betont, gleichzeitig sich aber von einem deterministischen bzw. objektivistischen Verständnis der Regelinterpretation und -anwendung distanziert. Nach Giddens werden kognitive und normative Regelvorgaben im Verlauf der sozialen Reproduktion fortlaufend kognitiv vermittelten Neuinterpretationen und entsprechenden handlungsförmigen Modifikationen unterzogen, wobei sich dieser kontinuierliche Transformationsprozeß als abhängig sowohl von dem jeweiligen Handlungswissen der Akteure als auch von deren Handlungsmächtigkeit im Sinne ihrer Zugriffsmöglichkeiten auf handlungsnotwendige politische und ökonomische Ressourcen darstellt.

Der Giddenssche Fokus auf die kognitiv-interpretativen Kapazitäten und Flexibilitäten sozialer Akteure liegt auch seinem Verständnis sozialer Entwicklungs- und Lernprozesse zugrunde, das er insbesondere im Zusammenhang seines methodologischen Diktums einer ‘doppelten Hermeneutik’ erläutert (1984, S. 182ff., 1992, S. 383ff). Nach Giddens kann eine auf einem ‘Verstehen’ sozialer Wissensformen aufbauende soziologische Analyse zu verallgemeinernden Aussagen (‘Erklärungen’) hinsichtlich globaler Struktur- und Systemzusammenhänge führen und so spezifische Entwicklungstendenzen bzw. -pfade des sozialen Wandels aufzeigen. Solche soziologischen Aussagen verkörpern allerdings nach Giddens keine allgemeingültigen Modelle, Regeln oder Gesetze, wobei er ähnlich wie Suchman (vgl. 2.2.4.) - hier allerdings in kritischer Auseinandersetzung mit funktionalistischen und/oder evolutionären Analysen des sozialen Wandels - den unvorsehbaren, d.h. insbesondere nicht evolutionär gerichteten, sondern fortlaufend situativ durch die Akeure korrigierbaren Charakter sozialer Entwicklungsprozesse betont. Hieran anschließend müssen Annahmen einer sozialen Selbstorganisation, wie sie entweder von Maes im Rahmen des reaktiven Paradigma oder von Luhmann im Rahmen seines Autopoiesis-Konzept vertreten werden, die auf kognitiven Reflexionsprozessen beruhenden Gestaltungspotentiale sozialer Akteure entgegengehalten werden. Anschließend legt es die Strukturierungstheorie im Unterschied zu den innerhalb der MAS etablierten Formen eines nicht auf höherwertigen kognitiven Kompetenzen beruhenden maschinellen Lernens (vgl. 3.3.) nahe, soziale Lernprozesse als Produkt kognitiv-interpretativ konstituierter Leistungen der Akteure zu verstehen. Ob jeweilige Innovationen im Sinne eines sozialen Lernens als adäquate bzw. ‘wünschenswerte’ neue Problemlösungen übernommen oder als nicht-adäquate bzw. nichtwünschenswerte Abweichungen abgelehnt, d.h. ob und inwieweit sie innerhalb eines Kollektiv als intelligent oder als nichtintelligent anerkannt werden, sollte folglich auf - durch Computerprogramme nicht umsetzbare - kognitiv vermittelte Interpretationen und Bewertungen der teilnehmenden sozialen Akteure zurückgeführt werden.

3.6. Zusammenfassung: Soziologische Einordnung und Bewertung der MAS

Das zentrale Ziel der MAS ist es, "eine künstliche Gesellschaft (...) zu konstruieren, die sich unabhängig von ihren Architekten und Designern in autonomer Selbstbewegung reproduzieren kann." (Florian 1998, S. 336) Im Rahmen des reflexiven Paradigmas wird

hierbei zwar von autonomen Entscheidungen und Verhaltensweisen der Einzelagenten ausgegangen, allerdings müssen deren Planungs- und Schlußfolgerungskompetenzen weitestgehend vorprogrammiert, d.h. spezifische Umweltbedingungen vor dem Systemverlauf antizipiert und definiert werden (vgl. 3.4.1.). Demgegenüber wird sowohl im Rahmen des reaktiven und als auch des sozialen Agentenparadigmas von einem nicht nur von Eingriffen der Programmierer und Nutzer, sondern auch von den kognitiven Orientierungen der Einzelagenten unabhängigen Selbstlauf von Multiagentensystemen ausgegangen. Dies wird innerhalb der MAS durch die bei Maes wie auch bei Conte/Castelfranchi ausgeführte - aber auch Gassers Ansatz implizit zugrundeliegende – Vorstellung einer emergenten Funktionalität betont. Emergente Funktionalität meint zuallererst, daß keiner der Einzelagenten die kollektiven Effekte des Zusammenwirkens plant bzw. intendiert, aber trotzdem genau solche Effekte im Systemverlauf des Agentensystems entstehen ('emergieren'), die die Koordination und Reproduktion des Gesamtzusammenhangs ermöglichen und sich in diesem (weiten) Sinne als 'funktional' erweisen (vgl. auch Schulz-Schaeffer 1998, S. 135/136). Annahmen eines autonomen Selbstlaufes von Multiagentensystemen berufen sich hierbei auf unterschiedliche, auf die eine oder andere Weise selbstorganisationstheoretisch - 'rotationstheoretisch' im Sinne von Ellrich/Funken 1998 - argumentierenden Analysen einer 'Emergenz des Sozialen'. Maes rekurriert im Rahmen des reaktiven Paradigmas (vgl. 3.4.2.) auf die auch von der Soziologin Knorr-Cetina nahegelegte Vorstellung einer selbstorganisatorischen Emergenz von Sozialität aus den 'mikro-sozialen' Wechselwirkungen und Anpassungsleistungen subkognitiver Agenten. Gassers explizit soziologisches MAS-Konzept (vgl. 3.5.2.) erläutert die Emergenz von Sozialität als das Produkt wechselseitig aufeinander abgestimmer Sozialmodelle sozialer Agenten. Conte/Castelfranchis Analyse einer Emergenz des Sozialen (vgl. 3.5.3.1.) beruft sich zwar programmatisch auf die soziologische Strukturierungstheorie, bleibt aber m.E. widersprüchlich und legt im Zuge der Annahme einer Koevolution von individuellen Handlungsausrichtungen und makrosozialen Strukturzusammenhängen ein an die Systemtheorie Luhmanns anschließendes soziologisches Ordnungsverständnis nahe (vgl. 3.5.3.2.). Insbesondere Luhmanns Behauptung eines autopoietischen 'Selbstlaufes' sozialer Kommunikationsprozesse von gegenüber den jeweiligen individuellen Akteurs- bzw. Agentenorientierungen unabhängigen und sich evolutionär stabilisierenden Lernprozessen stellt sich ein geeignetes Fundament für Ansprüche der MAS hinsichtlich einer Konstruktion künstlicher Gesellschaften dar.[14] Darüber hinaus legt es Luhmanns Ansatz in Analogie zu anderen

14. Angesichts der oben dargestellten (soziologisch betrachteten) Einseitigkeiten der von Maes propagierenten reaktiven Agenten, der Widersprüchlichkeiten von symbolisch-interaktionistisch argumentierenden Modellierungen bei Gasser und den Uneindeutigkeiten von Conte/Castelfranchis strukturierungs-theoretischem Konzept stellt sich m.E. Luhmanns Autopoiesis-Konzept als der aussichtsreichste Kandidat einer soziologischen Fundierung der MAS dar. In diese Richtungen tendieren m.E. auch die 'sozionischen Begründungsversuche' der MAS in Malsch (1998). Interessant ist in diesem Zusammenhang z.B. der an Bourdieus Theorie sozialer Praxis anschließende Theoriekonstruktionsversuch von Florian, der einerseits eingesteht "daß jede theroretische Modellierung aus Bourdieus Sicht die eigensinnige und unscharfe 'Logik der Praxis' zerstört" (1998, S. 336). Andererseits will Florian die Möglichkeit "eine künstliche Gesellschaft als eine soziale Praxis zu konstruieren" (ebd.) einer zukünftigen empirischen Überprüfung überlassen. Anschließend liest Florian Bourdieus Praxismodell durch eine 'luhmannianische Brille', wobei er die 'autopoietisch' verstandene Autonomie der Agentengesllschaft gegenüber den Agenten betont. So

sozialkonstruktivistischen, antiontologisch argumentierenden Positionen (vgl. 2.2.4. sowie auch 4.2.) nahe, Akteurseigenschaften - hier Akteure als Mitteilungsinstanzen und Adressaten von Kommunikationen - als ‘in interactu’ überhaupt erst hervorgebrachte soziale Phänomene zu verstehen (vgl. Fuchs 1991).[15]

---

beschreibt er "Bourdieus ‘Theorie der Praxis’ als eine Theorie über die Grundlagen, wie sich eine soziale Praxis durch Konditionierung individueller Praktiken (Habitus) und durch die Strukturierung sozialer Institutionen (Feld) gewißermaßen ‘autopoietisch’ (sic!, K.S.) durch den Vollzug individueller Handlungen zu reproduzieren vermag." (ebd., S. 338) Anschließend mahnt Florian weiteren Theoriekonstruktions- und Formalisierunsbedarf (interessanterweise auf Seite der VKIler) an: "Die möglichen Früchte, die sich aus einer Umorientierung von der Autonomie der Agenten auf die Autonomie der Agentengesellschaft ergeben könnten, kann die VKI für sich nur selber ernten, aber erst, nachdem sie sich auf die eigenständige Arbeit an der Übersetzung des Habitus-Feld-Konzeptes in entsprechende Programmierkonzeptionen eingelassen hat." (ebd., S. 339) Auch Malsch (1998a) und Ellrich/Funken (1998) beziehen sich auf das mit den Grundabsichten der MAS weitestgehend kompatible Autopoiesis-Konzept von Luhmann und fordern - im Gegensatz zu der von mir in diesem Zusammenhang für notwendig erachteten Theorieentscheidung - an dieses anschließend weitere Theoriekonstruktionsleistungen. Sie reflektieren im Zuge strukturierungstheoretischer bzw. pragmatistischer (Malsch ebd., inbesondere S. 290ff.) oder auch allgemein handlungs- bzw. transformationstheoretische Überlegungen (Ellrich/Funken) die Einseitigkeiten eines emergentinistischen bzw. selbstorganisatorischen Verständnis des Sozialen und melden Bedarf an einem zwischen Handlungs-, Transformations- bzw. Konstitutionstheorien und Selbstorganisations- bzw. Rotationstheorien vermittelnden (dieses synthetisierenden) Ansatz an. So beschreiben Ellrich/Funken am Ende ihres Textes ihre zukünftigen Ambitionen folgendermaßen: "Das Ziel einer ambitionierten Theorie des Umgangs mit Problemen, die emergente Phänomene als Verknotung unterschiedlicher Lösungsstränge ver-steht, muß die Synthese von Transformations- und Rotationstheorien sein." (Ellrich/Funken 1998, S. 387) Ähnlich enden die auch die Ausführungen von Malsch mit der Aufforderung zu weiterer Theorie-konstruktion, die insbesondere an der Modellierung und Formalisierung sozialer Lernprozesse ansetzen soll: "Wie lassen sich gesellschaftliche Lernprozesse, die zur Auflösung von überholten und zur Institutionalisierung von angemessenen Koordinationsverfahren führen, im Medium der Multiagenten-Technologie modellieren und implementieren?" (Malsch 1998, S. 292)

15. Peter Fuchs (1991) versteht im Anschluß an Luhmann individuelles Bewußtsein und soziale Kommunikation als zwar strukturell gekoppelte, aber im Sinne einer Koevolution unabhängig voneinander ihre je eigene Autopoiese vollziehende Systeme, wobei - so die Fuchs/Luhmannsche antiontologische Position - Akteure als Mitteilungsinstanzen und Adressaten erst in der Kommunikationn hervorgebracht bzw. ‘konstruiert’ werden. Hieran anschließend argumentiert Fuchs auf der einen Seite wie die ontologische ‘Hollow-Shell’-Kritik der KI gegen die Vorstellung von Computern als Kommunikationspartnern, insofern ihnen die Kompentenz eines sinnhaften (semantischen) Verstehens - in Luhmannscher Terminologie: das Unterscheiden-Können zwischer einer Mitteilung bzw. einem Mitteilungshandeln und ihrem bzw. seinem Inhalt - nicht zukommt. Auf der andereren Seite ist nach Fuchs damit noch nichts über die epistemologische Vergleichbarkeit bzw. Substituierbarkeit von menschlicher und künstlicher Intelligenz gesagt, die gemäß der Luhmannschen Theoriearchitektur gleichermaßen in der Umwelt sozialer Kommunikationssysteme verortet werden. Wenn es - so Fuchs - KI-Systemen im Zuge ihrer Sprachperformanz bzw. -imitation gelingt, Anhaltspunkte für die kommunikative Unterstellung von Selbstreferenz bzw. Selbstbezüglichkeit zu liefern, können sie durchaus in soziale Kommunikationsverläufe integriert, d.h. von ihren Anwendern als Kommunikationspartner akzeptiert und folglich soziale Kommunikationen, z.B. innerhalb von Agentensystemen ohne die Beteiligung menschlichen Bewußseins, vollzogen werden. Elena Esposito (1993) verweist darauf, daß diese (potentiellen bzw. vermeintlichen) Imitationskompetenzen nicht nur Computeragenten, sondern grundsätzlich allen Computersystemen zukommen. Diese generieren in ihrer Eigenschaft als nicht nur informationsverbreitende Medien, sondern auch als informationsverarbeitende Maschinen unerwartete

Die von mir (3.6.3.3.) präferierte Theorieperspektive der Giddensschen Strukturierungstheorie grenzt sich von diesen Annahmen in zweierlei Hinsicht ab. Auf der einen Seite hält sie der nicht nur Luhmanns Ansatz, sondern auch den Konzepten von Maes und Gasser zugrundeliegenden Vorstellung einer von den (vorgängigen) Agentenorientierungen unabhängigen Emergenz von Sozialität und entsprechenden sebstorganisationstheoretischen Annahmen die Gestaltungspotentiale sozialer Akteure bei der Produktion und Reproduktion von Sozialität gegenüber. Auf der anderen Seite müssen gemäß der Strukturierungstheorie Teilnehmerperspektiven in Form sozialer sozialer Konstruktionen bzw. Deutungen - hier: Zuschreibungen von 'Intelligenz' auf Maschinen - durch eine soziologische Beobachterperspektive ergänzt werden, die empirisch-ethnomethodologisch untersuchend - oder sozialtheoretisch-ontologisch argumentierend - die originären Leistungen menschlicher Akteure in Form deren soziokognitiven Verstehens- und Interpretationsleistungen nachweist. Entsprechend muß anschließend an die KI-Kritik bei Collins, Suchman und Wolfe (vgl. 2.2.2. -2.2.4.) auch angesichts der Produkte der MAS an einer soziologischen Relativierung bzw. Kritik sowohl ontologischer Vergleiche von Mensch und Maschine als auch epistemologischer Ansprüche einer Simulation bzw. Substitution sozialer Akteure und entsprechender Gesellschaftsformen festgehalten werden. Computeragenten können keine dem Menschen vergleichbare Handlungs- bzw. Interaktionspartner darstellen, insofern ihnen keine sinnhafte Wissensformen sowie auf jenen aufbauende Fähigkeiten der gestaltenden Konstruktion von Sozialität im allgemeinen und von sozialem Lernen im besonderen zukommen. Auch der Hinweis auf die vorrangige Bedeutung eines nicht-bewussten Handelns im Sinne eines reaktiven Handelns oder regelförmig-routinisierter Handlungsvollzüge stellt aus strukturierungstheoretischer Perspektive kein plausibles Argument für die imitativen Ansprüche des Agentenparadigmas dar. Vielmehr legt es die Giddenssche Vorstellung eines 'reflex monitoring of action' in Übereinstimmung mit Collins' ethnomethodologischer Analyse der 'bevaviour specific acts' nahe, auch angesichts von routinisierten Handlungsformen von diesen zugrundeliegenden, interpretativ konstituierten und folglich nicht computational darstellbaren Anpassungs-, Reparatur- und/ oder Gestaltungsleistungen sozialer Akteure auszugehen.

Die hier vorgestellte strukturierungstheoretische Perspektive argumentiert im Gegensatz zu konstruktivistischen oder pragmatistischen Positionen (vgl. 2.2.4, siehe auch 4.2.) aus der soziologischen Beobachterperspektive sozialtheoretisch-ontologisch und verweist - potentiell von den Deutungen und Delegationen der Computernutzer abweichend - auf nicht durch Computerprogramme imitierbare bzw. substituierbare Aspekte und Kompetenzen menschlichen Handelns. Aus dieser Perspektive stellen sich die Ansprüche bzw. die Selbstbeschreibungen der MAS sowohl im Sinne ihrer 'starken' als auch ihrer 'schwachen' Agentendefinitionen (vgl. 3.2.2.) als problematisch dar. Sowohl die 'starke' Zuschreibung von Menschen vergleichbaren kognitiven, normativen oder auch emotionalen Eigenschaften als auch die 'schwache' Übertragung von soziologischen Vorstellungen und Begrifflichkeiten wie aktiv zielorientiertes Handeln (Pro-Aktivität), Kommunikation oder Lernen auf Agentensysteme stellt sich angesichts der Computerprogrammen nicht

---

bzw. überrraschende Ergebnisse und demzufolge gegenüber ihren Anwendern eine virtuelle Kontingenz, die allerdings - so argumentiert Esposito aus der soziologischen Beobachterperspektive - nicht mit der doppelten Kontingenz sozialer Situationen verwechselt werden sollte.

zugänglichen sinnhaften Aspekte dieser soziokognitiven Phänomene als irreführend dar. Eine solche strukturierungstheoretisch motivierte Relativierung der Forschungsziele der MAS zielt weniger auf deren konkrete Modelle und Anwendungen, sondern vielmehr auf deren Selbstbeschreibungen und vor allem auf Versuche einer soziologischen Fundierung der MAS, die unter dem Gesichtspunkt der 'Sozialadäquatheit' auf die Konstruktion von - im Vergleich zu nicht soziologisch fundierten Computerapplikationen leistungsfähigeren Systemen - abzielt (Florian 1998, S. 305). Insofern - so interpretiere ich Malschs Nachweis 'metaphysischer Mucken' von Agentensystemen (vgl. 3.5.2) – sozialtheoretisch gehaltvolle Unterscheidungen wie technisch/sozial, profan/heilig, kognitiv/normativ etc. aus der Perspektive von technischen Systemen nicht rekonstruiert werden können, können soziologische Begrifflichkeiten nicht ohne Bedeutungsverlust auf technische Systeme übertragen und sie demzufolge auch nicht als Kriterien für die Bewertung deren (systeminterner) Leistungsfähigkeit herangezogen werden. Aufgrund der kategorialen Differenz von technischen und sozialen Systemen - so meine Vermutung - garantiert eine vermeintliche 'Sozialadäquatheit' der Modellbildung der MAS per se weder erweiterte Problemlösungskapazitäten der MAS hinsichtlich spezifischer Aufgabenstellungen noch im Sinne einer potentiellen Sozialverträglichkeit deren 'bessere' Akzeptanz durch die Nutzer.

4. Potentielle Fragestellungen für den Forschungsbereich Sozionik

4.1. ‘Schwache Sozionik’: Metaphernmigration zwischen Informatik und Soziologie?

Die hier vorgestellte strukturierungstheoretische Analyse der MAS kommt zu einer eher ‘pessimistischen’ Einschätzung hinsichtlich von in der Tradition der KI erhobenen Ansprüchen der technischen Konstruktion, Imitation und/oder Simulation von menschlichen Akteuren oder sozialen Kollektiven vergleichbaren Problemlösungsleistungen. Vielmehr geht sie davon aus, daß dementsprechende Ansprüche der KI im allgemeinen wie auch der VKI bzw. MAS im besonderen mit verkürzten Vorstellungen hinsichtlich spezifischer Eigenschaften menschlichen Handelns und Interagierens (Kreativität, Emotionalität, Reflexivität, (Selbst-) Bewußtsein, Verantwortungs- bzw. Vertrauensfähigkeit, etc.) einhergehen. Allerdings erscheint eine solche Kritik im Rahmen der VKI bzw. MAS wie auch der ‘Sozionik’ (vgl. 1.2.) nicht notwendigerweise erforderlich, insofern eine Vielzahl deren Protagonisten solche starken ontologischen und/oder epistemologischen Ansprüche gar nicht erheben und vielmehr die Differenzen zwischen Sozialsystemen und künstlichen Agentensystemen reflektieren (vgl. Florian 1998, S. 336, Malsch 1998a, S.286ff.). Gleichzeitig zieht aber die hier vorgestellte strukturierungstheoretische Analyse der MAS einige Einschätzungen hinsichtlich der innerhalb der Sozionik angestrebten Forschungsziele nach sich (vgl. 1.2.).

Erstens stellt die hier gewählte strukturierungstheoretische Theorieperspektive der im Rahmen einer ‘starken’ Sozionik anvisierten Simulation von Sozialprozessen und potentiell aus diesen abgeleiteten Sozialprognosen die Einsicht in den grundsätzlich offenen und kontingenten Charakter sozialer Handlungsprozesse, d.h. die "eigensinnige und unscharfe ‘Logik der Praxis’" (Florian 1998, S. 336) gegenüber. Zwar geht auch Giddens von dem Wert und der Möglichkeit von von jeweiligen Handlungsorientierungen der Akteure abstrahierenden ‘institutionellen’(potentiell formalisierbaren und computational darstellbaren) Analysen aus, reflektiert aber gleichzeitig deren vorläufigen, einseitigen sowie ergänzungsbedürftigen Charakter einer solchen soziologischen Herangehensweise. Vor allem das Ziel der Sozionik, aus Computersimulationen - und nicht z.B. aus empirisch-historischen Untersuchungen - Evidenzen für die soziologische Theoriebildung zu gewinnen, stellt sich nicht nur aus der Perspektive der Strukturierungstheorie, sondern letztlich aus jeder nicht systemtheoretisch-kybernetisch orientierten Theorieperspektive als fragwürdig dar.

Zweitens muß auch aber die im Rahmen einer ‘schwachen Sozionik’ postulierte Aufgabe der Soziologie als einer "Hilfs- bzw. Grundlagenwissenschaft für die KI" (Malsch/ Müller 1998, S. IV) gemäß der hier vertretenen Position weiter differenziert und teilweise relativiert werden. Die von Florian (1998) vor dem Hintergrund (vermeintlicher) "struktureller Homologien zwischen menschlichen und künstlichen Gesellschaftsformen" (ebd., S. 303) eingeforderte soziologische Fundierung der VKI bzw. MAS, die im Zuge einer nicht ‘mimetischen’ (quasi-metaphorischen), sondern ‘analogen’ Übertragung soziologischer Begriffe und Theorien auf Agentensysteme zu leistungsfähigeren KI-Applikationen gelangen will, stellt sich m.E. aufgrund des ‘Bedeutungsverlustes’ soziologischer Begrifflichkeiten bei ihrer Umsetzung durch technische Systeme als irreführend dar. Vielmehr ziehen solche Absichten - so wird hier vermutet - überhöhte Erwartungen bezüglich einer Simulation des Sozialen, d.h. letztlich ein inadäquates

Verständnis der MAS-Applikationen sowohl hinsichtlich ihrer menschlichen und/oder sozialen Kapazitäten potentiell weit übertreffenden Leistungspotentiale als auch ihrer grundsätzlichen Grenzen bzw. Schwächen nach sich.

Aber auch wenn die Sozionik wie bei Malsch (1998b, S. 48ff.) die Unmöglichkeit von (1:1-) Übertragungen soziologischer Kategorien auf Agentensysteme reflektiert, sich z.B. in Auseinandersetzung mit Gassser (vgl. 3.5.2.) von der Annahme einer soziologischen Fundierbarkeit der MAS distanziert und von einem komplex-vielschichtigen Übersetzungs- bzw. Transformationsprozeß im Sinne einer ‘Metaphernmigration’ zwischen Soziologie und Informatik ausgeht, bleibt auch hier die Rolle der Soziologie bezüglich der MAS unklar. So schlägt Malsch vor, nicht die Adäquatheit der in der MAS verwendeten Sozialvorstellungen zu bewerten, sondern vielmehr an computerwissenschaftlichen Kriterien anzusetzen:"Rechenzeit und Tempogewinn, Speicherbedarf und -optimierung, algorithmische Effizienz, architektonische Eleganz und softwaretechnische Wartbarkeit von Sprachen, Programmen und Werkzeugen sind die hier gültigen computerreferenziellen Bewertungskriterien, an denen sich die innovatorischen Leistungen der VKI (...) messen lassen müssen." (ebd., S.50) Anschließend stellt sich Malsch die Frage, "ob sich die VKI in ähnlicher Weise durch soziale Metaphern anregen lassen kann wie die klassiche KI oder die Artificial-Life-Forschung durch Geist- und Lebensmetaphern oder die Bionik durch Vorbilder aus der Biologie". Allerdings gibt es für sein anvisiertes Ziel, die "innovatorische Inspirationskraft" (ebd, S. 52), d.h. den computerwissenschaftlichen Erkenntnisgewinn sowie den technologischen Innovationswert von Sozialmetaphern innerhalb der MAS zu bewerten, strenggenommen gar keine soziologischen Begrifflichkeiten und Kriterien. Im Unterschied zu Malschs Absichten einer sozionischen Integration computerwissenschaftlicher Fragen impliziert m.E. die von ihm als Kontingenzthese referierte Annahme, derzufolge "der technische Fortschritt der KI-Forschung nicht mit Fragen einer angemessenen Rezeption soziologischer Ansätze und Sachverhalte zu tun hat" (ebd. 51), daß diesbezügliche Fragen den Informatikern als den Experten für Fragen der Formalisierbarkeit spezifischer Konzepte sowie der technischen Machbarkeit, Funktionalität und Innovativität von jeweiligen Applikationen überlassen bleiben.

Demgegenüber sollte sich m.E. eine originär soziologische Perspektive auf die MAS vor dem Hintergrund des Nachweises einer kategorialen Differenz von technischen und sozialen Systemen darauf fokussieren, ob und inwieweit sich die Innovationen der MAS bei ihrer Implementation z.B. in jeweilige Arbeitsprozesse als erfolgreich oder als (wie auch immer näher zu spezifizierend) ‘sozialverträglich’ erweisen. Im Zuge eines solchen sozionischen Perspektivenwechsels rücken die Antizipation und die rekonstruktive Analyse der Ordnungsformen von - gleichermaßen Agenten bzw. Agentensysteme und menschliche Akteure als User umfassenden - Hybridsystemen in den Vordergrund. Gemäß der konstruktivistischen Techniksoziologie gehen bereits in die Definition wie auch die Gestaltung von technischen Artefakten explizit oder implizit Vorstellungen des - faktisch und gewünschten - sozialen Verhaltens der menschlichen Anwender ein. Sozialkulturelle Modelle und Sichtweisen werden den Formen und Funktionsweisen technischer Artefakte z.B. vermittels sogenannter ‚Leitbilder’ ‘eingeschrieben’ (vgl. zusammenfassend Akrich 1992). Entsprechende soziologische Reflexionen und Stellungnahmen, die entweder auf sozialthereotischen Überlegungen oder auch auf empirischen Untersuchungen der jeweiligen Anwendungskontexte beruhen, können dann, wenn sie in den

Konstruktionsprozeß der Informatiker eingebracht werden, zwar weniger Modifikationen der konkreten technischen Programmarchitekturen, wohl aber Veränderungen deren Annahmen in Bezug auf die zu erwartetenden oder auch wünschenswerten soziotechnischen Effekte, d.h. bezüglich der Modellbildung der zu implementierenden soziotechnischen Systeme nach sich ziehen. Im Rahmen eines "reflexive mode of technology development" (Rammert 1998b, S.13) sollte dementsprechend die Soziologie versuchen, Einsichten z.B. in die kategorialen Diffferenzen von Agentensystemen und Sozialsystemen in Analysen der sozialen Verfasstheit (Handlungstypen, Organisationsformen, etc) von spezifischen Anwendungskontexten zu überführen sowie (im Sinne von Stars Durkheim-Test, vgl. 3.1.) die divergierenden Einschätzungen von Systementwicklern und -usern hinsichtlich der ‘sozialweltlichen Brauchbarkeit’ von computationaler Innovationen in die Systementwicklung der MAS miteinzubringen. Eine solche Herangehensweise stellt sich in Abgrenzung von den Perspektiven der Informatiker als eine genuin soziologische dar, insofern über die Bewertung der technischen Funktionalitäten hinaus nur eine soziologische Analyse der Anwendungszusammenhänge die dort vorherrschenden sozialen Probleme (Kommunikationstörungen, Interessenkonflikte, Machthierachien), Bedürfnisse und Werte aufzeigen und somit die Erwartungen bzw. Anforderungen der User an die Implementation von neuen Softwaretechnologien interpretieren kann. Eine strukturierungstheoretische Soziologie legt es in diesem Zusammenhang nahe, z.B. durch ethnomethodologische Analysen (teilnehmende Beobachtung etc.) die praktisch-praktisch-impliziten Handlungs- und Wissensformen sowie die vorwiegend nicht-explizierten Bedürfnisse und Interessen der User innerhalb jeweiliger Anwendungskontexte in den Blickpunkt zu rücken. Eine solche Analyse der jeweiligen ‘soziotechnischen’ Anwendungskontexe differeriert somit potentiell auf der einen Seite von expliziten Wahrnehmungen und Bewertungen der (zukünftigen) Computernutzer und auf der anderen Seite - so meine Vermutung vor dem Hintergrund entsprechender Überlegungen zur Organisationsform ‘Krankenhaus’ (vgl. Burkhard/Rammert 1999) - von den Einschätzungen und Modellen der MAS-Konstrukteure, die oft entweder implizit oder explizit - sowie entweder empirisch oder präskriptiv-normativ - Rational-Choice-Modelle oder systemtheoretisch-kybernetische Modelle der Anwendungskontexte bevorzugen.

4.2. Mensch-Maschine-Interaktion in hybriden Systemen

Die VKI bzw. MAS als das avancierteste Projekt der KI liefert Softwareprogramme und/ oder Roboter, deren Operationsweisen sich in einem zunehmenden Maße menschlichen Verhaltensweisen annähern. Allerdings stellen diese Computeragenten - so wurde in dem vorliegenden Text argumentiert - keine sozialen Akteure im Sinne von dem Menschen vergleichbaren Handlungs- bzw. Interaktionspartner dar. Gemäß der hier vertretenen Position sollten dann vor dem Hintergrund der Einsicht in die Differenz von menschlichem Handeln und machinellem Operieren - d.h. aufbauend auf einem ‘engen’ bzw. ‘anspruchsvollen’ soziologischen Handlungsbegriff - die spezifischen neuen Qualitäten der Operationsweisen technischer Agenten und die anschließenden neuen Ordnungsformen von

- technische Agenten, menschliche Akteure als User sowie spezifische Mensch-Maschine-Kooperationen umfassenden - ‘Hybridgemenschaften’ analysiert werden.

Der Analyse der Operationsweisen von Agenten nähert sich Schulz-Schaeffer (1998) mit der Abgrenzung unterschiedlicher Handlungstypen. Er bevorzugt hierbei allerdings einen eher ‘breiten’ bzw. ‘anspruchslosen’ Handlungsbegriff und will im Gegensatz zu einem ‘engen’ sozialtheoretisch-ontologischen Handlungsbegriff Handeln als Produkt jeweiliger sozialer Delegationen verstehen. So setzt er an der ‘Actant-Network-Theory’ von Callon/ Latour (vgl. Latour 1987, 1988, 1991, 1998, Callon/Latour 1992) an, die Netzwerke bzw. ‚Assoziationen’ von menschlichen und technischen Entitäten aus einer semiotischen Theorieperspektive untersucht und die analysierten (Alltags-) Technologien im Sinne eines strikten ‘Symmetrieprinzips’ als ihren menschlichen Gegenübern bzw. Usern ‘gleichwertige’ (semiotische) Aktanten beschreibt. Hierbei argumentiert Schulz-Schaeffer (ebd., S. 143ff.), daß die technischen Agenten der MAS im Unterschied zu bei Callon/Latours als Aktanten analysierten Alltagstechnologien nicht nur zu einem ‘effektiv situiertem’, sondern darüber hinaus zu einem ‘genetisch situiertem’ Handeln fähig sind. Ein effektiv-situiertes Handeln eines Artefaktes - z.B. eines bei Latour (1988) untersuchten automatischen Türschließers - liegt dann vor, wenn dessen sozial zugeschriebene Bedeutungen oder auch seine faktischen Wirkungen von der jeweiligen ‘sozialen’ Situation abhängen. Demgegenüber setzen nach Schulz-Schaeffer die Produkte der MAS ein genetisch-situiertes Handeln um, insofern hier der Agent im Unterschied zum Falle des effektiv situierten Handelns "über unterschiedliche Handlungsoptionen und eine wie auch immer rudimentäre Sprache verfügt, die es ihm erlaubt, bestimmte Gegebenheiten der Situation als Information bei der Handlungswahl zu nutzen" (ebd. 161).

Hieran anschließend gelangt Schulz-Schaeffer wiederum zur der in einem ‘engen’ Handlungsbegriff ausgedrückten Einsicht in die Differenz zwischen auf ‘rudimentäre’, d.h. insbesondere technisch-kybernetische (Agenten-) Sprachen rekurrierenden Computeragenten auf der einen und innerhalb soziokultureller Zusammenhänge sprach- bzw. kommunikationsfähigen, d.h. zu interpretativen Deutungen und gestaltenden Delegationen befähigten menschlichen Akteuren auf der anderen Seite. "Die Deutung der Situation erfolgt nicht durch die nichtmenschlichen Aktanten selbst, sondern durch menschliche Akteure. (...) Es gibt keine gemeinsame Sprache der menschlichen und nichtmenschlichen Aktanten, sondern nur die Sprache der menschlichen Akteure, die beschreibt, was menschliche Akteure oder an ihrer Stelle technische Artefakte zum Ge- oder Mißlingen soziotechnischer Zusammenhänge beitragen." (ebd., S. 148). Somit unterscheidet Schulz-Schaeffer – wenn ich ihn richtig verstehe - drei Handlungstypen, nämlich ein einfachen Alltagstechnologien zukommendes effektiv-situiertes Handeln, ein autonomen Agenten insbesondere im Hinblick auf ihre informationsverarbeitenden Operationswahlen zukommendes genetisch situiertes Handeln sowie - in Übereinstimmung mit den Intentionen eines ‘engen Handlungsbegriffs - ein originär menschliches, auf Sprache bzw. Welt interpretierenden Fähigkeiten aufbauendes soziales Handeln.

Ähnlich wie Schulz-Schaeffer argumentiert auch Rammert (1998a, b), auch wenn er vorrangig auf die Interaktions- und Kooperationsformen zwischen technischen Agenten und menschlichen Akteuren in hybriden Gemeinschaften fokussiert. Auch Rammert stellt im Sinne eines ‘weiten’ Handungsbegriffs auf eine sozialkonstruktivistische Perspektive ab, die weniger auf die sozialtheoretische Bestimmung der Eigenschaften (und Unterschiede)

von Computeragenten und menschlichen Akteuren, sondern vielmehr - wie z.B. auch Fuller (1994), Pickering (1993) und Knorr-Cetina (1992) - empirisch bzw. pragmatistisch auf die Formen und Bedingungen der Delegation von ‘Handlungsträgerschaften’ (agency) auf technische Artefakte fokussiert. "Gegenstand einer soziologischen Untersuchung (...) sollen also die Emergenz und die Verteilung von menschlicher und materieller Handlungsträgerschaft sein, wie sie aus den sozialen Praktiken der menschlichen Nutzer, den sozialen Operationen der technischen Objekte und aus ihren wechselseitigen Verknüpfungen entstehen. Nicht die vorgängige Zuteilung von Kompetenzen auf die unterschiedlichen Agenten, sondern das Studium der Performanzen und die nachträgliche Zurechnung von ‘Agency’ stehen zur Debatte." (1998a, S. 118) Im Anschluß an diese Auffassung von ‘agency’ als einer in jeweiligen Kontexten aufteilbaren und in diesem Sinne nicht qualitativ bestimmbaren, sondern ‘quantifizierbaren‘ Ressource reflektiert Rammert im Anschluß an Pickerings Latour-Kritk (vgl. Pickering 1995) die Nicht-Symmetrie von Aktant und Akteur und verweist ähnlich wie m.E. auch Schulz-Schaeffer auf einzig menschlichen Akteuren zukommende intentionale Interpretations- bzw. Deutungsleistungen, die seines Erachtens allerdings nicht wie bei Schulz-Schaeffer sprachlich bzw. sozial-kommunikativ, sondern vielmehr gemäß phänomenologischen Grundannahmen (vgl. 2.2.2.) vorrangig körperlich konstituiert sind. "Nur die menschlichen Agenten scheinen eine Intentionalität aufzuweisen, Ziele zu konstruieren, die sich auf zukünftige Zustände beziehen, und diese anzustreben.(...) Nur bei ihnen kann aufgrund ihres körperlichen Weltbezugs erwartet werden, daß sie über die Reflexivität verfügen, die Bedeutung von Operationen und Informationen für sich und die Welt zu verstehen."(ebd., S. 119) Anschließend grenzt Rammert die auf Intersubjektivität beruhende Interaktion zwischen menschlichen (körperlich konstituierten) Akteuren von sogenannten ‘Interaktivitäten‘ zwischen menschlichen und technischen Agenten ab, denen er wiederum die Vorstellung quasi-interaktionaler Wechselwirkungen zwischen Einzelagenten innerhalb eines Agentensystems in Form einer "Interobjektivität" (ebd., S.121) gegenüberstellt. Mit dem Begriff ‘Interobjektivität’ reflektiert er einerseits auf der Hardwareebene von Computern die Wechselwirkungen physikalischer Spannungszustände zwischen Computeragenten und andererseits auf der Softwareebene im Sinne von ‘Intertextualität’ die quasi-eigenständigen bzw. teilautonomen Potentiale der (semiotisch verstandenen) Informations- bzw. Zeichenverarbeitung von Computeragenten. "Agenten sind demnach Sprachprogramme, die teilautonom und teilintelligent ihre Zeichenstrukturen ohne direkte Kontrolle verändern. Man könnte dieses Verhältnis als eines der ‘Intertextualität’ bezeichnen, weil sich Texte gleichsam selbständig zu anderen Texten in Relation setzen und sich wechselseitig verändern. Schließlich bedarf es jedoch der herstellenden, intervenierenden und auslösenden Interaktion (...) mit handelnden und wahrnehmenden menschlichen Körpern, wie in Technikтheorien und in Konstruktionspraktiken so leicht vergesssen wird." (ebd., S. 122)

Rammert betont mit seinen Überlegungen zu den semiotischen Kapazitäten von technischen Computeragenten - entsprechend der Zielrichtung von Schulz-Schaeffers Unterscheidung von einem effektiv-situierten Handeln und einem Computeragenten vorbehaltenen genetisch-situiertem Handeln - die spezifisch neuen, autonom-eigenständigen und sich potentiell Eingriffsversuchen der menschlichen Nutzer widersetzenden Eigengesetzlichkeiten und -dynamiken von Computeragenten. Hinsichtlich

der hier relevanten Frage der ‘Interaktivität’ zwischen Mensch und Maschine legt es diese Argumentation nahe, innerhalb jeweiliger Hybridgemeinschaften von einer sukzessiven Verschiebung hin zu den Agenten als ‘Aktivitätsinstanzen‘ und einem potentiell ‘gleichartigen’ Charakter der wechselseitigen Anpassungsleistungen zwischen Agenten und Usern auszugehen. Eine solche Interaktivitäts-Analyse der in spezifischen (zeitlich begrenzten) Kontexten ‘emergierenden’ Verteilung von Handlungsträgerschaften zwischen Menschen und Artefakten will angesichts der Kapazitäten der Produkte der MAS nicht im vorhinein ausschließen, daß menschliche Akteure einen Großteil ihrer Handlungsträgerschaften erfolgreich an Agenten delegieren und darüber hinaus als Nutzer bereit sind, im Sinne einer ‘experimentellen Interaktivität’ (Rammert 1998b) sich zumindest bis zum Falle der Enttäuschung so zu verhalten, ‘als ob’ technische Computeragenten dem Menschen vergleichbare Handlungspartner darstellen.

Gemäß der obigen strukturierungstheoretischen Analyse der Potentiale der MAS müssen allerdings Rammerts Überlegungen in zweierlei Hinsicht kommentiert werden. Auf der einen Seite dürfen die Informations- bzw. Zeichenverarbeitungskompetenzen von technische Agenten nicht mit sozial eingebetteten Kommunikationskompetenzen, d.h. insbesondere mit Sprache bzw. ‘social worls’ sinnhaft hervorbringenden und rezipierenden Fähigkeiten menschlicher Akteure verwechselt werden.[16] Auf der anderen Seite muß der Fokus auf ‘agency’ als einem situativen Zuschreibungsprodukt, d.h. die soziologische Teilnehmerperspektive auf die Deutungen und Delegationen der Teilnehmer durch den Fokus auf, nur aus einer (ethnomethodologischen) Beobachterperspektive zugängliche Sinnkonstruktions- und Sinnrezeptionsleistungen sowie auf entsprechende Gestaltungspotentiale sozialer Akteure ergänzt werden. Diese originär menschlichen Kapazitäten entziehen sich nicht nur einer Umsetzung durch die Softwareagenten und Roboter der MAS, vielmehr kommt ihnen auch eine konstitutive Rolle innerhalb jeweiliger Mensch-Maschine-Interaktivitäten zu. Wie Collins gezeigt hat (vgl. 2.2.4.), erklären die den Teilnehmern selbst meist nicht bewußten Anpassungs- und Reparaturleistungen sozialen Handelns, warum KI-Systeme - hier: Computeragenten - trotz ihres faktisch differenten Operierens und entsprechenden Grenzen einer Imitation oder auch Substitution menschlichen Handelns von den Anwendern als dem Menschen vergleichbar ‘intelligent’ akzeptiert und entsprechend in jeweilige Arbeits- und Interaktionszusammenhänge integriert werden. Darüberhinaus kann die Neigung der User, Computer als dem Mensch vergleichbare Partner zu akzeptieren, sozialpsychologisch auf ihre Eigenschaft als vielschichtige Projektionsmedien im Sinne von ‘evokativen Objekten’ (Turkle 1984) oder ‘Grenzobjekten ‘ (Schachtner 1993) zurückgeführt werden.

Analysen von Mensch-Maschine-Interaktivitäten in Hybridgemeinschaften sollten der hier vertretenen Position folgend auf Einsichten sowohl in die spezifisch neuen Qualitäten von Computeragenten als auch in die grundsätzlichen Differenzen von menschlichem Handeln und maschinellen Operieren aufbauen und dann sowohl pragmatistische, ethno-methodologische wie auch sozialpsychologische Überlegungen integrieren. Anschließend

16. Rammerts Argumentation legt es E. an dieser Stelle nahe, aufgrund ihrer phänomelogischen und nicht sprachlich-komunikationstheoretischen Herleitung menschlichen Sinnverstehens die Differenz zwischen einer sprachlich konstituierten und soziokognitiv eingebetten symbolischen Interaktion zwischen menschlichen Akteuren und einer zeichenhaft (semiotisch) vermittelten ‘Intertextualität’ zwischen Computeragenten (wiederum) zu ‘verwischen’.

sollten sie auf die Entwicklung von soziologischen Kriterien abzielen, anhand derer bestimmt werden kann, in welchen Kontexten die Delegation von Handungsträgerschaften an Agenten sich als (pragmatisch) sinnvoll und in welchen Kontexten sie sich als nicht sinnvoll darstellt. Inbesondere im Sinne einer Unterstützung der Konstruktion und Implementation von MAS-Produkten und der entsprechenden Gestaltung von soziotechnischen Zusammenhängen kann dann die Soziologie durch empirische Rekonstruktionen oder durch modellhafte Antizipationen jeweils zukünftiger Anwendungszusammenhänge die Vorzüge wie auch Deffizienzen von Computeragenten zu identifzieren versuchen. Ziel sozionischer Analysen von Agent-Akteur-Hybriden wäre es dann zu bestimmen, "wie in einer hybriden Sozialform die jeweiligen Vorzüge am besten genutzt und bewahrt und gleichzeitig die jeweiligen Schwächen am günstigsten kompensiert werden können." (Rammert ebd., S. 119)

4.3. Vor- und Nachteile des Agentenparadigmas

Das Agentenparadigma stellt ein immenses Innovationspotential im Rahmen der Konstruktion von neuen Softwaretechnologien dar. Allerdings ist umstritten, ob und inwieweit die Orientierung an menschlichen und/oder sozialen Vorbildern bei der Konstruktion von Computersystemen sinnvoll ist. Auf der einen Seite wird das Paradigma ‘intelligenter Agenten’ insbesondere im Zusammenhang mit den neusten Formen eines objektorientierten Programmierens als die Zukunftstechnologie der Informatik dargestellt (Brenner u.a. 1998). Auf der anderen Seite finden sich generelle Einwände gegen das KI-Paradigma im allgemeinen und das Agentenparadigma im besonderen. So plädiert Shneiderman (1997) für ein erweitertes Verständnis von neuen Computertechnologien, das sich von der Orientierung an bzw. dem Vergleich mit menschlichen oder sozialen Vorbildern löst. "Computer supported cooperative work, hypertext/hypermedia, multimedia, information visualization, and virtual reality are powerful technologies that enable humans users to accomplish tasks that no human has ever done. If we describe computers in human terms, we run the risk of limiting our ambition and creativity in the design of future computer capabilities."(ebenda, S. 98)

Auch hinsichtlich der Wünschbarkeit von autonom-eigenständig operierenden Softwareagenten und Robotern gibt es Pro- und Kontraargumente, die sich in den Computerwissenschaften zuallerst auf die im Rahmen des Agentenparadigmas vollzogene Abschaffung der Möglichkeiten einer ‘direkter Manipulation’ der Computeroperationen durch die User beziehen (Lanier/Maes 1996, Sheiderman/Maes 1997). Im Hintergrund steht hierbei die Frage, ob und inwieweit sich eine Konstruktion autonomer Agentensysteme als sinnvoll darstellt, welche nicht nur die Emergenz neuer Problemlösungsleistungen, sondern auch eine unvorhersehbare und somit potentiell nichtgewünschte bzw. dysfunktionale Systementwicklungen miteinkalkuliert und somit letztlich die Ablösung eines traditionellen Technikbegriffs - etwa im Sinne von Technik als wiederholbaren maschinellen Ausführungen gewünschter Effekte (Schulz-Schaffer 1999) - nach sich zieht. Die Protagonisten des Agentenparadigma verstehen das Agentenparadigma als einen entscheidenden Schritt in Richtung von fehlertoleranten, flexiblen sowie vor allem an den User anpassungsfähigen bzw. ‘lernfähigen’ Technologien (vgl. Maes 1994a, Negroponte

1997). Hierbei wird gefordert, die Vorstellung von Computerprogrammen als passiven, Schritt für Schritt auf direkte Manipulationen in Form von Benutzerinstruktionen angewiesene Entitäten durch im Rahmen des Agentenparadigmas bereitgestellte Konzepte der Delegation von Agentenkompetenzen abzulösen. "Direct manipulation will have to give away to some form of delegation. (...) Instead of exercising complete control (and taking responsibilitiy for every move the computer makes), people will be engaged in a cooperative process in which both human and computer agents initiate communication, monitor events and perform tasks to meet a user's goal." (Maes 1996. S. 1) In diesem Zusammenhang wird auch betont, daß Agenten weniger menschliches Handeln im Sinne der KI substituieren, sondern vielmehr die Handlungspotentiale der Nutzer 'vermehren', wie Maes am Beispiel des sogenannten 'Personal Digital Assistant' erläutert."An agent is not a replacement for human intelligence. Agents are meant to augment the user. In particular, the agent's role includes making personalized suggestions (based on patterns in the user's actions as well as patterns among users), making sure the user doesn't miss things he's obviously be interested in, helping the user remember past actions, and so on".(Maes in Lanier/Maes 1996, 4/S.2) Demgegenüber fragen die Kritiker des Agentenparadigmas, ob es überhaupt wünschenswert ist, technische Artefakte nach dem Vorbild sozialer Prozesse so zu modellieren, daß sich ihr Operieren ähnlich wie das sozialer Agenten als unvorhersagbar und somit potentiell 'Out of control' (Brooks) darstellt. Sie gehen wie z.B. Lanier davon aus, daß Agenten keine menschlichen Lernformen vergleichbaren Fähigkeiten erlangen können und folglich deren Kontrollier- bzw. Steuerbarkeit teilweise durch vorgängige Programmierung, aber vor allen Dingen durch die Eingriffsmöglichkeiten der Anwender fortlaufend gewährleistet sein müsse. "Functionality buries in an 'agent' would become more useful if it were replaced under conscious control of the user, but this requires the very difficult discipline of good user interface design." (Lanier in ebd., 1/S. 1-2) Weiterhin befürchtet Lanier wie auch die soziologischen KI-Kritiker Collins, Suchman und Wolfe, daß (sozialpsychologisch erklärbare) 'anthropomorphe' Auffassungen im Anschluß an beschränkte Eingriffsmöglichkeiten der User auch deren Verantwortlichkeiten herabsetzen und somit die Vernachlässigung spezifischer Fertigkeiten der Anwender, d.h. z.B. der oben genannten Eigenschaften menschlichen Handelns (Kreativität, Emotionalität, Selbstbewußtsein, etc.) nach sich ziehen. "With agents, we are building inadaequate ideas about ourselves into the functional fabric of our actions. Agents are just an insidious way of getting entangled in our own fantasies."(ebd., 8/S.2)

Diese Frage der Vor- und Nachteile der autonomischen Selbstständigkeit von Agenten stellt sich nicht nur hinsichtlich der Programmierung der Operationsweisen von Computeragenten, sondern auch bei deren Repräsentation gegenüber den Usern vermittels (potentiell die faktischen Operationsformen der Programme nicht abbildender oder gar konter-karierender) graphischer Darstellungen des Mensch-Maschine-Interface. So scheinen eine Reihe von empirischen Analysen der Mensch-Maschine-Kooperation bzw. des Mensch-Maschine-Interface zu zeigen, daß menschliche User von der Mensch- bzw. Sozialähnlichkeit sowohl der Operationsweisen als auch und vor allem der graphischen Gestaltungen von Computeragenten - z.B. in Form von Emotionen verkörpernden 'gesichtsähnlichen' Darstellungen - profitieren (vgl. z.B. Sproull u.a. 1997). Umgekehrt kritisiert Shneidermann (ebd.) sowohl die Programmierung autonomer Operationsweisen

von Agenten als auch deren ‘anthropomorphen’ Darstellungen und behauptet: „Users want the feeling of mastery, competence, and understanding that come frome a predictable and controllable interface." (ebenda, S.98)

Eine weitere interessante Frage in diesem Zusammenhang ist, ob und inwieweit die Zuschreibung eines autonomen Agentenstatus auf Computeragenten auch ethische und rechtliche Attribute, d.h. z.B. deren Status als Rechtspersonen oder als (rechtlich anerkannte) Transaktionspartner innerhalb ‘elektronischer Märkte’ nach sich ziehen sollte. Innerhalb der MAS finden sich eine Vielzahl diesbezüglicher Überlegungen, die auf der einen Seite Agenten als von ihren Programmierern und Nutzern unabhängige Träger von sozialen Rechten und Verpflichungen anerkennen wollen und auf der anderen Seite programmierungstechnisch darauf abzielen "(...) to build into agents an attitude about such rights."(Krogh 1996, S. 1 und ff., vgl. auch Huhns/Singh 1998)

Eine soziologische Bewertung der Vor- und Nachteile des Agentenparadigma sollte m.E. alle diese Fragen aufnehmen und fragen, in welchen Kontexten die Autonomie und somit potentielle Unkontrollierbarkeit, die graphische Darstellung bzw. Inszenierung von ‘Menschenähnlichkeit’ wie auch ein eigenständiger rechtlicher Status von Computeragenten angemessen ist. Allerdings sollten m.E. anthropomorphe Semantiken bei der Beschreibung von Computeragenten vermieden sowie hieran anschließende Zuschreibungen und Delegationen von rechtlichen Attributen auf Agenten mit Vorsicht behandelt werden. Auch wenn Handlungsträgerschaften im Sinne von Fuller (1994) teilbare und je nach Kontext delegierbare Ressouren darstellen, stellt sich die Vorstellung von Computeragenten als ethischen und rechtlichen Akteuren als irreführend dar, insofern für jene aufgrund ihrer Nichtpartizipation an sinnhaften Sozialwelten die sozialen Rechten zugrundeliegenden Wertbindungen und Verpflichtungen - wie auch mit Rechten und Pflichten zusammenhängende sozialpsychologische Phänomene wie Schuld, Vertrauen, etc. - keine Rolle spielen. Die Vorstellung von ‘agency’ als einem teilbaren Gut zieht auch und vor allem im Falle der Zuschreibung rechtlicher Verpflichtungen auf Computeragenten die gleichzeitige Abwertung bzw. Minimierung rechtlicher Verpflichtungen auf Seiten der Programmierer und/oder Nutzer nach sich, auf deren Aufwertung aber m.E. eine soziologische Reflexion der Informatik im allgemeinen und eine soziologische Analyse von Mensch-Maschine-Interaktivitäten gerade abzielen sollte.

Zitierte Literatur

Ahrweiler, P./Gilbert, N. (1998) "Computer simulations in science and technology studies", Springer, Berlin/NY.

Avouris, N./Gasser, L. (ed.)(1992a) "Distributed artificial intelligence: Theory and praxis", Kluwer Academic Publisher, Dordrecht u.a.

Avouris, N./Gasser, L. (1992b) "Introduction" in dieselben (1992a).

Bamme, A. u.a. (1983) "Die Maschine, das wir selbst sind. Zur Grundlegung einer Sozialpsychologie der Technik" in "Psychosozial", Vol. 18.

Bates, J. (1994) "The role of emotion" in "Communications of the ACM", Vol. 37/7.

Bond, A./Gasser, L. (ed.) (1988) "Readings in DAI", Morgan Kaufman, San Mateo.

Bradshaw, J.(1997a) "An introduction to software Agents" in derselbe (ed.) (1997b).
Bradshaw, J. (ed.) (1997b) „Softwareagents“, AAAI Press/The MIT Press, Menlo Park u.a..
Braun, H. (1998) "The role taking of technology" in Malsch (Hg.) (1998).
Brenner, W. u.a. (1998) "Intelligente Softwareagenten - Grundlagen und Anwendungen", Springer, Berlin.
Brooks, R. (1991)" Intelligence without representation" in "Artificial intelligence", Vol. 47, 1991.
Burkhard, H.D. (1993) "Theoretische Grundlagen (in) der Verteilten künstlichen Intelligenz" in Müller, J. (1993).
Burkhard, H.D./Rammert, W. (1999) "Integration kooperationsfähiger Agenten in komplexen Organisationen - Möglichkeiten und Grenzen der Gestaltung hybrider offener Systeme", unveröffentlichtes Manuskript, TU/HU-Berlin.
Callon, M./Latour, B. (1992) "Don’t throw the baby out with the Bath school! A reply to Colllins and Yearley" in Pickering (1992).
Castelfranchi, C. (1990) "Social power - A missed point in Multi-agent, DAI and HCI" in Demazeau/Müller (ed.) (1990).
Castelfranchi, C. (1995) "Commitments: From individual intentions to groups as organizations" in Lesser (1995) "ICMAS-95", AAAI-Press. SF, Menlo Park.
Castelfranchi C./Conte, R: (1992) "Emergent functionality among intelligent systems: Cooperation within and without minds" in "AI and Society", Vol. 6.
Castelfranchi, C./Conte, R. (1996) "DAI and social sciences: Critical issues" in O’Hare/Jennings (ed.) (1996.)
Castelfranchi, C./Werner, E. (ed.) (1994) "Artificial social systems", Springer, Berlin u.a.
Castelfranchi, C./Müller, J.P. (ed.) (1995) "From the reaction to cognition - 5th workshop of modelling an agent in a multi-agent world", Springer, Berlin u.a..
Cavedon, L./Rao, A./Wobcke, W.(ed.) (1997)"Intelligent agent systems", Springer, Berlin u.a.
Churchland, B./Churchland, P. (1990) "Could a machine think - Classical AI is unlikely to yield conscious machines- Systems that mimic the brain might" in "Scientific American", Vol. 262/No.1.
Clarke, J. u.a. (ed.) (1989) "Anthony Giddens - Consensus and controversy", MacMillan, Houndsmills/ London.
Cohen, I. (1989) "Structuration theory and social order: Five issues in brief" in Clarke u.a. (ed.) (1989).
Cohen, P.R./Levesque, H.J. (1990) "Intention is choice with commitment" in "Artificial Intelligence", Vol. 42.
Collins, H.M. (1990) "Artificial experts - Social knowledge and intelligent machines"; MIT Press, Cambridge.
Collins, H.M. (1994) "Scene from afar" in "Social studies of science", Vol. 24.
Collins, H.M. (1995) "Science studies and machine intelligence" in Jasanoff u.a. (ed.) (1995).
Collins, H.M. (1996) "Embedded or embodied? A review of Hubert Dreyfus’ ‘What computers still can’t do" in "Artificial Intelligence", Vol. 80.
Collins, H.M/Yearley, S. (1992) "Epistemological chicken" in Pickering (ed.)(1992)
Conte, R./Castelfranchi, C. (1995a) "Cognitive and social action", UCL Press, London.
Conte R./Castelfranchi, C. (1995b) "Norms as mental objects. From normative beliefs to normative goals" in Castelfranchi/Müller (ed.) (1995).
Conte R./Castelfranchi, C. (1996) "Distributed Artificial Intelligence and social science: Critical issues" in O’Hare/Jennings (ed.) (1996).

Dautenhahn (ed.) (1997) "Socially intelligent agents - papers form AAI Fall Symposium“, Menlo Park, Cal..
Davis, R./Smith, R.G. (1983) "Negotiation as a metaphor for distributed problem solving" in "Artificial Intelligence", Vol .20/1.
Demazeau, Y./Müller, J.P. (1990) "Decentralized artificial intelligence", Cambridge U.Press, Cambridge.
Dennett, D. (1984) "Cognitive wheels: The frame problem of AI" in Hookway (1984) „Minds, Machines, and Evolution“.
Dennett, D. (1987) "The intentional stance", MIT Press, Cambridge.

Doran, J.E./Franklin, S. /Jennings, N.R. /Norman, T.J. (1997) "On cooperation in multi-agent systems" in "The knowledge engineering review", Vol. 12/3, 1997, zitiert nach http://www.elec.qmw.ac.uk/dai/pubs/tomas.html.

Dreyfus, H.L./Dreyfus, S.E. (1987) "Künstliche Intelligenz. Von den Grenzen der Denkmaschine und dem Wert der Intuition", Rowohlt, Reinbek.

Dreyfus, H.L./Dreyfus, S.E. (1988) "Making a mind versus modeling the brain: AI back at a branch-point" in "Artificial Intelligence", Vol. 117/1.

Dreyfus, H.L. (1996) "Response to my critics" in "Artificial Intelligence", Vol. 80/1.

Ellrich, L./Funken, C. (1998) "Problemfelder der Emergenz - Vorüberlegungen zur informatorischen Anschlußfähigkeit soziologischer Begriffe" in Malsch (Hg.) (1998).

Ferber, J. (1996) "Reactive AI: Principles and applications" in O'Hare/Jennings (ed.) (1996).

Florian, M. (1998) "Die Agentengesellschaft als sozialer Raum - Vorschläge zur Modellierung von 'Gesellschaft' in VKI und Soziologie au sder Sicht des Habitus-Feld-Konzeptes von Pierre Bourdieu" in Malsch (Hg.) (1998).

Fox, M.S. (1988) "An organizational view of distributed systems" in Bond/Gasser (ed.) (1988).

Franklin, S./Graesser, A. (1997) "Is it an agent, or just a program?: A Taxonomy for autonomous agents" in Wooldridge/Müller/Jennings (ed.) (1997).

Friedman, B. (ed.) (1997) "Human values and the design of computer technology", CSLI Publications, Cambridge.

Fuchs, P. (1991): "Kommunikation mit Computern? Zur Korrektur einer Fragestellung" in "Sociologica Internationalis", Vol. 29/1.

Fuller, S. (1994) "Making agency count - A brief foray into the foundation of social theory" in "American Behavioural Scientist", Vol. 37/6.

Galison, P./Stump,D. (ed.) (1996) "The disunity of science: Boundaries, contexts and power", Stanford U.Press, Stanford.

Gasser, L. (1991) "Social conceptions of knowledge and action: DAI foundations and open system semantics" in "Artificial Intelligence", Vol. 47.

Gasser, L. (1992) "Boundaries, Identity, and Aggregation: Plurality issues in multiagent systems" in Werner/Demazeau (ed.) (1992).

Gasser, L. u.a. (1987) "MACE: A flexible testbed for Distributed AI Research" in Huhns (ed.) (1987).

Giddens, A. (1977) "Notes on structuration" in derselbe "Studies in social and political theory", Hutchington, London, 1977.

Giddens, A. (1984) "Interpretative Soziologie - Eine kritische Einführung", Campus, Frankfurt.

Giddens, A. (1992) "Die Konstitution der Gesellschaft", Campus, Frankfurt

Giddens, A. (1995) "Konsequenzen der Moderne", Suhrkamp, Frankfurt.

Gilbert, N. (1995) "Emergence in social simulation" in Gilbert/Conte (1995).

Gilbert, N./Conte, R. (ed.) (1995) "Artificial societies - The computer simulation of social life", UCL Press, London.

Gold, P. (1998) „Philosophische Aspekte Künstlicher Intelligenz" in Gold/Engel (Hg.) (1998)

Gold, P./Engel, A. (Hg.) (1998) "Der Mensch in der Perspektive der Kognitionswissenschaften", Suhrkamp, Frankfurt.

Green et.al (1997) "Software review" (Report der 'Intelligent Agent group' am Trinity College Dublin), zitiert nach http://www.cs.tcd.ie/research_groups/aig/iag/pubreview.zip.

Haddahi, A. /Sundermeyer, K. (1996) "Belief-Desire-Intention agent architectures" in O' Hare/Jennings (ed.) (1996)

Hayes-Roth, B. (1988) "A blackboard architecture for control" in Bond/Gasser (ed.) (1988).

Heintz, B. (1993) "Die Herrschaft der Regel", Campus, Frankfurt.

Heintz, B. (1995) "Die Innenwelt der Mathematik", unveröff.Habil, Fu-Berlin.
Hewitt, C.E. (1977) "Viewing control structures as pattern of message passing" in "Artificial intelligence", Vol. 8.
Hewitt, C.E. (1991) "Open information systems semantics for distributed artificial intelligence" in "Artificial Intelligence", Vol. 47.
Hintikka, J. (1962) "Knowledge and belief", Cornell U.Press, New York.
Hirschauer, S. (1994) "Towards a methodology of investigations into the strangeness of one's own culture: A response to Collins" in "Social studies of science", Vol. 24.
Huhns, M.N. (ed.) (1987) "Distributed artificial intelligence, Vol. 1", Pitman, London.
Huhns, M.N./Gasser, L. (ed.) (1989) "Distributed artificial intelligence, Vol. 2", Pitman, London.
Huhns, M.N. Singh, M.P. (1998) "Agent jurisprudence" in "ICEE - Internet computing", Vol. X, March/April 1998.
Jasanoff u.a. (ed.) (1995). "Handbook of science and technology studies", Sage, Thousand Oaks.
Jennings, N. (1996) "Coordination techniques for DAI" in O'Hare/Jennings (ed.) (1996).
Joas, H. (1992a) "Die Kreativität des Handelns", Suhrkamp, Frankfurt.
Joas, H. (1992b) "Eine soziologische Transformation der Praxisphilosophie - Giddens' Theorie der Strukturierung" in Giddens (1992).

Kirn, S. (1995) "Verteilte künstliche Intelligenz - ein (noch unvollständiges Lexikon)", zitiert nach http://www.wirtschaft.tu-ilmenau.de/~stk/veroeff/veroeff1.html (Stand Aug.1998).
Kirn, S. (1996) "Organization intelligence and DAI" in O'Hare/Jennings (ed.) (1996).
Knorr-Cetina, K. (1981) "The microsociological challgence of macrosociology" in dieselbe/Cicourel "Advances in social theory and methodology - toward an integration of micro- and macro-sociologies" , Routledge and Kegan, Boston.
Knorr-Cetina, K. (1990) "Zur Doppelproduktion sozialer Realität - Der konstruktivistische Ansatz und seine Konsquenzen" in "Österreichische Zeitschrift für Soziologie", 15Jg./3.
Knorr-Cetina, K. (1992) "The couch, the cathedral, and the laboratory: On the relationship between experiment and laboratory in science" in Pickering (ed.) (1992).
Knorr-Cetina, K. (1996) "The care of the self and blind variation: An ethnography of the empirical in two sciences" in Galison, P. /Stump (ed.) (1996).
Kornfeld W.A./Hewitt, C.E.(1988) "The scientific community metaphor" in Bond/Gasser (ed.) (1998).
Krämer, S. (1994a) "Einleitung" in Krämer (Hg.) (1994)
Krämer, S. (Hg.) (1994b) "Geist – Gehirn - Künstliche Intelligenz", de Gruyter, Berlin.
Krogh, C. (1996) "The rights of agents" in Wooldridge/Müller/Tambe (ed.) (1996).

Lanier, J./Maes, P. (1996) "Intelligent agents - Stupid humans" in ‚Hotwired', zitiert nach http://www. Hotwired.com/braintennis/96/29/index0a.html.
Latour, B. (1987) "Science in Action", Harvard U. Press, Harvard.
Latour, B. (1988) "Mixing humans and Nonhumans together. The sociology of a Door-closer" in "Social problems" Vol. 35/3.
Latour, B.(1991) "Technology is society made durable" in Law (ed.) (1991).
Latour, B. (1998) "Über technische Vermittlung - Philosophie, Soziologie, Genealogie", unveröffentlichtes Manuskript, FU-Berlin.
Law (ed.) (1991) "A sociology of monsters: Essays on power, technology and domination", Routledge and Kegan, London.
Lesser, V.R. /Corkhill, D.D. (1983) "The distributive vehicle monitoring testbed: A tool for investigating distributed problem solving networks" in "The AI Magazine"/Vol. 4..
Leydesdorff, L. (1993) "'Structure'/'Action' contingencies and the model of parallel distributed processing" in "Journal of the theory of social behaviour", Vol. 23/1.
Luhmann, N. (1984) "Soziale Systeme", Suhrkamp, Frankfurt.
Luhmann, N. (1990) "Wissenschaft der Gesellschaft" , Suhrkamp, Frankfurt.

Maes, P. (ed.) (1990) "Designing autonomous agents", MIT Press, Cambridge.
Maes, P. (1994) " Agents that reduce work and information overload" in "Communications of the ACM", Vol. 37/7.
Maes, P. (1995) "Artificial life meet entertainment: Life like autonomous agents" in "Communications of the ACM", Vol. 38/11.
Malsch, T. (1997) "Die Provokation der ‘Artificial societies‘- Warum sich die Soziologie mit den Sozialmetaphern der VKI beschäftigen sollte" in "Zeitschrift für Soziologie", Vol. 26/1.
Malsch, T. (Hg.) (1998) "Sozionik - Soziologische Ansichten über künstliche Sozialität", Edition Sigma, Berlin.
Malsch T./Müller H.J. (Hg.) (1998) "Was VKI und Soziologie voneinander lernen können?", Hamburg-Harburg.
Malsch, T./Schulz-Schaffer, I. (1997) "Generalized media of interaction and inter-agent-coordination" in Dautenhahn (ed.) (1997).
Minsky, M. (1990) "Mentopolis", Klett-Cotta, Stuttgart.
Moulin, B./Chaib-Draa, B. (1996) "An overview of distributed artificial intelligence" in O’Hare/Jennings (ed.) (1996).
Müller, H.J. (Hg.) (1993) "Verteilte künstliche Intelligenz - Methoden und Anwendungen", BI Wissenschaftsverlag Mannheim u.a..
Müller, J.P. (1996) "The design of agents - a layered approach", Springer, Berlin u.a.
Müller, J.P. (1997) "Control structures for autonomous and interacting agents: A survey" in Cavedon, L./Rao, A./Wobcke, W. (ed.) (1997).
Münch (Hg.) (1992). "Kognitionswissenschaft, Suhrkamp, Frankfurt.

Negroponte, N. (1997) "Agents: From direct manipulation to delegation" in Bradshaw (ed.) (1997)
Newell, A./Simon, H. (1992) "Computerwissenschaft als empirische Forschung: Symbole und Lösungssuche" in Münch (Hg.) (1992).
Nwana, H.S. u.a. (1997) "An introduction to agent technology" in Nwana, H.S. (ed.) (1997)
Nwana, H.S. (ed.) (1997) "Software Agents and soft computing - towards enhancing machine intelligence - Lecture notes in AI 1198", Springer, Berlin u.a.

O’Hare, G.M./Jennings, N.R. (ed.) (1996) "Foundations of distributed artificial intelligence", Wiley and Sons, New York.

Perram, J.W./Müller, M.P. (ed.) (1996) "Distributed software - Agents and applications", Springer, Berlin u.a..
Pickering, A. (ed.) (1992) "Science as practice and culture", Chicago U.Press, Chicago.
Pickering, A. (1993) "The mangle of praxis: Agency and emergence in the sociology of science" in "American Journal of sociology", Vol. 99/3.
Pickering, A. (1995) „The mangle of praxis: Time, agency, and science”, Chicago U.Press, Chicago.

Rammert, W. (1998a) "Giddens und die Gesellschaft der Heinzelmännchen - Zur Soziologie technischer Agenten und Systeme Verteilter Künstlicher Intelligenz" in Malsch (Hg.) (1998).
Rammert, W.(1998b) "From single knowledge machines to multi-agent systems, or: How social theory and software construction may benefit from one another" in Malsch/Müller (Hg.) (1998).
Rao, A.(1996) „AgentSpeak L: BDI-Agents speak in out in a logical computable language“ in Perram/Müller (ed.) (1996).
Rosenschein, J.S./Genesereth, M.R. (1988) "Deals among rational agents" in Bond/Gasser (ed.) (1988).

Saam, N. (1996) "Computergestützte Theoriekonstruktion in den Sozialwissenschaften - Konzeptbasierte Simulation eines theoretischen Modells am Beispiel militärischer Staatsstreiche in Thailand", Society for Computer Simulation International, San Diego u.a..

Schachtner, C. (1993) "Die Geistmaschine", Suhrkamp, Frankfurt.

Schulz-Schaeffer, I. (1998) "Akteure, Aktanten und Agenten - Konstruktive und rekonstruktive Bemühungen um die Handlungsfähigkeit der Technik" in Malsch (Hg.) (1998).

Schulz-Schaeffer, I./Malsch, T.(1998) "Das Koordinationsproblem künstlicher Agenten aus der Perspektive der Theorie künstlich generalisierter Interaktionsmedien" in Malsch (Hg.) (1998).

Schulz-Schaeffer, I. (1999) „Die Doppelstruktur von Technik – Zur sozialen Bedeutung gegenständlicher Technik“, Dissertation Universität Bielefeld.

Searle, J. (1986) "Geist, Hirn und Wissenschaft“, Suhrkamp, Frankfurt, 1986.

Searle, J. (1990) "Is the brain’s mind a computer program? - No. A program merely manipulates symbols, whereas a brain attaches meaning to them" in "Scientific American", Vol. 262/No. 1.

Searle, J. (1997) "The mystery of consciousness", NY Review of books, New York.

Shneiderman, B. (1997) "Direct manipulation versus agents: Paths to predictable, controllable, and comprehensible interfaces" in Bradshaw (ed.) (1997).

Shneiderman, B. /Maes, P. (1997) "Direct manipulation versus interface agents" in "Interactions", Vol.4/No.6.

Shoham, Y. (1997) „An overview of Agent-Oriented Programming“ in Bradshaw (ed.) (1997).

Smith, R.G./Davis, R. (1988) "Frameworks for cooperation in distributed problem solving" in Bond/Gasser (ed.) (1988).

Sproull, L. et.al (1997) "When the interface is a face" in Friedman (ed.)(1997)

Star, S.L. (1989) "The structure of ill-structured solutions: Boundary objects and heterogeneous distributed problem solving" in Huhns/Gasser (ed.)(1989).

Steels, L. (1990) "Cooperation between distributed agents through selforganization" in Demazeau/Müller (ed.) (1990).

Strübing, J. (1998) "Multi-Agenten Systeme als ‘Going concern‘ - Zur Zusammenarbeit von Informatik und Interaktionismus auf dem Gebiet der Verteilten Künstlichen Intelligenz" in Malsch (Hg.)(1998).

Suchman, L. (1987) "Plans and situated action – The problem of human-machine communication“, Cambridge U. Press, Cambridge/Mass.

Suchman, L. (1997) "Do categories have politics? The language/action perspective reconsidered" in Friedman (ed.) (1997).

Turing, A. (1950) "Computing machinery and intelligence" in "Mind", Vol. LIX/No.2.

Turkle, S. (1997) "Leben im Netz. Identität im Zeitalter des Internet", Rowohlt, Reinbek.

von Martial, F. (1993) "Planen im Multiagentensystemen" in Müller (Hg.)(1993).

Werner, E. (1996): "What ants cannot do" in Perram/Müller (ed.) (1996).

Werner, E. /Demazeau, Y. (ed.) (1992) "Dezentralized AI", North Holland, Amersterdam.

Winograd, T./Flores, F. (1986) "Erkenntnis, Maschinen, Verstehen. Zur Neugestaltung von Computersystemen", Rotbuch, Berlin.

Wolfe, A. (1991) "Mind, self and society: AI and the sociology of mind" in "American Journal of Sociology, Vol. 96/5

Wolfe, A. (1993) "The human difference - Animals, computers, and the necessity of social science", U. of California Press, Berkely.

Wooldridge, M. J. (1997) "Agent as a Rohrschach Test: A response to Franklin and Grasser" in Wooldridge/Müller/Jennings (ed.) (1997).

Wooldridge, M. J./Jennings, N.R. (1995a): "Intelligent agents: theory and practice" in "The knowledge engineering review", Vol. 10/2.

Wooldridge M. J./Jennings, N.R. (1995b): "Agent theories, architectures, and languages: A survey" in Wooldridge/Jennings (ed.) (1995c).

Wooldridge, M. J./Jennings, N.R. (ed.) (1995c) "Intelligent Agents 1 - Agent theories, architectures and languages" , Springer, Berlin, u.a..

Wooldridge, M. J./Jennings, N.R. (1996) "Towards a theory of cooperative problem solving" in Perram, Müller, J.P. (ed.) (1996).

Wooldridge, M. J /Müller, J.P./Jennings, N.R. (ed.) (1997) "Intelligent Agents 3 - Agent theories, architectures and languages", Springer, Berlin u.a..

Wooldridge, M.J./Müller, J.P./Tambe, M. (ed.) (1996) "Intelligent agents 2 – Agent theories, architectures and languages", Springer, Berlin u.a..