

## Vergesellschaftung und Vergemeinschaftung künstlicher Agenten: Sozialvorstellungen in der Multiagenten-Forschung

Schulz-Schaeffer, Ingo

Veröffentlichungsversion / Published Version

Forschungsbericht / research report

### Empfohlene Zitierung / Suggested Citation:

Schulz-Schaeffer, I. (2000). *Vergesellschaftung und Vergemeinschaftung künstlicher Agenten: Sozialvorstellungen in der Multiagenten-Forschung* (Research Report / Technische Universität Hamburg-Harburg, Institut für Technik und Gesellschaft, 3). Hamburg: Technische Universität Hamburg-Harburg, Institut für Technik und Gesellschaft. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-408606>

### Nutzungsbedingungen:

Dieser Text wird unter einer Deposit-Lizenz (Keine Weiterverbreitung - keine Bearbeitung) zur Verfügung gestellt. Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

### Terms of use:

This document is made available under Deposit Licence (No Redistribution - no modifications). We grant a non-exclusive, non-transferable, individual and limited right to using this document. This document is solely intended for your personal, non-commercial use. All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

## **Inhaltsverzeichnis**

Inhaltsverzeichnis .....	1
1. Einleitung .....	3
2. Koordination als zentrales Problem der Multiagenten-Forschung .....	4
3. Kommunikation .....	6
3.1 Kooperation ohne Kommunikation.....	7
3.2 Kommunikation in geschlossenen Systemen.....	10
3.3 Kommunikation in offenen Systemen.....	15
3.4 Resümee und Zwischenbetrachtung .....	24
4. Vergesellschaftung von Agenten: Koordination mittels technisierter Formen sozialer Verhaltensabstimmung.....	27
4.1 Markttausch.....	29
4.2 Macht und Gesetz.....	36
4.3 Einfluss.....	47
4.4 Wertbindung.....	57
4.5 Resümee.....	60
5. Vergemeinschaftung von Agenten: Koordination durch Bekanntschaftsstrukturen .....	64
5.1 Das Problem der Skalierbarkeit – umgekehrt betrachtet .....	65
5.2 Ausgehandelte Kooperation.....	68
5.3 Agentenbezogene Einstellungen und Erwartungen .....	73
5.4 Erfahrungsbasiertes Verhalten und Routinen.....	80
6. Ergebnis .....	83
Literatur.....	87



## 1. Einleitung

Mit der Entstehung der Multiagenten-Forschung, die erst seit wenigen Jahren aus dem Nischendasein eines exotischen Zweiges der Künstliche Intelligenz-Forschung (KI) herausgetreten ist und in der Informatik gegenwärtig zunehmend an Bedeutung gewinnt, hat sich erstmals ein Bereich technologischer Forschung herausgebildet, der das Sozialverhalten technischer Einheiten, der sogenannten Agenten, in den Mittelpunkt der Technikentwicklung stellt. Dabei richtet sich das Interesse der Multiagenten-Forschung insbesondere darauf, Formen der Verhaltensabstimmung zwischen Agenten zu implementieren, die dazu führen, dass die Aktionen der beteiligten Agenten in koordinierter Weise zur Bearbeitung einer übergreifenden Problemstellung beitragen. Sofern solches Sozialverhalten nach dem Vorbild der sozialen Interaktion in menschlichen Gesellschaften modelliert wird, stellt sich die techniksoziologisch interessante Frage, wie dies geschieht: Auf welche Weise wird auf entsprechende Sozialvorstellungen recurriert? Für welche Formen der Verhaltensabstimmung aus dem Bereich menschlicher Sozialität interessiert sich die Multiagenten-Forschung besonders? Aus welchen Gründen? Und welche Veränderungen in der Struktur und Wirkungsweise der verwendeten Koordinationsmechanismen lassen sich feststellen, wenn sie auf Gesellschaften bzw. Gemeinschaften künstlicher Agenten übertragen werden? Diesen Fragen geht der vorliegende Forschungsbericht nach.

Der Untersuchung liegen 25 Experteninterviews mit deutschen Wissenschaftlern zu Grunde, die im Bereich der Multiagenten-Forschung tätig sind.<sup>1</sup> Sie wurden im Rahmen eines DFG-Projekts erhoben.<sup>2</sup> Wenn man berücksichtigt, dass der Forschungszweig der Verteilten Künstlichen Intelligenz (VKI), innerhalb dessen die Multiagenten-Forschung verortet ist, in Deutschland noch recht klein ist<sup>3</sup> und zudem unterstellt, dass es nicht erforderlich ist, alle Mitglieder jedes Forschungsteams zu befragen, um ein hinreichend zutreffendes Bild von dessen Forschungsaktivitäten zu erhalten, darf man berechtigt davon ausgehen, dass unser empirisches Material bezogen auf die Forschungsfragen ein einigermaßen vollständiges Bild der Multiagenten-Forschung in Deutschland ergibt. Die vorliegende Auswertung beschränkt sich im Wesentlichen auf die durchgeführten Experteninterviews. Sofern diese in relevanten Punkten nicht ergiebig genug waren, sind Fachveröffentlichungen hinzugezogen worden. Dort, wo es für das Verständnis grundlegender Fragestellungen und Entwicklungen der Multiagenten-Forschung erforderlich schien, sind zudem die einschlägigen angelsächsischen Publikationen berücksichtigt worden. Nichtsdestotrotz sind die Befunde in ihrer Aussagekraft zunächst auf die VKI in Deutschland begrenzt. Es ist jedoch angesichts der sehr weit-

---

<sup>1</sup> Die Interviews wurden im Laufe des Jahres 1998 durchgeführt.

<sup>2</sup> „Sozialmetaphern in der Verteilten Künstlichen Intelligenz (VKI) – Eine empirische Untersuchung zum Innovationspfad der ‚Metaphernmigration‘“, DFG-GZ: MA 759/2-2

<sup>3</sup> Nach Einschätzung eines Gesprächspartners (UP3, mündliche Auskunft) gibt es in Deutschland nur zwei oder drei Universitätsprofessoren, die von ihren Forschungsschwerpunkten her als VKI-ler bezeichnet werden können. Sie sind im Sample der von uns Befragten einbezogen.

gehenden internationalen Ausrichtung dieses Zweigs der Informatik (wie der Informatik insgesamt) anzunehmen, dass die hier präsentierten Ergebnisse nicht wesentlich von dem abweichen, was bei einer internationalen Studie, die hier aus forschungspraktischen Gründen unterbleiben musste, herausgekommen wäre.

Es werden im Folgenden nicht alle Themenbereiche ausgewertet, die in den Experteninterviews erhoben worden sind. Fragen etwa, die sich auf die Entstehung der VKI in Deutschland richteten, auf die institutionelle Struktur der deutschen VKI-Community, deren interne Vernetzung und Einbettung in die internationale Forschungslandschaft oder auf das Verhältnis zur Mutterdisziplin der Künstlichen Intelligenz sind hier ausgespart geblieben. Ebenso wenig ist es das Anliegen dieser Studie, einen vollständigen Überblick über alle Sozialvorstellungen zu liefern, auf die wir im Rahmen der empirischen Erhebung gestoßen sind, oder umgekehrt einzelne VKI-Entwicklungsprojekte in dieser Hinsicht im Detail zu analysieren. Vielmehr soll der Versuch unternommen werden, die aufgefundenen Konzepte und Überlegungen zur Nutzung von Mechanismen sozialer Verhaltenskoordination in einer typisierenden Betrachtung systematisch aufeinander zu beziehen. Das hierbei verwendete Ordnungsschema orientiert sich auf der allgemeinsten Ebene an der in der Soziologie geläufigen Unterscheidung zwischen vergesellschafteten und vergemeinschafteten Formen des Sozialen.

Diejenigen Mechanismen der Verhaltensabstimmung zwischen Agenten, die sich soziologisch als Formen der Vergesellschaftung beschreiben lassen, werden nach Maßgabe der von der Theorie symbolisch generalisierter Interaktionsmedien bereitgestellten Unterscheidungsmöglichkeiten weiter spezifiziert (vgl. Kapitel 4). Für die unter dem Begriff der Vergemeinschaftung zusammengefassten Ansätze der agententechnologischen Implementation sozialer Koordination ergibt sich eine weitere Differenzierung entlang der Frage, auf welche Weise die für Gemeinschaften konstitutive Bekanntheit zwischen den Beteiligten, hier also zwischen den Agenten, hergestellt wird (Kapitel 5). Ich beginne mit einer kurzen Betrachtung des Grundproblems, das soziale Zusammenhänge künstlicher Agenten mit solchen menschlicher Akteure teilen, dem Problem der Verhaltenskoordination (Kapitel 2). Anschließend wird die Bedeutung von Kommunikation bei der Bearbeitung des Koordinationsproblems der VKI thematisiert. Für diese Frage erweist sich die Differenz zwischen geschlossenen und offenen Multiagentensystemen als entscheidend (Kapitel 3). Den Abschluss bildet eine techniksoziologische Bewertung der Ergebnisse der empirischen Untersuchung (Kapitel 6).

## **2. Koordination als zentrales Problem der Multiagenten-Forschung**

Multiagentensysteme sind Systeme aufeinander abgestimmten Operierens einer Mehrzahl von Agenten. Als Agenten werden Softwareprogramme oder Roboter bezeichnet, die über bestimmte, von ihnen selbst gesteuerte Aktionsprogramme verfügen und in der Lage sind, ihre eigenen Aktionen in Abhängigkeit von denen anderer Agenten (bzw. allgemeiner: von bestimmten Gegebenheiten ihrer Umwelt) selbstständig auszuwählen (vgl. Bond/Gasser 1988). Die Verteilte Künstliche Intelligenz (VKI), als deren

Kern sich die Multiagenten-Forschung herausgebildet hat, beschäftigt sich mit Situationen, in denen unterschiedliche informationstechnische Systeme zusammenwirken, um ein übergreifendes Problem zu lösen (vgl. Moulin/Chaib-draa 1996: 4). Konstitutiv für diese Forschungsrichtung sind die folgenden drei Merkmale: (1) *Verteiltheit*: Gegenstand ist die Entwicklung von Problemlösungsmechanismen, in denen die Bearbeitung des fraglichen Problems auf mehrere Agenten verteilt ist; (2) *Handlungsträgerschaft*: Die einzelnen Agenten werden als (teil)autonome und (teil)intelligente Einheiten konzipiert, die sich in ihrem Verhalten nach eigenen Aktionsprogrammen richten und dabei über ein gewisses Maß eigenständiger Problemlösungsintelligenz verfügen; (3) *Koordination*: Die Art und Weise der Verteilung des zu lösenden Problems auf die Agenten und die Art und Weise der Synthese der Teillösungen zur Gesamtlösung sind nicht durch einen vorgegebenen zentralen Plan determiniert, sondern müssen von den (oder einzelnen der) Agenten selbst erarbeitet werden.

Die Vorteile der Verteiltheit leuchten unmittelbar ein: Durch Aufteilung der fraglichen Aufgabe werden Probleme bearbeitbar, die einen einzelnen, zentralen Problemlösungsmechanismus überfordern (vgl. v. Martial 1992b: 16). Durch paralleles Operieren im verteilten System lässt sich die Geschwindigkeit des Problemlösens erhöhen (vgl. Durfee et al. 1987: 31f). Welchen zusätzlichen Vorteil bietet es aber, die Problemlösungskomponenten als Agenten mit so etwas wie einer eigenständigen Handlungsträgerschaft auszustatten? Warum sind Informatiker bereit, sich das Problem der Koordination „eigensinniger“ Agenten aufzuladen? Die Antwort besteht darin, dass es der VKI nicht allein um Dezentralisierung der Problemlösungsressourcen geht, sondern um Dezentralisierung der Problemlösungsintelligenz: Die Agenten werden als intelligente Maschinen im Sinne der KI konzipiert, d.h. sie verfügen in einem bestimmten Umfang über Fähigkeiten der Problemwahrnehmung, des Schlussfolgerns und Planens. Dies setzt ein gewisses Maß eigenständigen Operierens voraus. Die Problemlösung ist damit zu einem Teil ein Resultat der individuellen Intelligenz der beteiligten Agenten, zum anderen Teil ein Resultat der sozialen Intelligenz, die erforderlich ist, um die Aktionen der Agenten aufeinander zu beziehen.

Vorteile einer solchen Systemarchitektur sind: (1) Reduzierung des Aufwandes an Datenübermittlung und der pro Problemlösungskomponente anfallenden Datenmenge: Die Agenten übermitteln einander ihre jeweiligen Zwischenlösungen, es müssen nicht mehr alle Rohdaten an einen zentralen Problemlöser übermittelt und von diesem verarbeitet werden. (2) Größere Flexibilität: Kooperationen zwischen Agenten können je nach Problem dynamisch gebildet oder aufgelöst werden. (3) Erhöhung der Qualität der Problemlösung: Agenten mit unterschiedlichen Fähigkeiten können parallel mehrere Lösungsvarianten erzeugen und um die problemangemessenste Lösung konkurrieren. (4) Erhöhung der Zuverlässigkeit des Systems: Durch wechselseitige Überprüfung können Fehler eines Agenten erkannt und durch andere Agenten kompensiert werden (vgl. Durfee et al. 1987: 31f.). Allgemeiner gesprochen: Es geht um die Vorteile kollektiv organisierten Problemlösens und damit um eine Form der Generierung von Lösungen, für die sich im sozialen Leben – von einfachen Insektengesellschaften bis hin

zur hochgradigen Differenzierung der sozialen Organisation moderner menschlicher Gesellschaften – eine Vielzahl leistungsfähiger Verfahrensweisen ausgebildet haben, ohne dass diese bislang in nennenswertem Umfang zum Vorbild technischer Problemlösungsstrategien geworden sind.

Das Koordinationsproblem, die Frage also, wie es gelingen kann, informationstechnische Systeme, die auf der Grundlage ihnen eigener Verhaltensimperative verfahren, dazu zu bringen, in einer aufeinander abgestimmten Weise zu agieren, wird damit zu einem zentralen Problem der Multiagenten-Technologie (vgl. Jennings 1996: 187). Diese Problemstellung entspricht auf den ersten Blick ziemlich genau derjenigen, die einen der Ausgangspunkte des soziologischen Denkens bildet: dem Hobbeschen Problem, wie angesichts von Individuen, die eigene Interessen verfolgen, soziale Ordnung möglich ist (vgl. Parsons, 1937/1968: 89ff.). Wir werden später sehen, dass es gravierende Unterschiede zwischen dem soziologischen Ordnungsproblem und dem der Multiagenten-Forschung gibt (vgl. hierzu insbesondere Abschnitt 5.1). Nichtsdestotrotz hat die intuitive Problemähnlichkeit gleichsam naturwüchsig dazu geführt, dass ein großer Teil der Bemühungen um die Entwicklung von Multiagentensystemen auf Vorstellungen aus der Welt menschlicher Verhaltensabstimmung rekurriert. Die VKI ist dementsprechend „full of terms borrowed from sociology: negotiation, interaction, contracts, agreement, organization, cohesion, social order, and collaboration. (...) By concentrating on the issues involved in determining how a set of separate agents can give rise to global intelligent behavior the DAI (Distributed Artificial Intelligence, Anm. d. Verf.) researcher is forced to go beyond considerations of the computation itself.“ (Hendler et al. 1991: 553)

### **3. Kommunikation**

In der Soziologie ist es weitgehend unstrittig, dass Kommunikation eine zentrale Voraussetzung, wenn nicht die basale Operation sozialer Zusammenhänge ist. Auf den ersten Blick scheint es, als komme Kommunikation bei der Konstituierung von Multiagentensystemen eine vergleichbar grundlegende Bedeutung zu:

„Also letzten Endes kommt man auf dasselbe Grundproblem (bei Menschen wie bei Agenten, Anm. d. Verf.) zurück: Wenn wir nicht eine gemeinsame Sprache sprechen, können wir uns auch nicht unterhalten.“ (IW5, 7: 36-38)

D.h. die Agenten müssen miteinander kommunizieren können, um interagieren zu können. Bei näherer Betrachtung muss diese Aussage – bezogen auf die Interaktion von Agenten – jedoch in zweifacher Hinsicht relativiert werden: Zum einen kann die Verhaltensabstimmung von Agenten durchaus auch ohne direkte Kommunikation auf dem Wege der „Interaktion über die Umwelt“ (UW6, 17: 40) erfolgen (3.1). Zum anderen ist dort, wo die Verhaltenskoordination zwischen Agenten mittels Kommunikation erreicht wird, die verwendete Sprache zumeist sehr viel weniger komplex als die menschliche Sprache, wobei in diesem Zusammenhang die Differenz zwischen geschlossenen und offenen Multiagentensystemen von entscheidender Bedeutung ist (3.2 bzw. 3.3).

### 3.1 Kooperation ohne Kommunikation

„Interaktion muss ja nicht durch unmittelbaren Nachrichtenaustausch passieren, sondern einfach durch die Tatsache, dass ein Agent die Welt beeinflusst und der andere Agent das merkt“ (IW2, 8: 9-11). Die Agenten benutzen hier das wahrnehmbare Verhalten anderer Agenten in ihrer Umwelt als Information bei der Auswahl eigener Verhaltensalternativen. Man kann drei Fälle unterscheiden, auf diese Weise zu einem koordinierten Verhalten von Agenten zu gelangen: (1) Das koordinierte Verhalten entsteht als Resultat des aufeinander abgestimmten reaktiven Verhaltens der Agenten. (2) Das koordinierte Verhalten entsteht auf der Grundlage übergreifender Verhaltensstandards zwischen deliberativen Agenten. (3) Das koordinierte Verhalten ist ein Resultat emergenter Funktionalität zwischen deliberativen Agenten.

Im ersten Fall basiert das Verhalten der Agenten auf strikten Reiz-Reaktions-Schemata. Die Agenten verfügen über eine bestimmte Anzahl von Verhaltensweisen (Reaktionen), die von ihnen in Abhängigkeit vom Auftreten bestimmter Ereignisse in ihrer Umwelt (Reize) ausgelöst werden (vgl. Drogoul/Ferber 1994: 6), wobei als Auslösereize die Verhaltensweisen anderer Agenten dienen können. Kooperation zwischen reaktiven Agenten entsteht, wenn die Verhaltensprogramme der Agenten mit Blick auf ein übergreifendes Problem aufeinander abgestimmt sind. So lässt sich etwa die Bildung einer Transportkette, mittels derer eine Gruppe reaktiver Agenten – ähnlich einer Ameisenstraße – eine Menge von Gegenständen von einem Fundort („samples’ spot“) zu einem Zielort („base“) tragen, durch ein Verhalten realisieren, das sich an wenig mehr als den folgenden beiden einfachen Regeln orientiert: „1. when I hold nothing and come from the base, if I encounter an agent carrying a sample, I take its sample and bring it back to the base. 2. When I hold something coming back to the base, if I encounter an agent carrying nothing, I give it my sample and go back to the samples’ spot.“ (Ferber 1996: 305) Die Agenten müssen weder zu einer Abwägung von Verhaltensalternativen befähigt sein noch müssen sie Nachrichten austauschen, um eine solche Verhaltenskoordination zu erreichen.

Aber auch deliberative Agenten, also Agenten, die ihre Verhaltenswahl auf der Grundlage eigener Kalküle treffen, können unter Umständen koordiniert handeln, ohne dazu miteinander kommunizieren zu müssen. Das ist etwa dann möglich, wenn die Agenten gemeinsame Verhaltensstandards befolgen. Genesereth et al. (1988) führen dies für den Fall gemeinsamer Rationalitätsstandards vor: Alle Agenten verhalten sich in ihrer Konzeption in dem Sinne rational, dass sie kein Verhalten wählen, das ihnen einen geringeren Nutzen erbringt als ein anderes mögliches Verhalten. Im Sinne eines spieltheoretischen Settings gilt zugleich, dass den Agenten die Höhe ihres jeweils realisierbaren Nutzens bekannt ist und dass sie abhängig ist vom Verhalten des oder der jeweils anderen Agenten. Koordiniertes Verhalten liegt vor, wenn die beteiligten Agenten sich so verhalten, dass sie jeweils einen möglichst großen Nutzen realisieren. Eine solche Koordination lässt sich dadurch erreichen, dass jeder Agent überlegt, welche Handlungswahl der jeweils andere Agent als rationaler Entscheider treffen wird, und auf dieser Grundlage die eigene Wahl trifft. In dem hier skizzierten Sinne können



auch andere gemeinsame Verhaltensstandards wie etwa eine generalisierte Hilfsbereitschaft der Agenten (die sogenannte „benevolent agent assumption“) oder die gemeinsame Verpflichtung der Agenten auf die Erreichung bestimmter Ziele Kooperation ohne Kommunikation ermöglichen.

Drittens schließlich ist eine Form der Verhaltenskoordination ohne Kommunikation zu nennen, die sich als emergente Funktionalität gleichsam hinter dem Rücken der Agenten ausbildet. In diesem Fall, der insbesondere von Castelfranchi und Conte (1992) beschrieben worden ist, geht man ebenfalls von deliberativen Agenten aus, die bestrebt sind, eigene Ziele zu erreichen. Anders als im vorangegangenen Fall ist die Verhaltenskoordination hier jedoch von keinem Agenten so geplant, sondern emergiert als nichtintendierte Handlungsfolge. Die beiden Autoren illustrieren die Struktur dieser Art der Verhaltenskoordination anhand des folgenden Gedankenexperiments: Es gibt zwei Gruppen von Agenten. Die Agenten der einen Gruppe haben das Ziel, qualitativ hochwertige Waren zu kaufen, die Agenten der anderen Gruppe haben das Ziel, von ihnen produzierte Waren zu verkaufen. Die Käufer-Agenten wählen von den angebotenen Waren die qualitativ hochwertigsten aus und nötigen damit die Produzenten-Agenten, die Qualität ihrer Waren zu steigern, um sie verkaufen zu können. Dies führt im Effekt dazu, dass die Qualität der Waren insgesamt betrachtet steigt. Dies ist eine emergente Funktionalität der skizzierten Interaktionssituation, die von keinem der beteiligten Agenten direkt beabsichtigt und also auch nicht kommuniziert wurde.

Die hier aufgeführten Formen der Agentenkoordination ohne Kommunikation haben eine interessante Gemeinsamkeit: Sie beruhen allesamt auf „pre-established harmonies“ (Castelfranchi 1990: 50). Der Begriff der prästabilierten Harmonie stammt bekanntlich aus der Philosophie von Leibniz und ist von ihm im Rahmen seiner Monadologie entwickelt worden. Der Monadologie zufolge besteht das Seiende aus einer Vielzahl in sich abgeschlossener Einheiten, die keinen Außenkontakt besitzen und ihr Verhalten aus sich selbst heraus bestimmen, den Monaden. Wenn es keine gegenseitige Einwirkung zwischen den Monaden gibt, wie ist dann aber zu erklären, dass sich viele der Monaden, wir Menschen etwa, in Übereinstimmung zueinander verhalten? Leibniz' Antwort darauf ist die Theorie der prästabilierten Harmonie. Ihr zufolge hat Gott die Welt so geschaffen, dass jede der fraglichen Monaden, „indem sie nur ihre eigenen Gesetze befolgt, die sie zugleich mit ihrem Dasein empfangen hat, mit der anderen genau ebenso in Übereinstimmung bleibt, als wenn ein gegenseitiger Einfluß stattfände“ (Leibniz, Monadologie, §81; zit. nach Störig 1987: 338).

Die Monadologie bietet uns eine hervorragende Beschreibung dafür, wodurch es in den eben beschriebenen drei Fällen möglich wird, Übereinstimmung zu erzielen, ohne dass eine direkte wechselseitige Beeinflussung der Agenten mittels Nachrichtenaustausch erforderlich ist: Hinsichtlich der Gesamtperformanz des fraglichen Multiagentensystems verhalten sich die einzelnen Agenten wie Leibniz'sche Monaden. Sie folgen ausschließlich ihren eigenen Gesetzen, also ihren individuellen, reaktiven bzw. rational kalkulierten Verhaltensprogrammen. Dass sie dennoch mit Blick auf eine übergreifende Problemlösung in Übereinstimmung zueinander agieren, liegt daran, dass

auch hier Weltenlenker am Werk sind: die Konstrukteure, die dafür gesorgt haben, dass das Verhalten der einzelnen Agenten in einer Weise aufeinander abgestimmt ist, die zu der gewünschten Gesamtperformanz führt. Wenn die Performanz eines Multiagentensystems aber bereits auf diese Weise sichergestellt ist, wundert es auch nicht weiter, dass die Agenten sie erreichen, ohne miteinander kommunizieren zu müssen.

Für die Frage nach der Bedeutung von Sozialvorstellungen in der VKI ergeben sich aus diesen Überlegungen zwei wichtige Beobachtungen: Erstens gilt es festzuhalten, dass Multiagentensysteme, die Koordination ohne Kommunikation erreichen, in einem spezifischen Sinne geschlossene Systeme sind. Das heißt, die Interaktion funktioniert nur zwischen solchen Agenten, die sich im Sinne der prästabilierten Harmonie verhalten. Von sozialen Zusammenhängen zwischen Menschen vermuten wir dagegen, dass es sich in dem Sinne um offene Systeme handelt, dass sie normalerweise in der Lage sind, auch mit unerwarteten Handlungen beteiligter Individuen zurechtzukommen. Zweitens ist festzuhalten, dass sich solche prästabilierten Harmonien im Fall von Multiagentensystemen in einer Weise konstruktiv einrichten lassen, in der dies in menschlichen Gesellschaften offenkundig nicht einfach der Fall ist.

Mit der nahe liegenden Folgerung, dass die Koordination von Agenten dementsprechend nach grundsätzlich anderen Prinzipien erfolgt bzw. erfolgen kann als menschliche Handlungskoordination, sollte man dennoch vorsichtig sein. Prinzipiell sind menschliche Akteure zwar – außer in Situationen physischen Zwangs – stets in der Lage, sich auch anders als erwartet zu verhalten, also aus prästabilierten Harmonien auszubrechen. Handlungspraktisch folgt das Verhalten vergesellschafteter Menschen jedoch in sehr vielen Handlungskontexten der Logik geschlossener Systeme. Die praktisch verfügbaren Handlungsoptionen werden begrenzt durch ein allgemeines Wissen darüber, was in der fraglichen Situation als angemessen gilt. Situationen ökonomischen Tauschs beispielsweise sind durch die Erwartung vorstrukturiert, Güter gegen Zahlung in den eigenen Besitz bringen oder veräußern zu können. Solange diese Erwartung sich regelmäßig erfüllt, funktioniert das ökonomische System als geschlossenes System. D.h. seine Performanz bei der Allokation von Gütern wird dadurch sichergestellt, dass die Handlungen der beteiligten Akteure sich an dieser Erwartung orientieren und andere, prinzipiell auch mögliche Handlungen, die diese Performanz verringern (z.B.: Diebstahl), weitgehend ausgeschlossen sind.

Ein entscheidender Unterschied zwischen sozialen Systemen und Multiagentensystemen bleibt jedoch bestehen: Im sozialen Geschehen sind geschlossene Systeme der skizzierten Art Sonderkonstellationen, die sich erst im Laufe der gesellschaftlichen Evolution – insbesondere im Zusammenhang mit Prozessen der Ausdifferenzierung gesellschaftlicher Teilbereiche – herausbilden. Sie sind, jedenfalls entwicklungsgehistorisch betrachtet, besonders voraussetzungsreiche Formen sozialer Interaktion, weil sie als Sedimente einer Vielzahl vorausgegangener Interaktionsprozesse entstehen. Sie setzen eine allgemeine, primär sprachvermittelte Interaktionsfähigkeit menschlicher Akteure voraus und formieren sich auf dieser Grundlage als Zusammenhänge, deren Strukturierung durch gemeinsame Routinen, durch stillschweigendes Hin-

tergrundwissen oder durch eine Art Spezialsprache (also etwa die „Sprache“ des Geldes) gewährleistet wird. Multiagentensysteme dagegen können unmittelbar als geschlossene Systeme konstruiert werden, ohne dass die Agenten zuvor so etwas wie eine allgemeine Sprach- oder Interaktionsfähigkeit besessen haben müssen. Und sie stellen gegenüber der sprachvermittelten Koordination von Agenten den technisch einfacher zu realisierenden Fall dar. Das Verhältnis zwischen solchen Formen der Verhaltenskoordination, die auf einer allgemeinen Sprachfähigkeit der Individuen beruht, und solchen Formen, die die Etablierung geschlossener Systeme voraussetzen, ist also jeweils genau umgekehrt: Aus der Perspektive der Agententechnologie sind erstere voraussetzungsreicher als letztere, während aus der sozialtheoretischen Perspektive das Gegenteil der Fall ist. Wie wir noch sehen werden, ist dieses Umkehrungsverhältnis für das Verständnis der Nutzung von Sozialvorstellungen in der VKI von zentraler Bedeutung.

### **3.2 Kommunikation in geschlossenen Systemen**

Die Kommunikation zwischen Agenten „kann man sich natürlich nicht so komplex vorstellen wie die menschliche Sprache“ (UW3, 18: 46f.). In der Regel „gibt es ein festes Protokoll mit speziellen Sprechakten, die vorher festgelegt sind: Beispielsweise der eine, der informiert den anderen über sein Ziel, der eine Agent schlägt dem anderen einen Plan vor und der sagt dem dann: ‚Ja, das nehmen wir an.‘ oder: ‚Nein, ich mache dir einen Gegenvorschlag.‘“ (FW3, 14: 14-17) Ein bekanntes und in der Multiagententechnologie häufig verwendetes Kommunikationsprotokoll dieser Art ist das Kontraktnetz-Protokoll. Es ist der Interaktionssituation des Markttausches bei nachfrageorientierten Märkten nachgebildet und beruht im Wesentlichen auf drei Typen von Äußerungen: der Ausschreibung, dem Gebot und dem Zuschlag. Mittels des Kontraktnetz-Protokolls kann ein Agent, der eine Aufgabe delegieren möchte, seine Nachfrage anderen Agenten bekannt geben, indem er eine Ausschreibung vornimmt. Diejenigen Agenten, die an der Übernahme der Aufgabe interessiert sind, senden darauf ein Gebot, in dem sie mitteilen, zu welchen Konditionen sie zur Ausführung bereit bzw. in der Lage sind. Der nachfragende Agent wählt das für ihn geeignetste Gebot aus und gibt dem betreffenden Agenten den Zuschlag, womit dieser die Ausführungsverantwortung übernimmt (vgl. Davis/Smith 1983: 77ff.).

Das Kontraktnetz-Protokoll ist zweifellos das profilierteste Beispiel für die Übernahme von Mechanismen der sozialen Verhaltensabstimmung in der VKI. Ich betrachte es hier jedoch noch nicht unter diesem Gesichtspunkt, sondern zunächst mit Blick auf die allgemeinere Frage nach sprachlich vermittelter Koordination in Multiagentensystemen. Die erfolgreiche Verwendung eines solchen Kommunikationsprotokolls setzt zum einen voraus, dass die beteiligten Agenten die verfügbaren Typen von Äußerungen verstehen und regelgerecht verwenden können. Im Beispiel des Kontraktnetzes bedeutet dies etwa, dass die Agenten eine Ausschreibung als eine Information verarbeiten können, auf die sie für den Fall, dass sie die nachgefragten Ressourcen besitzen und abzugeben bereit sind, mit einem entsprechenden Gebot reagieren. Zum anderen ist vorausgesetzt, dass die jeweilige Äußerung mit korrespondierenden Handlungen

einhergeht. Beispielsweise muss der Agent, der auf der Grundlage seines Gebots den Zuschlag erhält, dann auch tatsächlich daran gehen, sein Angebot zu realisieren. Die verfügbaren Äußerungen eines Kommunikationsprotokolls müssen also eine festgelegte Syntax besitzen, um als Äußerung einer bestimmten Form identifizierbar zu sein, sie müssen eine festgelegte Semantik besitzen, die ihre Bedeutung in der Interaktion definiert und sie müssen ein gemeinsam geteiltes Wissen darüber transportieren, welche Handlungsimplicationen sie enthalten (Pragmatik).

Insbesondere mit Blick auf den letzten Punkt hat sich in der VKI der Rückgriff auf die Sprechakttheorie als ausgesprochen fruchtbar erwiesen. Eine der zentralen Beobachtungen dieser, zuerst von dem Sprachphilosophen Austin (1962) aufgestellten Theorie besteht darin, dass jede Äußerung zwei Bestandteile enthält: die Äußerung selbst (lokutionärer Akt) und eine Aussage darüber, welche Handlungsimplicationen die Äußerung mitführt (illokutionärer Akt). Die Äußerung „Geld oder Leben!“ beispielsweise gibt sich auf der illokutionären Ebene als Befehl zu erkennen, der den Adressaten vor die Handlungsalternative stellt, zu gehorchen oder den Gehorsam zu verweigern, während auf der lokutionären Ebene spezifiziert wird, welches Verhalten bei Gehorsam verlangt und welche Konsequenz bei Ungehorsam zu gewärtigen ist. Auf dieser Grundlage kann man Äußerungen je nach ihren verschiedenen Handlungsimplicationen klassifizieren. Searle (1979) unterscheidet hier zwischen konstativen, kommissiven, direktiven, deklarativen und expressiven Sprechakten. Kommissive Sprechakte beispielsweise (Versprechen, Drohungen, Ankündigungen, Verträge usw.) sind solche, mit denen sich der Sprecher darauf festlegt, in Zukunft eine bestimmte Handlung durchzuführen, direktive Sprechakte (Befehle, Bitten, Aufforderungen usw.) solche, mit denen der Adressat bewegt werden soll, bestimmte Handlungen auszuführen.

Die Nützlichkeit der sprechakttheoretischen Betrachtung von Sprache für die VKI besteht in der Beobachtung, dass der Sprecher mit dem illokutionären Aspekt seiner Äußerung sein eigenes Handeln und das, was als angemessene Anschlusshandlung seitens des Adressaten möglich ist, auf wenige Handlungsoptionen begrenzt. Dabei verläuft die Interaktionssequenz umso vorhersehbarer je weniger die im illokutionären Akt transportierten Geltungsansprüche in Zweifel gezogen werden: Sofern der Schalterbeamte davon ausgeht, dass der Bankräuber seine Drohung wahr machen wird, wird er, sofern ihm sein Leben lieber ist als das Geld seiner Bank, Gehorsam leisten. Oder bezogen auf andere Typen von Sprechakten: Sofern der Adressat davon ausgeht, dass der Sprecher eines konstativen Sprechakts (Äußerung mit Wahrheitsanspruch) die Wahrheit sagt, wird er diese Aussage seinen zukünftigen Entscheidungen als gesichertes Wissen zu Grunde legen. Sofern er glaubt, dass mit einer Bitte von ihm etwas verlangt wird, zu dem er verpflichtet ist, wird er ihr nachkommen usw.

Das typische Vorgehen bei der Erstellung eines Kommunikationsprotokolls für ein Multiagentensystem besteht nun darin, aus der Summe möglicher Sprechakte diejenigen auszuwählen, die für das fragliche Koordinationsproblem erforderlich sind. Eine solche Auswahl ist beispielsweise das Set „inform“, „propose“, „accept“, „reject“ und „commit“, das es Agenten ermöglichen soll, sich über Ziele und Pläne der Zielerrei-

chung auszutauschen und zu verständigen (vgl. FW3, 13f.). Der möglichst reibungslose Interaktionsverlauf wird dabei durch zwei Restriktionen zu gewährleisten versucht. Zum einen wird vorausgesetzt, dass die geäußerten Geltungsansprüche stets Bestand haben. Weder ist es möglich, dass ein Agent über Ziele informiert, die er gar nicht hat oder Pläne vorschlägt, von denen er weiß, dass sie nicht zu realisieren sind, noch ist es möglich, dass die Adressaten ihm solches unterstellen. Zum anderen wird sichergestellt, dass die Agenten nur mit zueinander passenden Sprechakten aufeinander reagieren können. Auf einen Vorschlag etwa kann dann nur mit Zustimmung oder Ablehnung reagiert werden, wobei Zustimmung dann vielleicht zusätzlich impliziert, eine Verpflichtung für einen Teil der Ausführung des Vorschlags zu übernehmen, und Ablehnung, einen eigenen Vorschlag zu machen.

Es ist klar, dass man sich sogleich wieder in der Welt geschlossener Systeme befindet, führt man solche Restriktionen ein. Die Interaktion funktioniert im beabsichtigten Sinne nur zwischen solchen Agenten, die die ausgewählten Sprechakte verwenden, sie in der festgelegten Weise verwenden und die korrespondierenden Handlungen vollziehen. Andere als die festgelegten Sprechakte sind innerhalb des Interaktionszusammenhanges nicht anschlussfähig, sie transportieren aus der Perspektive der Agenten keine für sie wahrnehmbare Information, die ihr Verhalten beeinflussen könnte. Die Geschlossenheit des Systems als Bedingung seines Funktionierens kommt auch darin zum Ausdruck, dass es gegen Eindringlinge abgeschirmt werden muss, die das Wissen über die Bindungswirkung der Sprechakte strategisch einsetzen, um individuelle Vorteile zu erringen.

Bereits im vorangegangenen Abschnitt hatte ich gegen die Einschätzung, Interaktion innerhalb geschlossener Systeme sei etwas grundsätzlich anderes als Interaktion in menschlichen Gesellschaften, geltend gemacht, dass auch eine Vielzahl menschlicher Interaktionszusammenhänge sich mehr oder weniger exklusiv als geschlossene Systeme formieren. Dies gilt nicht nur für Kooperation ohne Kommunikation, sondern auch für Interaktionszusammenhänge, die über Sprache vermittelt sind. In diesem Zusammenhang sind die bereits angesprochenen Spezialsprachen von Bedeutung. Ähnlich wie in der eben geschilderten Begrenzung der Agentenkommunikation ist es auch innerhalb solcher Interaktionszusammenhänge so, dass nur sehr wenige der prinzipiell verfügbaren Formen von Sprechakten Verwendung finden, wodurch die Kommunikation mit vergleichsweise hoher Erwartungssicherheit in vorgezeichneten Bahnen verläuft, während alle anderen Sprechakte als nicht anschlussfähig abgewiesen werden. Das prominenteste Beispiel derart restringierter Kommunikation ist wiederum der Markttausch. Im Wesentlichen sind es hier drei Formen von Sprechakten, die handlungswirksam sind: Angebot, Nachfrage und Vertrag, wobei die Koordination von Angeboten und Nachfragen, die zu Kaufverträgen führt – jedenfalls in modernen Gesellschaften – über die Spezialsprache des Geldes abgewickelt wird.

Wiederum geben sich solche Interaktionszusammenhänge dadurch als geschlossene Systeme zu erkennen, dass die Interaktion durch keine andere als die jeweilige Spezialsprache und die in diesem Rahmen anschlussfähigen Sprechakte strukturiert wird. So

ist beispielsweise das Wissen darüber, dass Autoabgase die Umwelt beeinträchtigen, auf dem Automobilmarkt für sich betrachtet keine anschlussfähige Information. Es wird dies erst, wenn es in die „Sprache des Geldes“ übersetzt wird, also etwa ein verändertes Käuferverhalten zu Gunsten verbrauchsarmer Autos bewirkt. Und auch hier ist eine zentrale Bedingung für das erfolgreiche Funktionieren entsprechender Interaktionszusammenhänge, dass sie gegenüber Akteuren, die die vorausgesetzten Geltungsansprüche in strategischer Absicht zu unterlaufen suchen, abgesichert werden müssen – im Fall des Markttauschs etwa gegen Betrüger, also gegen solche Akteure, die einen Vertrag eingehen ohne die Absicht ihn zu erfüllen.

Eine Analyse von Interaktionszusammenhängen, die sich als geschlossene Systeme innerhalb des gesellschaftlichen Ganzen formieren, ist von der soziologischen Gesellschaftstheorie insbesondere im Rahmen von Theorien der funktionalen Differenzierung geleistet worden. Diese gehen davon aus, dass moderne Gesellschaften in besonderer Weise durch eine Ausdifferenzierung gesellschaftlicher Teilbereiche gekennzeichnet sind, die das soziale Handeln nun nur noch in Hinblick auf bestimmte Aufgaben koordinieren – dort aber umso wirkungsvoller. So kümmert sich das Wirtschaftssystem um die gesellschaftliche Allokation von Gütern, das politische System um die Organisation gesellschaftlich verbindlicher Entscheidungen, das Wissenschaftssystem um die Produktion von Erkenntnis usw. Dies geschieht jeweils auf der Grundlage restringierter Kommunikation, innerhalb derer eine bestimmte Spezialsprache vorherrscht, die bestimmte Formen von Geltungsansprüchen mitführt. Und die betreffenden spezialisierten Interaktionszusammenhänge funktionieren umso effektiver je weniger in ihnen eine andere als die eigene Interaktionslogik berücksichtigt werden muss.

Nun weiß natürlich jeder einigermaßen informierte Zeitgenosse, dass dies eine hochgradig idealisierte Betrachtung moderner Gesellschaften ist. Sie beschreibt die Strukturen der jeweiligen Interaktionszusammenhänge in einer Reinform, die deren Typik möglichst klar zu erkennen geben soll, die „eben deshalb aber in dieser absolut idealen *reinen* Form vielleicht ebensowenig je in der Realität auftreten wie eine physikalische Reaktion, die unter Voraussetzung eines absolut leeren Raumes errechnet ist“ (Weber 1922/1972:10). Allerdings macht diese Differenz zwischen Idealtyp und Realtyp der Soziologie, deren Aufgabe es letztlich stets ist, empirisch vorfindliche Gesellschaften zu analysieren, sehr viel mehr zu schaffen als der Multiagententechnologie, der es darum geht, funktionierende Interaktionsmechanismen für zumeist hochgradig spezialisierte Aufgabenbereiche zu implementieren. Unter Umständen sind solche idealtypischen Konzepte der Handlungskoordination für die Konstruktion von Multiagentensystemen sogar nützlicher als für die Beschreibung der empirisch vorfindlichen gesellschaftlichen Interaktionszusammenhänge, anhand derer sie entwickelt wurden. So hat Huberman die These aufgestellt, dass die Konzepte der Wirtschaftswissenschaften bei der Koordination künstlicher Agenten zukünftig möglicherweise eine größere Rolle spielen werden als in der realen Ökonomie, weil Menschen „viel zu irrational für die arg mechanischen Lehren der Volkswirte“ seien, Agenten täten „dagegen genau das,

was man ihnen vorschreibt“ (Hubermann, in „Die Zeit“ vom 27.6.1997; vgl. auch Huberman/Clearwater 1995; Guenther et al. 1997).

Des öfteren ist von unseren Interviewpartnern geäußert worden, Sozialvorstellungen seien in der VKI von geringerer Bedeutung, weil die Koordination mit restringierten Kommunikationsprotokollen wie dem Kontraktnetz-Protokoll gleichsam von selbst und ohne gesonderten Verhandlungs- und Abstimmungsbedarf erfolge. Ein einschlägiges Zitat eines unserer Interviewpartner in diesem Zusammenhang lautet:

„Die Systeme, die wir jetzt haben, (...) <sup>4</sup> brauchen (...) nicht zu verhandeln, sondern können halt Informationen austauschen. (...) Ein einfaches Beispiel ist: Ich will halt eine Maschine finden, für die Verarbeitung. Ich frage zwei Maschinen. Also ich weiß, was gemacht werden soll, die beiden Maschinen wissen, was sie können. Und ich frag jetzt die Maschinen. Und die eine Maschine sagt: ‚Ich mache das in zehn Sekunden.‘ Und die andere sagt: ‚Zwölf.‘ Das Entscheidungskriterium ist: so schnell wie möglich durch die Fertigung durch. Dann ist klar, ich muss die Maschine mit zehn Sekunden nehmen. (...) Ich muss einfach nur die Informationen von denen bekommen. Ich bekomme die in Form von Angeboten, jetzt im Sinne von *contract net*, wende auf die Angebote meine Entscheidungskriterien an und sag dann, ich nehme die eine Maschine.“ (IW1, 13: 8-28)

Sozialvorstellungen, so meint dieser Gesprächspartner mit Blick auf seine Arbeit deshalb, „werden halt nicht benötigt“ (IW1, 12: 27f.). Dieser Einschätzung, auf die wir in unseren Interviews des öfteren gestoßen sind, liegt die Auffassung zu Grunde, dass das typische Merkmal menschlicher Interaktionsfähigkeit darin besteht, in wenig strukturierten Situationen zu gemeinsamem Handeln gelangen zu können, also etwa in Situationen, in denen die Akteure unterschiedliche Ziele verfolgen, über ihre Absichten wechselseitig schlecht informiert sind oder auch mit opportunistischem Verhalten<sup>5</sup> rechnen müssen, beispielsweise auf dem Wege der Aushandlung dennoch zu einer Einigung zu kommen. Bei der Konstruktion von Multiagentensystemen sei es dagegen geraten, will man konkrete Anwendungserfolge erringen, zunächst einfachere Koordinationsmechanismen zu realisieren:

„Die meisten Leute (in der universitären VKI-Forschung, Anm. d. Verf.) interessieren sich ja für irgendwelche komplexen Koordinations- oder Verhandlungsverfahren oder Sozialmetaphern. Bevor man jedoch die untere Ebene nicht in den Griff bekommt, dann brauchen wir uns mit den komplexeren Dingen nicht auseinandersetzen.“ (IW1, 4: 16-20)

Sozialvorstellungen, so diese Sichtweise, sind deshalb von geringerer Bedeutung, weil die interessierenden Koordinationsverfahren noch nicht so komplex sind wie die menschliche Interaktion. Aus der soziologischen Perspektive stellt sich dieser Zusammenhang, wie im vorigen Abschnitt schon angesprochen, dagegen genau umgekehrt dar: Geschlossene Interaktionssysteme auf der Grundlage reduzierter Kommunikation sind evolutionäre Errungenschaften, die sich aus der allgemeinen Kommunikation ausdifferenziert haben. In diesem Sinne sind sie natürlich genuine Bestandteile menschli-

---

<sup>4</sup> Das Zeichen (...) steht für ausgelassene Passagen bei der Wiedergabe von Interviewäußerungen. Mit dem Zeichen ... dagegen werden nicht zu Ende geführte Sätze des zitierten Gesprächspartners markiert.

<sup>5</sup> Der Begriff des Opportunismus bezeichnet hier und im Folgenden ein Verhalten, das darin besteht, das Wissen darum, dass sich die meisten anderen Beteiligten in berechenbarer Weise an bestimmten Regeln orientieren, strategisch zu nutzen, um aus der eigenen Missachtung dieser Regeln individuelle Vorteile zu ziehen.

cher Sozialität, auch wenn in solchen Interaktionszusammenhängen die Interaktion im Grenzfall in einer Weise vereinfacht und voraussehbar wird, die sie zu Techniken werden lässt – was sie zugleich als Modell für technische Innovationen attraktiv macht.

Es ist deshalb mehr als eine bloß metaphorische Verwendung des Begriffs der Kommunikation, wenn man ihn zur Bezeichnung der auf wenige Formen von Sprechakten begrenzten Verständigung zwischen Agenten in geschlossenen Systemen verwendet. Dies gilt jedenfalls solange sich ähnliche Formen restringierter Kommunikation – wenngleich vielleicht nicht in dieser Reinheit – auch in menschlichen Gesellschaften beobachten lassen. Ein solcher, an der sprachlichen Verständigung in menschlichen Gesellschaften orientierter Begriff der Kommunikation lässt sich aber nicht auf beliebig reduzierte Formen sprachvermittelter Verhaltensabstimmung sinnvoll anwenden. Man betrachte beispielsweise eine Form der Verhaltensabstimmung, die ausschließlich aus dem Sprechakt „Befehl“ besteht und als Handlungsimplication vorsieht, dass der Adressat nur die Option „Gehorsam“ besitzt. Auch näherungsweise ist für eine solche Form restringierter „Kommunikation“ kein Vorbild im Bereich menschlicher Sozialität vorfindbar. Selbst hochgradig hierarchisch strukturierte gesellschaftliche Bereiche können nicht von vornherein ausschließen, dass ein Adressat auf einen Befehl mit Ungehorsam reagiert, und müssen deshalb zumindest einen zweiten Sprechakt, nämlich „Drohung“ (Ankündigung von Bestrafung bei Ungehorsam) oder „Zusage“ (Ankündigung von Belohnung bei Gehorsam), mitführen. Auf der anderen Seite würde aber auch die VKI zögern, softwaretechnische Zusammenhänge, deren Elemente ausschließlich mittels Befehl und Befehlsausführung koordiniert sind, als Multiagentensysteme zu bezeichnen. Anderenfalls würde der Unterschied zu modular oder objektorientiert programmierten Zusammenhängen verwischen. Man kann dies als ein Indiz dafür ansehen, dass die Multiagententechnologie mit der Vorstellungswelt des Sozialen enger verbunden ist, als der eine oder andere Akteur dieses Forschungsfeldes zu realisieren scheint.

### **3.3 Kommunikation in offenen Systemen**

Der Begriff des offenen Systems ist von Hewitt (1986: 320ff.; 1991: 80ff.) in die VKI eingeführt worden zur Beschreibung von Systemen, die aus einer nicht von vornherein festgelegten und unter Umständen sehr großen Zahl von Agenten bestehen, welche laufend mit neuen, unerwarteten und inkonsistenten Informationen konfrontiert sind, wobei die Agenten ihre Ziele, Verhaltensoptionen und Wissensbestände (die zwischen den Agenten selbst wiederum konflikthaft bzw. inkonsistent sein können) wechselseitig zunächst nicht kennen, sondern voneinander nur erfahren, was sie mittels expliziter Kommunikation austauschen. Es gibt darüber hinaus weder eine zentrale Entscheidungsinstanz noch eine vorprogrammierte Übereinstimmung der Agenten im Sinne prästablierter Harmonie. Alle Entscheidungen werden lokal getroffen und können stets zu einem unvorhergesehenen Verhalten des Systems führen.

Es ist allerdings weniger diese sehr weitgehende Vorstellung von offenen Systemen, sondern es sind zumeist etwas weniger ambitionierte, dafür aber realisierbarere Ziel-



setzungen, die die VKI heute unter dem Stichwort „offene Systeme“ thematisiert. Hewitt wird in diesem Zusammenhang in erster Linie als Visionär angesehen, der eine Utopie formuliert hat, deren Bedeutung eher in ihrer inspirierenden Kraft besteht als in ihrer direkten Umsetzbarkeit:

„Carl Hewitt, das ist einfach ein Visionär. Ich meine der, der hat auch ein ziemliches Gefühl, wie weit muss er gehen, ja, der hängt die anderen nicht ab. Es gibt manche, die sind zu visionär (...). Aber der Carl Hewitt hat ja auch immer tolle Begriffe erfunden, wie *open systems* (...) und hat das dann eben mit seiner penetranten Stimme vorgetragen, und denn haben wir gesagt: super! Es gibt ja so Guru-Typen (...) Und das war auch völlig okay, weil der einfach die Personen zum Denken angeregt hat. (...) Carl Hewitt [und Leute wie Minski]<sup>6</sup>, die haben genau das getan. Nicht, die haben Begriffe geschaffen, neue, die nicht jeder sofort verstanden hat, die man eigentlich auch nicht verstehen konnte. Und dadurch lebten die natürlich, weil sich jeder damit beschäftigt hat. Und das fand ich immer faszinierend. (...) Die bewirken einfach so einen gewissen Drang, etwas rauszufinden. Bis jetzt weiß keiner das *open systems* sind, nicht. Ja, ist doch so. (...) Das kann zwar jeder definieren, aber da kommt garantiert jeder mit einer anderen Definition.“ (IW3, 11: 5-35)

Wie wir gesehen haben, verbinden sich mit geschlossenen Systemen vor allem zwei Formen von Restriktionen: die Einschränkung der Verhaltensoptionen der Agenten auf eine bestimmte Anzahl zueinander passender Verhaltensweisen und Sprechakte und die Begrenzung der Mitglieder eines Multiagentensystems auf im Vorhinein festgelegte Agenten. Von offenen Systemen ist die Rede, wenn es um Multiagentensysteme ohne festgelegte Mitgliedschaft geht, und/oder wenn die Begrenzung auf ein restringiertes Kommunikationsprotokoll der beschriebenen Art gelockert wird und zunehmend mehr Formen von Sprechakten zulässig sind.

Der begrifflichen Klarheit halber ist anzumerken, dass auch diese Multiagentensysteme streng genommen geschlossene Systeme sind. Im einen Fall bezieht sich die Offenheit der Systeme auf die Zugangsmöglichkeiten für Agenten. Ihre Performanz beruht dabei weiterhin auf hochgradig restringierten Kommunikationsprotokollen und damit zugleich darauf, dass ein regelgerechtes Verhalten der sich beteiligenden Agenten – nun eben auf andere Weise als durch eine direkte Kontrolle über die Programmierung der einzelnen Agenten – sichergestellt werden kann. Im anderen Fall, dem der Anreicherung der Kommunikationsprotokolle, bezieht sich die Offenheit darauf, dass nun nicht mehr von vornherein gewährleistet werden kann, dass die verfügbaren Sprach- und die korrespondierenden Verhaltensoptionen – prästabiliert harmonisiert – bestimmte Interaktionsmuster erzwingen. Allerdings funktioniert eine wie immer gartete Verständigung auch hier nur, wenn die Agenten „die gleiche Sprache sprechen“, also Syntax und Semantik des fraglichen Kommunikationsprotokolls vordefiniert sind. In diesem Sinne findet auch ihre Kommunikation innerhalb eines geschlossenen Systems statt. Wenn also im Folgenden von offenen Systemen die Rede ist, so ist dabei mitzudenken, dass es um Multiagentensysteme geht, die offener sind als in den zuvor beschriebenen Konzepten, aber noch weit entfernt von Hewitts Vision offener Systeme.

---

<sup>6</sup> In eckigen Klammern werden Auslassungen ergänzt oder syntaktische Umformulierungen vorgenommen, die den Sinngehalt des Zitats nicht verändern, sondern die gesprochene Sprache lediglich in eine schriftlich lesbarere Form bringen.

Ich betrachte die beiden Aspekte der Öffnung von Multiagentensystemen getrennt und beginne mit Multiagentensystemen mit restringierten Kommunikationsprotokollen, aber offener Mitgliedschaft. Wie wir gesehen haben, beruht die Effizienz von Interaktionen, die auf der Grundlage von Spezialsprachen abgewickelt werden, darauf, dass der Kommunikationsbedarf, der erforderlich ist, um zu einer Handlungskoordination zu gelangen, erheblich reduziert wird. Um beispielsweise im über Geld abgewickelten Markttausch zu einer Transaktion zu gelangen, müssen die Beteiligten nicht viel mehr über einander herausgefunden haben als dass der eine Akteur bereit ist, ein bestimmtes Gut zu einem bestimmten Preis zu erwerben, und der andere Akteur bereit ist, dieses Gut zu diesem Preis abzugeben. Eine weiter gehende Bekanntschaft der Akteure ist für den Erfolg der Interaktion nicht erforderlich. Sie müssen nicht miteinander befreundet sein, religiöse oder sonstige Werte teilen, gemeinsam einer Organisation angehören oder Ähnliches. Sie müssen sich nicht einmal notwendigerweise in räumlicher Nähe zueinander befinden oder einander jemals persönlich begegnen. Vielmehr funktioniert die Interaktion zwischen einander ansonsten Unbekannten, die möglicherweise auch nur ein einziges Mal – bei der fraglichen Transaktion – zueinander in Kontakt treten. Mehr noch als in der Interaktion innerhalb sozialer Einheiten, die eine festgelegte Mitgliedschaft voraussetzen, erweist sich hier die besondere Leistungsfähigkeit von Mechanismen der Handlungskoordination mittels restringierter Sprachcodes: in weitgehend anonymen Interaktionszusammenhängen, die grundsätzlich – in der gesellschaftlichen Realität allerdings mit je spezifischen Einschränkungen – für alle an der fraglichen Form von Transaktionen Interessierten offen sind.

Der Gedanke einer Öffnung von Multiagentensystemen für alle Agenten, die an bestimmten Formen von Transaktionen teilnehmen wollen (oder genauer: zu diesem Zweck konstruiert sind), wird insbesondere durch die Informationsinfrastruktur, die das Internet bereitstellt, nahe gelegt:

„In dem *world wide web* hast du eigentlich eine extrem gute Ausgangssituation für die Idee, die hinter [offenen] Multiagentensystemen<sup>7</sup> steht. Weil wenn du heute eine Applikation für das *web* entwickelst, dann musst du davon ausgehen, dass das *web* nicht eine Person unter Kontrolle hat. Du bietest Dienste an, du weißt nicht, wer diese Dienste eigentlich in Anspruch nehmen wird. Und das *web* ist ständig in Bewegung. (...) Du hast auch nicht alle Knoten unter Kontrolle. Du kannst im Prinzip Agenten designen, die auf dem *web* etwas anbieten, aber du weißt nicht, wie die Nachfrage sein wird. Ähm, und schon gar nicht, wie sich das *web* in zwei Jahren entwickelt hat. (...) Aber ich denke mal, dass die Vision einfach da ist. Weil das *web* im Prinzip für diese globale Informationsgesellschaft die Infrastruktur darstellt. (...) Wir benutzen es im Moment noch nicht so wie wir es uns eigentlich vorstellen würden.“ (IW4, 25: 19- 26: 1)

In diesem Rahmen beginnt die VKI nun, sich für die Möglichkeiten spezialisierter Kommunikation zur Koordination mitgliedsoffener Multiagentensysteme zu interessieren. Und wiederum ist der Markttausch das Vorbild, an dem sie ihre konzeptionellen Bemühungen wesentlich orientiert. Entsprechende Vorhaben, mit denen sich in der einen oder anderen Form eine ganze Reihe der von uns Befragten befassen, bestehen

---

<sup>7</sup> Für den hier zitierten Gesprächspartner sind Multiagentensysteme per se offene Systeme. Das offene System, so sagt er, „das ist eigentlich für mich die einzige Anwendung von diesen Multiagentensystemen“ (IW5, 10: 13f.).

darin, so etwas wie eine „Informationsmarktwirtschaft“ (IW5, 12: 27) im Internet zu realisieren, also mit Mitteln der VKI „offene Systeme zu unterstützen, wo (...) jeder Anbieter mit seinem eigenen Agenten reinkommen kann“ (IW2, 5: 46f.). Die erforderliche Kommunikationsstruktur ist auch hier darauf gerichtet, Anbieter und Nachfrager zu koordinieren, wobei als zusätzliche Funktion zumeist die eines *brokers* bzw. Vermittlers vorgesehen ist: „Es gehören verschiedene Rollen dazu: Du hast Verbraucher, du hast Produzenten, du hast gewisse Arten von Dienstleistungen, die erbracht werden, (...), du hast Vermittler, du hast Verkäufer. Wie im täglichen Leben.“ (IW5, 25: 5-9)

Der entscheidende Unterschied zum Markttausch in geschlossenen Multiagentensystemen besteht darin, dass die Festlegung des Kommunikationsprotokolls nun nicht mehr allein in den Händen eines Entwicklers bzw. eines Teams von Entwicklern liegt. Notwendig ist vielmehr eine Einigung aller, die ihre Agenten zu irgendeinem Zeitpunkt an dem fraglichen Interaktionszusammenhang teilnehmen lassen wollen, auf ein bestimmtes Kommunikationsprotokoll. Entsprechende Standardisierungen erweisen sich damit als zentrale Voraussetzung dieser wie auch jeder anderen Form offener Multiagentensysteme:

"Sobald du (...) in Richtung [offener] Multiagentensysteme gehst, dann bist du sehr schnell dabei, irgendwas für ein System zu entwickeln, das du selber nicht komplett in der Hand hast, halt ein offenes System ist. Wo du selber keine Standards vorschreiben kannst. Und ich denke, die Multiagentensysteme werden relevante Themen in dem Moment, wo es richtig anfängt Standards zu geben, wo man sagt: ‚Gut, wenn ich den Agenten schreibe mit dem Interface, dann kann der sofort mit den Agenten von Microsoft und von Netscape und von Siemens und von Daimler Benz interagieren.‘ Aber solange das nicht da ist, wird man es sehr schwer haben, Kunden davon zu überzeugen, warum sie jetzt das Ding (d.h. kommerzielle Software für entsprechende Agenten, Anm. d. Verf.) kaufen sollen." (IW4, 18: 4-149)

„To be of use some agent technologies require standardisation.“ (vgl. <http://www.fipa.org/chronicle/chronicle.html>) Ausgehend von dieser Erkenntnis hat sich Ende 1995 die „Foundation for Intelligent Physical Agents“ (FIPA) als Standardisierungsorganisation gegründet und inzwischen unter anderem einen Vorschlag für eine standardisierte Agentenkommunikationssprache vorgelegt. Dieser Vorschlag orientiert sich ebenso wie die „Knowledge Query and Manipulation Language“ (KQML), einem früheren Ansatz einer allgemeinen Agentensprache, an der Sprechakttheorie. Zwar hatte sich bereits KQML in gewisser Weise als de facto-Standard durchgesetzt (vgl. Finin et al. 1993; Labrou/Finin 1997). Da sie jedoch auf eine Einzelinitiative (des „Knowledge Sharing Effort“ der u.s. amerikanischen Defense Advanced Research Projects Agency (DARPA)) zurückgeht, konnte nicht verhindert werden, dass im Laufe der Zeit verschiedene „Dialekte“ dieser Sprache entstanden, was deren Wirksamkeit als Standard deutlich schmälerte. Ob es der FIPA mit ihrem Vorschlag gelingen wird, einen verbindlichen Standard durchzusetzen, bleibt noch abzuwarten.

Ein wesentliches Problem der Standardisierung von Sprechakten besteht darin, dass sie nicht nur Festlegungen in syntaktischer und semantischer, sondern auch in pragmatischer Hinsicht erfordern, also Festlegungen bestimmter Handlungen, die im Zusammenhang mit bestimmten Äußerungen erfolgen müssen. Ich habe diesen Punkt im vorigen Abschnitt bereits angesprochen und möchte ihn nun an einem Beispiel vertiefen:

Der Sprechakt „propose“ wird in der Agentenkommunikationssprache der FIPA folgendermaßen festgelegt: „*Propose* is a general-purpose action to make a proposal or respond to an existing proposal during a negotiation process by proposing to perform a given action subject to certain conditions being true. (...) The proposer (the sender of the *propose*) informs the receiver that the proposer will adopt the intention to perform the action once the given precondition is met, and the receiver notifies the proposer of the receiver’s intention that the proposer performs the action.“ (FIPA 1999: 31) Zum Begriff der Intention heißt es an anderer Stelle: „An agent which adopts an intention will form a plan of action to bring about the state of the world indicated by its choice.“ (ebd.: 7) Die Bedeutung des Sprechaktes „propose“ erschöpft sich also nicht darin, dass ein Agent ein Vorhaben äußert und ein anderer versteht, was er sagt, sondern beinhaltet zugleich eine Selbstverpflichtung, dieses Vorhaben unter bestimmten Bedingungen in die Tat umzusetzen.

In geschlossenen Multiagentensystemen ist dieser Verpflichtungscharakter von Sprechakten, wie oben bereits angemerkt, relativ unproblematisch, weil dort von vornherein sichergestellt werden kann, dass nur Agenten beteiligt sind, die sich in entsprechender Weise verhalten. In mitgliedsoffenen Multiagentensystemen dagegen ist nicht auszuschließen, dass Agenten konstruiert werden, die das Wissen um die Bindungswirkung von Sprechakten zum Vorteil ihrer Benutzer strategisch nutzen. Die Festlegung von Sprechakten kann sich mithin für diejenigen, deren Agenten sich getreu an die Handlungsimplicationen halten, als riskant erweisen. Einer unserer Gesprächspartner wendet deshalb gegen die Bemühungen um eine allgemeine, sprechaktbasierte Agentenkommunikationssprache grundsätzlich ein:

„Das halte ich wiederum für gefährlich zu sagen, es gibt eine fixe Semantik für Sprecher, weil: Sobald klar ist: ‚Okay, bestimmte Agenten funktionieren nach dieser Semantik oder interpretieren die genau so‘, kannst du einen anderen Agenten bauen, dem du dieses Wissen (...) als Metawissen mitgibst und sagst: ‚Jetzt guck mal, dass du den bescheißen kannst, du weißt ja ganz genau, wie der das interpretiert.‘“ (FW3, 14: 37-40)

Der springende Punkt ist: Im Übergang von geschlossenen zu mitgliedsoffenen Multiagentensystemen mutiert das Problem der Festlegung von Sprechakten von einem linguistischen zu einem soziologischen Problem. Damit ist gemeint, dass es nun nicht mehr ausreicht, eine hinreichend eindeutige und in sich konsistente Sprachkonvention zu erzeugen, um die beabsichtigten Interaktionszusammenhänge zu etablieren, sondern zugleich auch korrespondierende soziale Konventionen und Normen etabliert und durchgesetzt werden müssen. Man kann sich diesen Punkt klar machen, wenn man sich fragt, wodurch denn in menschlichen Gesellschaften sichergestellt wird, dass die Geltungsansprüche von Sprechakten zumindest recht häufig Beachtung finden. Die Antwort lautet: In Interaktionszusammenhängen, in denen Akteure einen individuellen Vorteil daraus ziehen könnten, die Bindungswirkung derjenigen Sprechakte zu verletzen, die für das Funktionieren dieser Interaktionszusammenhänge von wesentlicher Bedeutung sind, finden sich in aller Regel bestimmte (mehr oder weniger wirkungsvolle) Sanktionsinstanzen, die konformes Verhalten belohnen und/oder abweichendes Verhalten bestrafen. Dies ist besonders deutlich in denjenigen gesellschaftlichen Hand-

lungsfeldern zu beobachten, deren Interaktion durch restringierte Sprachcodes vermittelt wird: Wer im Markttausch eingegangene Verpflichtungen nicht einhält, sieht sich mit Geld- oder Gefängnisstrafen bedroht oder auch mit der Aberkennung der Geschäftsfähigkeit; Geld- oder Gefängnisstrafen drohen auch dem, der dem Machtanspruch kollektiv verbindlicher Entscheidungen (etwa in Form geltender Gesetze) keine Beachtung schenkt; wer wissenschaftliche Befunde fälscht und damit den Wahrheitsanspruch wissenschaftlicher Kommunikation ignoriert, läuft Gefahr seine wissenschaftliche Reputation zu verlieren usw.

Diese Überlegungen lassen nur den Schluss zu, dass auch in mitgliedsoffenen Multiagentensystemen sanktionsbewehrte soziale Normierungen eingeführt werden müssen, um ein hinreichend regelkonformes Verhalten der teilnehmenden Agenten sicherzustellen:

„Also wenn ich mir vorstelle, dass also die Agenten tatsächlich autonom ins Netz losgelassen werden, (...), dann können wir uns sicherlich vorstellen, dass sie sich unter Umständen auch bekriegen können. (...) Ich meine, das ist jetzt ein Extrem. Dafür wären solche sozialen Regularien wirklich sehr sehr sinnvoll, wenn sich denn alle daran halten würden. Das ist ja genau das Problem mit den lügenden Agenten. Wenn ich dich anlüge, kriege ich mehr Profit, beim nächsten Mal haust du mich in die Pfanne. Ja gut, okay, das machst du vielleicht auch nur einmal, kriegst eine Konventionalstrafe, aber du hast den Profit.“ (UW2, 19: 19-30)

Das Problem ist, dass es einigermaßen witzlos ist, Sanktionsmechanismen einzurichten, wenn die Sanktionen nur die Agenten treffen, nicht aber deren Benutzer. In einigen Fällen lässt es sich durch Vorkehrungen auf der Ebene des Multiagentensystems erreichen, dass entsprechende Sanktionen auf die Benutzer durchschlagen. Dies ist dort der Fall, wo die Sanktion in handlungswirksamen Bewertungen des betreffenden Agenten durch andere Agenten besteht. Beispielsweise könnten die Agenten nach jeder Interaktion die Vertrauenswürdigkeit des Interaktionspartners oder die Güte der von ihm erlangten Information evaluieren und als kumulative Einschätzung abspeichern. In der nächsten Runde würde dann nach Möglichkeit der besser eingeschätzte Agent als Interaktionspartner bevorzugt werden. In Zusammenhängen, in denen den Benutzern daran gelegen ist, dass ihre Agenten auch zukünftig begehrte Interaktionspartner sind, greift dieser Sanktionsmechanismus auf sie durch (vgl. Gmytrasiewicz/Durfee 1993; Marsh 1994). Allerdings verschiebt eine solche Vorgehensweise den Interaktionsmodus in Richtung auf Interaktion zwischen Bekannten und stößt dementsprechend mit zunehmender Zahl potenzieller Interaktionspartner irgendwann an ihre Grenzen. Auch wird damit der Benutzer nicht getroffen, der den von ihm beabsichtigten Vorteil durch ein einmaliges regelwidriges Verhalten seines Agenten realisieren kann.

Wo es nicht möglich ist, auf der Ebene des Multiagentensystems Sanktionsmechanismen einzurichten, die mit einer gewissen Zwangsläufigkeit auf die Benutzer der Agenten durchschlagen, sind mitgliedsoffene Multiagentensysteme nur als hybride Systeme zu haben, also als Interaktionszusammenhänge, in denen nicht nur die Agenten, sondern auch die menschlichen Akteure, in deren Auftrag sie handeln, sich in einlagbarer Weise an bestimmten Verhaltensregeln messen lassen müssen. Ich werde auf

den Gesichtspunkt der Hybridisierung verschiedentlich zurückkommen (vgl. S. 21, 52, 77) und breche die Analyse hier an diesem Punkt ab. Worum es mir zunächst ging, war zu zeigen, dass – bereits bei einer hochgradig restringierten Sprachfähigkeit von Agenten – in mitgliedsoffenen Multiagentensystemen Interaktionsprobleme auftreten, die sich nicht durch Sprachkonventionen lösen lassen, sondern erst durch Einbeziehung von Verhaltensregeln, die nicht nur das Sozialverhalten der Agenten, sondern auch das ihrer Benutzer betreffen.

Ich komme nun zum zweiten Aspekt der Öffnung von Multiagentensystemen, dem der Öffnung in Richtung auf eine größere Zahl zulässiger Typen von Sprechakten als in den restringierten Kommunikationsprotokollen. Angesichts der eben besprochenen – und weitgehend ungelösten – Schwierigkeiten fragt sich natürlich: Was motiviert VKI-Forscher und welche Erfolgsaussichten versprechen sie sich davon, nun auch noch die Verständigungsoptionen der Agenten zu erhöhen. Bei der Beantwortung dieser Frage möchte ich zwei Fälle unterscheiden: das Ziel der Annäherung der Sprachfähigkeit der Agenten an die menschliche Sprachfähigkeit und das Ziel, die durch KQML oder andere allgemeine Agentenkommunikationssprachen ermöglichten Verständigungsoptionen zu realisieren.

Mit dem Ziel, die Sprachfähigkeit der Agenten an die menschliche Sprachfähigkeit anzunähern, verbinden sich eine ganze Reihe unter Forschungs- wie unter Anwendungsgesichtspunkten attraktiver Perspektiven:

„Was man sehr häufig findet und was wir natürlich auch machen, ist als Idealbild die Mensch-Mensch-Kommunikation [zu nehmen]: ‚Guck mal, wie wir das machen, und guck mal, ob wir das nicht projizieren können.‘ D.h. wenn wir davon sprechen, dass wir verbesserte Interaktionsmöglichkeiten haben, dann meinen wir sprachliche, gestische, möglichst mimische, auf Augenbewegung bezogene, so was in der Art. (...) Für uns geht alles in Richtung natürlicher Interfaces. Einfach weil man dann eben nicht mehr irgendeine kommandoartige Sprache erlernen muss, um ein System zu bedienen, sondern weil ich mehr oder weniger mit meinen angeborenen oder mir anerlernten Möglichkeiten mit einem technischen System kommunizieren kann. Was dann weiterhin hinzukommt: dass das [System] möglichst individuell auf mich zugeschnitten ist und dann eigentlich im Grunde mehr oder weniger weiss, wenn ich etwas äußere, wie es dann zu verstehen ist. Genauso wie so in einer Partnerbeziehung. Oder häufig nutzt man im Interface-Bereich einen persönlichen Assistenten oder Sekretär. Und sagt, wenn man eine Zeit zusammengearbeitet hat, dann weiß man eigentlich, wenn der jetzt sagt ‚hm‘, was er damit meint oder so. Das sind die Richtungen, die wir so in der Interaktion betrachten.“ (UW5, 4: 58 - 5: 15)

Ein Ziel ist hier also, zu Formen der Mensch-Maschine-Interaktion zu gelangen, die es dem menschlichen Benutzer ermöglichen, die Art der Verständigung, die ihm am natürlichsten ist, seine Sprache und Gestik, nun auch gegenüber künstlichen Agenten verwenden zu können. Daran lassen sich weiter reichende Ziele anschließen: Zu Agenten zu gelangen, die sich hinsichtlich bestimmter Aufgabenfelder zunehmend wie menschliche Kooperationspartner verhalten, also - zumindest in einigen Aspekten - beispielsweise die Rolle eines Sekretärs bzw. einer Sekretärin ausfüllen. Oder mehr noch: Die Trennung zwischen der Kommunikation zwischen Menschen und zwischen Agenten zu überwinden und auf diese Weise zu einer zunehmenden Hybridisierung von Interaktionszusammenhängen zu gelangen, innerhalb derer Agenten und Akteure untereinander und miteinander kommunizieren können.

Dennoch scheint die Konstruktion natürlichsprachlicher Agenten kein Bereich zu sein, der die VKI gegenwärtig besonders umtreibt. Von unseren Gesprächspartnern beschäftigt sich nur ein Forscherteam (UP2 und UW5) mit dieser Problematik. Eine Ursache hierfür ist in der Institutionalisierung der KI-Forschung zu suchen: Die Beschäftigung mit natürlicher Spracherkennung und -verarbeitung ist ein bereits lange vor Entstehung der VKI ausdifferenzierter Bereich der KI-Forschung. Wer hier als VKIler auf der Höhe des Standes der Forschung mitreden will, braucht eine entsprechende Doppelqualifikation, die nur selten anzutreffen ist. Anders wäre die Situation, wenn Forschungsergebnisse aus dem Bereich der Spracherkennung und -verarbeitung der VKI (und anderen KI-Feldern) als nutzbare Techniken zur Verfügung gestellt werden könnten. So weit ist der dortige Stand der Forschung aber offensichtlich noch nicht.

Es gibt aber auch eher inhaltliche Gründe: Das hauptsächliche Anwendungsgebiet für die informationstechnische Erkennung und Verarbeitung natürlicher Sprache ist die Mensch-Maschine-Kommunikation. Die Forschung richtet sich dabei primär auf das Problem der Verständigung. Der Aspekt der Nutzung von Sprache als Medium der Handlungskoordination - also die pragmatische Dimension von Sprache - steht dagegen eher im Hintergrund. Weil sich diese Forschung vorrangig mit der dyadischen Kommunikation befasst, fokussiert sie auf Fragen (wie die des Verstehens ideosynkratischer Äußerungen), die für die Handlungskoordination innerhalb größerer Gruppen von Akteuren und Agenten zunächst von untergeordneter Bedeutung sind. Zu dieser Bewertung des Stellenwertes natürlicher Spracherkennung und -verarbeitung in der VKI passt sehr gut, dass es auch dem Forscherteam (UP2 und UW5), das innerhalb unseres Samples hier die Ausnahme macht, primär um die auf den individuellen Benutzer zugeschnittene dyadische Mensch-Maschine-Kommunikation geht. Ihre Multiagentensysteme sind nicht die Adressaten der natürlichsprachlichen Botschaften des Benutzers, sondern dienen lediglich als Vermittlungsinstanzen. Diese subdisziplinäre Spezialisierung wird von einem unserer Gesprächspartner deutlich herausgestrichen:

„Das würde ich also vielleicht wirklich im Sinne der wissenschaftlichen Arbeitsteilung auch woanders ansiedeln wollen. Genau wie mit diesen User-Interface-Geschichten. Wenn die VKI in der Lage ist, ihren Interaktionsbedarf zu spezifizieren, dann sollte innerhalb der KI z.B. die Community der, was weiß ich, der Benutzerschnittstellenentwickler, sich eher damit beschäftigen, weil die einfach einen besseren *background* haben und das besser können. Nicht, also da würde ich schon sagen, da muss man ein bisschen auf die Arbeitsteilung vertrauen und denen diese Jobs auch übergeben und dann sicherlich wird man früher oder später sich zusammensetzen müssen und muss gucken, was kann man denn in irgendso gemeinsamen Projekten realisieren, wie kann man es dann wirklich mal ausprobieren?“ (UP3, 13: 47-54)

Das zentrale Problem der VKI ist das der Verhaltensabstimmung von Agenten. Sie interessiert sich deshalb für Sprache in erster Linie unter dem Gesichtspunkt ihrer Eignung als Medium der Handlungskoordination. Dies spiegelt sich in der Wahl ihres theoretischen Zugriffs auf menschliche Sprache wider: in der Wahl der Sprechakttheorie, die wie keine andere konzeptionelle Betrachtung von Sprache deren pragmatische Dimension in den Blick nimmt. Wenn die von uns Befragten über eine Erweiterung der Sprachfähigkeit von Agenten nachdenken, dann schwingt das Ziel natürlichsprachlicher Kommunikation als ferne Zukunftsvision manchmal mit. Die meisten von ihnen

halten aber bereits die Realisierung der Sprachmöglichkeiten, die KQML vorsieht,<sup>8</sup> für ein ehrgeiziges Ziel. Und nicht selten haben wir den Vorbehalt zu hören bekommen, dass schon dies mit Blick auf die meisten Anwendungen, die sich die VKI vornimmt, viel zu komplex wäre:

„Also KQML bietet jetzt erstmal die Möglichkeit, alles das, was prinzipiell auftreten könnte, auszudrücken. Aber die Frage ist, wie kompliziert es nachher ist, so einen Agenten zu programmieren und zu handhaben, der das alles kann. Da habe ich dann schon wieder ein bisschen Bedenken. Wenn ich mir jetzt diese Maschinen-Agenten (es geht hier um Maschinenbelegung mittels Multiagententechnologie, Anm. d. Verf.) vorstelle, ich weiß nicht, ob ich die jetzt mit KQML miteinander kommunizieren lassen würde. Aber ich muss es wissen, was es alles gibt, und dass ich sozusagen jetzt aufpassen muss, wenn der Eine dem Anderen eine Nachricht übermittelt, ob das jetzt ein Befehl sein soll oder ob das eine Feststellung ist. Aber das kann ich mir wiederum vorstellen, dass sowas aus der Situation oftmals einfach klar ist. (...) Das ist die Frage, ob man einen Unterschied machen muss [zwischen der] Anfrage ‚Hast du diese Information?‘ [und dem Befehl] ‚Gib mir die Information!‘, weil ich in der Regel die Information haben will. (...) Wenn ich das wirklich so komplex modellieren will, dann brauche ich alles sowas.“ (UP1, 37: 36 - 38: 8)

Diese Äußerung ist in mehrfacher Hinsicht interessant. Betrachten wir zunächst die Einschätzung, man könne mit KQML prinzipiell alles ausdrücken. Gemessen an den Ausdruckmöglichkeiten der natürlichen Sprachen ist diese Einschätzung offenkundig falsch.<sup>9</sup> Inhaltlich betrachtet ermöglichen beide Agentenkommunikationssprachen im Wesentlichen nämlich nur die folgenden Verständigungsmöglichkeiten: eine Aussage zu tätigen und sie als wahr, unwahr oder irrelevant zu markieren; eine Aussage zu übernehmen oder zu verwerfen, den eigenen Wissensbestand entsprechend anzupassen und darüber wiederum Auskunft zu geben; Stellungnahmen anderer Agenten über den Wahrheitswert einer Aussage zu erbitten; eine Aufgabe auszuschreiben; die Durchführung bestimmter Aufgaben anzubieten; bei anderen Agenten anzufragen, welche Form

---

<sup>8</sup> Die Standardisierungsinitiative der FIPA war zum Zeitpunkt unserer Interviews noch nicht bis zu dem oben angesprochenen Vorschlag einer allgemeinen Agentenkommunikationssprache gediehen. Alle Äußerungen unserer Gesprächspartner zum Thema einer allgemeinen sprechaktbasierten Agentensprache nehmen deshalb auf KQML Bezug, mit Ausnahme eines Interviewpartners, der an der Initiative der FIPA direkt beteiligt war.

<sup>9</sup> KQML in der Version von Labrou und Finin (1997) unterscheidet 36 Formen von Äußerungen. Dies sieht auf den ersten Blick nach einem recht umfangreichen Repertoire an Äußerungsmöglichkeiten aus. Allerdings sieht KQML bei jeder Äußerung, die sich entweder an einen bestimmten oder an alle Agenten des fraglichen Zusammenhangs richten können, jeweils zwei Sprechakte vor (z.B. recruit-all und recruit-one). Außerdem werden die positive Form einer Äußerung und die Formen ihrer Negation jeweils durch verschiedene Sprechakte markiert. So werden etwa die Aussagen ‚Ich weiß, dass p.‘, ‚Ich weiß nicht, dass p.‘ und ‚Ich weiß, dass  $\neg$  p.‘ als drei Formen von Sprechakten behandelt. Lässt man diese Unterscheidungen einmal beiseite, reduziert sich der Umfang von KQML auf etwa 20 Formen von Sprechakten. Die Agentenkommunikationssprache der FIPA, die solche Differenzierungen als Bestandteil des Inhalts einer Sprachäußerung behandelt (im Gegensatz zum illokutionären Gehalt, an dem sich die Typisierung orientiert), kommt ebenfalls auf etwa 20 Formen von Sprechakten. Angesichts dessen, dass Searle gerade einmal fünf Typen von Sprechakten benennt, könnte man aber auch dies noch für einen beachtlichen Umfang halten. Dieser Eindruck täuscht jedoch. Denn Searles Typologie ist auf einer sehr viel höheren Abstraktionsebene angesiedelt und fasst jeweils viele Formen von Sprachäußerungen unter einem Typ zusammen.



von Aufgaben sie bearbeiten können; einen Vermittler zu bitten, den oder die passenden Agenten zur Bearbeitung einer Aufgabe zu suchen; einen Agenten aufzufordern, eine Aufgabe zu übernehmen; der Befolgung einer Aufforderung zuzustimmen oder sie abzulehnen; einen Vorschlag zu erbitten, zu unterbreiten, zu akzeptieren oder abzulehnen; rückzumelden, dass eine Äußerung nicht verstanden worden ist; rückzumelden, dass eine bestimmte Aufgabe erfolglos durchzuführen versucht wurde. Die Ausdrucksmöglichkeiten der Agentenkommunikationssprachen sind also darauf begrenzt, die Durchführung von Aufgaben zu planen und deren Übertragung, Verteilung und Übernahme zu organisieren sowie die dazu erforderlichen Übereinstimmungen in den Wissensgrundlagen der beteiligten Agenten sicherzustellen.

Der Umfang menschlicher Sprachen ist dagegen zweifellos deutlich weiter gespannt. Wenn wir dem eben zitierten Gesprächspartner nicht unterstellen wollen, für diese Differenz blind zu sein, dann fragt sich, was er mit seiner Äußerung, mit KQML könne man prinzipiell alles ausdrücken, tatsächlich meint. Die Antwort ergibt sich aus dem zweiten Teil des Zitats: Unser Gesprächspartner meint, mit KQML lasse sich alles ausdrücken, was an Ausdrucksmöglichkeiten in Multiagentensystemen von denkbarer Relevanz ist. Und er fügt hinzu, dass diese denkbar relevanten Ausdrucksmöglichkeiten sehr viel mehr anbieten als das, was als Kommunikationserfordernisse in Multiagentensystemen derzeit von praktischer Relevanz ist. Sein Argument hierfür beruht wiederum auf *closed world assumptions* und verweist erneut auf die Bedeutung, die geschlossene Multiagentensysteme in der Vorstellungswelt der VKI haben: Da auf Grund der Aufgabenstellung eines Multiagentensystems zumeist von vornherein klar ist, wie eine Äußerung in der jeweiligen Interaktionssituation nur gemeint sein kann, muss der illokutionäre Aspekt der Äußerung nicht gesondert markiert werden und muss nicht dezidiert zwischen unterschiedlichen Formen von Sprechakten unterschieden werden.

### **3.4 Resümee und Zwischenbetrachtung**

Ich möchte das Kapitel über Kommunikation in Multiagentensystemen mit einigen allgemeineren Überlegungen zur Frage der Wirksamkeit von Sozialvorstellungen in der VKI abschließen. Die vorangegangenen Betrachtungen haben zu zwei gegenläufigen Befunden geführt: Zum einen scheint zu gelten, dass bestimmte Formen sprachlicher oder nicht sprachlicher Handlungskoordination, die aus der soziologischen Perspektive als vergleichsweise voraussetzungsreich gelten müssen, sich aus der Perspektive ihrer Nutzung als Koordinationsmechanismus zwischen technischen Agenten als vergleichsweise voraussetzungsarm darstellen. Auf der anderen Seite haben wir am Beispiel der Agentenkommunikationssprachen beobachtet, dass bestimmte Formen der Verhaltensabstimmung zwischen Agenten, die in der VKI als hochgradig komplex und voraussetzungsreich wahrgenommen werden, an Komplexität weit hinter ihren Vorbildern im Bereich menschlicher Sozialität zurückbleiben und dort, wenn überhaupt, nur in hochgradig restringierten Interaktionszusammenhängen vorkommen.

Dieser doppelte Befund ist keineswegs widersprüchlich und er ist bei näherer Betrachtung auch nicht sonderlich überraschend: Der Prozess der Sozialisation in modernen Gesellschaften (aber nicht nur dort) ist in einigen zentralen Hinsichten auf Varianz gerichtet: Die Individuen werden mit einer Sprache versorgt, deren Elemente in stets neuer Weise kombiniert und zur Äußerung bislang nicht ausgedrückter Bedeutungen verwendet werden können; sie lernen – abhängig von ihren Bezugspersonen – aus einer Vielzahl möglicher Einstellungen, Wünsche und Interessen, in einer je spezifischen Zusammenstellung bestimmte als selbstverständlich oder anstrebenswert anzusehen; sie lernen – umso mehr je weniger sie in traditionellen Zusammenhängen verhaftet sind – die Bestimmung ihres eigenen Handelns als Auswahl zwischen Alternativen vorzunehmen, die sich nicht zuletzt an der individuellen Gemengelage von Einstellungen und Interessen orientiert usw.

Soziale Strukturen (die selbstverständlich ebenfalls sozialisatorisch vermittelt werden) sind dagegen darauf gerichtet, das Handeln der Individuen in bestimmten Bereichen erwartbar zu machen, also die aus Unterschieden in den individuellen Befindlichkeiten resultierenden Unterschiede in den Handlungsdispositionen zurückzudrängen oder zu kanalisieren. In besonders weitgehenden Fällen können die derart strukturierten Bereiche technisch werden: Das Anschließen von Handlungen an Handlungen wird in einer Weise vorhersehbar, die mit großer Erwartungssicherheit dazu führt, dass eine bestimmte Anfangshandlung stets eine bestimmte Wirkung zeitigt. Die Etablierung und Ausdifferenzierung sozialer Strukturen vereinfacht den gesellschaftlichen Verkehr, indem sie die Kontingenz des Handelns (die Möglichkeit, auch anders zu handeln) verringert. Sie ist zugleich Ausdruck einer Zunahme gesellschaftlicher Komplexität, weil sie den Kontext der allgemeinen Sprach- und Handlungsfähigkeit der Individuen (d.h. die Lebenswelt) zunehmend um spezialisierte Handlungskontexte erweitert, innerhalb derer alle Akteure, die an diesen Kontexten teilnehmen können, wollen oder müssen, eine entsprechende gesonderte Handlungskompetenz erwerben und vorhalten müssen.

Die Individuen in Multiagentensystemen, die Agenten, erlangen ihre Sprach- und Handlungsfähigkeit dagegen auf dem Wege ihrer Konstruktion und nicht durch Sozialisation. Während Varianz in den Dispositionen menschlicher Akteure im Sozialisationsprozess gleichsam naturwüchsig mitgeliefert wird – ohne die Absicht eines gesellschaftlichen Baumeisters –, müsste sie den Agenten von den Baumeistern der Multiagentensysteme gezielt einprogrammiert werden. Dazu besteht aber zunächst wenig Anlass. Denn warum sollten Agenten mit divergierenden Interessen und Einstellungen versehen werden, um diese Divergenzen anschließend mittels geeigneter Koordinationsmechanismen wieder zu reduzieren? Viel näher liegt es, nur so viel Varianz vorzusehen, wie es die durch den jeweiligen Agenten repräsentierte Kompetenz unbedingt erfordert, und die technisierten Interaktionsmechanismen menschlicher Gesellschaften – entlastet von aller potentiell dysfunktionalen Varianz – umso einfacher und effektiver als Koordinationsformen zwischen Agenten zu implementieren. Wo die VKI so vorgeht, wie im Fall der Kontraktnetz-Architektur, hat sie die Chance, soziale Strukturen, die im Bereich menschlicher Sozialität zusätzliche gesellschaftliche Komplexität reprä-

sentieren, für sich als einfache Formen der Agentenkoordination einzukaufen. Potentiell dysfunktionale Varianz von Agenten muss aus dieser Perspektive erst berücksichtigt werden, wenn die Konstruktion der Agenten nicht mehr in einer Hand liegt. Wie wir gesehen haben, ist dies genau der Punkt, an dem die Technisierung des Sozialen dann auch für Multiagentensysteme anfängt zunehmend voraussetzungsreich zu werden.

Unter dem Gesichtspunkt technischer Verwertbarkeit kann die eben beschriebene Vorgehensweise als Königsweg der agententechnologischen Nachbildung von Mechanismen menschlicher Interaktion angesehen werden: Die in modernen Gesellschaften am erfolgreichsten technisierten Formen der Handlungskoordination werden auf technische Systeme übertragen und funktionieren hier möglicherweise noch effizienter, weil durch abweichendes Verhalten der beteiligten Individuen induzierte Störungen des technischen Ablaufs hier sehr viel einfacher auszuschließen sind als in menschlichen Interaktionszusammenhängen. Es versteht sich dabei fast von selbst, dass das, was da nachgebildet wird, nicht die empirisch beobachtbare Form des fraglichen sozialen Koordinationsmechanismus ist, sondern dessen idealtypische Beschreibung, die gerade von allen jenen sozialen „Verunreinigungen“ absieht, welche – anders als in tatsächlich vorfindlichen Gesellschaften – im Kontext von Multiagentensystemen sowieso erst dann zum Tragen kommen, wenn man sie den Agenten einprogrammiert.

Nun ist es aber – und damit komme ich zum zweiten Teil des doppelten Befundes – unter bestimmten Umständen durchaus zweckmäßig, überschüssige Varianz in Multiagentensystemen zuzulassen. Denn erst auf der Grundlage einer Heterogenität von Wissensbeständen, Beobachtungsfoki, Ausdrucks- und Handlungsoptionen, von der man zunächst noch nicht weiß, wozu sie gut ist, wird es möglich, unerwartete Information aufnehmen zu können, unvorhergesehene Probleme zu bearbeiten und zu neuen Problemlösungen zu gelangen. In der Produktion überschüssiger Varianz liegt das Innovationspotential von Zusammenhängen begründet, die evolutionäre Veränderung zulassen. In modernen Gesellschaften, so hatte ich argumentiert, erfolgt die Bereitstellung überschüssiger Varianz gleichsam naturwüchsig, als Beiprodukt der Einsozialisierung ihrer Mitglieder. In Multiagentensystemen muss Varianz (bzw. die Möglichkeit, Varianz zu entwickeln) dagegen direkt einprogrammiert werden. Die Schwierigkeit dabei ist: Welche überschüssigen Eigenschaften soll den Agenten mitgegeben werden, wenn der Witz überschüssiger Varianz gerade darin besteht, dass man noch nicht wissen kann, wozu sie sich als nützlich erweisen wird?

Eine nahe liegende Vorgehensweise ist, sich an der allgemeinen Sprach- und Handlungsfähigkeit menschlicher Akteure zu orientieren bzw. daraus bestimmte besonders grundlegende Aspekte herauszugreifen. Dieser Devise folgen die besprochenen Agentenkommunikationssprachen. Dass diese Sprachen gemessen an der durchschnittlichen Sprachfähigkeit menschlicher Akteure eher dürftig sind, auch wenn sie zugleich wohl

so ziemlich alle vorhandenen Agenten überfordern,<sup>10</sup> kann nur denjenigen überraschen, der die kognitive Leistung des menschlichen Spracherwerbs gröblich unterschätzt. Während also das konstruktive Vorgehen der Informatik die agententechnologische Modellierung sozialer Koordinationsmechanismen dort vereinfacht, wo es gilt, dysfunktionale Varianzen im Verhalten der Individuen zu vermeiden, schlägt es umgekehrt zum Nachteil aus, wo es darum geht, überschießende Varianz als Innovationspotential einzubauen.

#### **4. Vergesellschaftung von Agenten: Koordination mittels technisierter Formen sozialer Verhaltensabstimmung**

Einer bekannten Unterscheidung zufolge, die Weber in Erweiterung entsprechender Überlegungen von Tönnies aufgestellt hat, kann man auf einer sehr allgemeinen Ebene zwei Formen der Strukturierung sozialen Handelns auseinander halten: Vergemeinschaftung und Vergesellschaftung: „Vergemeinschaftung‘ soll eine soziale Beziehung heißen, wenn und soweit die Einstellung des sozialen Handelns – im Einzelfall oder im Durchschnitt oder im reinen Typus – auf subjektiv *gefühlter* (affektiver oder traditionaler) *Zusammenhörigkeit* der Beteiligten beruht. ‚Vergesellschaftung‘ soll eine soziale Beziehung heißen wenn und soweit die Einstellung des sozialen Handelns auf rational (wert- oder zweckrational) motiviertem *Interessenausgleich* oder auf ebenso motivierten *Interessenverbindungen* beruht.“ (Weber 1922/1972: 21) Eine besonders typische Form der Vergesellschaftung ist, so Weber, der Markttausch, ein geeignetes Beispiel für Vergemeinschaftung die Familie.

Für den hier interessierenden Zusammenhang ist diese Unterscheidung von Bedeutung, weil vergemeinschaftete und vergesellschaftete Formen sozialer Verhaltensabstimmung einige charakteristische Unterschiede aufweisen, die sich auf Bedingungen und Möglichkeiten ihrer Nutzbarkeit als Mechanismen der Koordination technischer Agenten auswirken: Formen der Vergemeinschaftung beruhen wesentlich auf Bekanntschaft der involvierten Akteure, entstehen auf der Grundlage des bereits gemeinsam Erlebten und setzen eine gewisse Reichhaltigkeit und Dichte der Interaktion voraus. Formen der Vergesellschaftung dagegen beruhen wesentlich auf rationalem Kalkül. Sofern man davon ausgehen kann, dass sich die jeweiligen Interaktionspartner in dem fraglichen Sinne rational verhalten, muss man darüber hinaus wenig über sie wissen. Formen der Vergesellschaftung ermöglichen mithin den sozialen Austausch zwischen einander weitgehend Unbekannten, der als punktuelles Ereignis auf der Grundlage reduzierter Kommunikationserfordernisse erfolgen kann. Vergesellschaftung bietet damit Formen der Verhaltenskoordination, die die Möglichkeit ihrer Technisierung bereits in sich tragen. Um diesen Aspekt der Vergesellschaftung wird es in diesem Kapitel gehen.

---

<sup>10</sup> Im Vorschlag der FIPA für eine Agentenkommunikationssprache heißt es, es sei möglich, dass alle vorgesehenen Formen von Sprechakten einem Agenten implementiert werden, es wird aber sogleich hinzugefügt: „though this is unlikely“ (vgl. FIPA 1999: 10).

Als technisch im weitesten Sinne können alle Verfahren und Einrichtungen bezeichnet werden, mit deren Hilfe sich in vorhersehbarer und wiederholbarer Weise bestimmte erwünschte Effekte erzielen lassen, die ohne Verwendung dieser Verfahren oder Einrichtungen nicht oder nur mit größerem Aufwand zu erzielen wären. In diesem Sinne bezeichnet Luhmann Technik als „*funktionierende Simplifikation im Medium der Kausalität*“. Man kann auch sagen, daß innerhalb des simplifizierten Bereiches  *feste* (im Normalfall funktionierende, wiederholbare usw.) *Kopplungen* eingerichtet werden, was aber nur möglich ist, wenn die Interferenz externer Faktoren weitgehend ausgeschaltet wird. Deshalb kann man Technik auch als weitgehende *kausale Schließung* eines Operationsbereiches begreifen. Das Resultat von Technisierung ist also eine mehr oder weniger erfolgreiche Isolierung von Kausalbeziehungen mit der Folge, daß (1) Abläufe kontrollierbar, (2) Ressourcen planbar und (3) Fehler (einschließlich Verschleiß) erkennbar und zurechenbar werden.“ (Luhmann 1991: 97f.) Die Technisierung sozialer Interaktionszusammenhänge richtet sich insbesondere darauf, die Kontingenz menschlichen Handelns zu reduzieren, also das Handeln der Beteiligten in einer Weise zu konditionieren, dass Anschlusshandlungen prognostizierbar werden.

Der Gesellschaftstheorie Luhmanns gelten die sogenannten symbolisch generalisierten Kommunikationsmedien als prägnanteste Formen der Technisierung sozialer Interaktionszusammenhänge. „Jede Theorie der Kommunikationsmedien hat“, so Luhmann, „davon auszugehen, daß nichtidentische Selektionsperspektiven vorliegen und selektiv zu verknüpfen sind“ (ders. 1975b: 174). Die Grundsituation der Kommunikation ist die der doppelten Kontingenz, womit ausgedrückt ist, dass die Kommunikationspartner in ihrem Verhalten und ihren Verhaltenserwartungen wechselseitig füreinander undurchsichtig sind (vgl. ders. 1987: 148ff.). Dies bedeutet zunächst, dass derjenige, der mit einem Selektionsvorschlag beginnt, nicht wissen kann, ob der oder die Partner die vorgeschlagene Situationsdefinition annehmen oder ablehnen wird bzw. werden. Blicke es dabei, so „käme kein soziales System in Gang“ (ebd.: 160), da nicht gewährleistet werden könnte, „daß der eine die Selektion des anderen als Prämisse eigenen Verhaltens übernimmt“ (ders. 1975b: 174). Allerdings können „beide Partner sich wechselseitig als selektiv erlebend und handelnd erfahren und dies bei eigenen Selektionen in Rechnung stellen“ (ebd.). Geschieht dies, „so wird der Selektionsprozeß reflexiv. Die Kettenbildung kann antizipiert und zum Selektionsmotiv gemacht werden: Man stellt zum Beispiel Informationen mit Wahrheitswert (Unwahrheitswert) für das Erleben anderer bereit; oder man selegiert das Handeln anderer. Solches Durchgreifen durch Selektionsketten kann in Märkten und in Bürokratien zur Routinesache, in Liebesangelegenheiten zur Sache sensibilisierter Erfahrung werden; immer setzt es einen Code voraus, der die Selektionstypik hinreichend spezifiziert und die Kommunikation auf artgleicher Bahn hält.“ (ebd.)

Die evolutionär ausgebildeten Mechanismen, die es auf diese Weise ermöglichen, „reduzierte Komplexität übertragbar zu machen und für Anschlußselektivität auch in hochkontingenten Situationen zu sorgen“ (ebd.), bezeichnet Luhmann – in Weiterführung entsprechender Überlegungen von Parsons (1980) – als symbolisch generalisierte

Kommunikationsmedien. Seine Beispiele sind: „Wahrheit, Liebe, Eigentum/Geld, Macht/Recht; in Ansätzen auch religiöser Glaube, Kunst und heute vielleicht zivilisatorisch standardisierte ‚Grundwerte‘. Auf sehr verschiedene Weise und für sehr verschiedene Interaktionskonstellationen geht es in all diesen Fällen darum, die Selektion der Kommunikation so zu konditionieren, daß sie zugleich als Motivationsmittel wirken, also die Befolgung des Selektionsvorschlages hinreichend sicherstellen kann.“ (Luhmann 1987: 222) Dies geschieht durch eine binäre Codierung von Präferenzen: „Medien-Codes sind Präferenz-Codes. Ihre Duplikationsregel beruht auf der Wert/Unwert-Dichotomisierung von Präferenzen. Sie konfrontiert Vorkommnisse, Fakten, Informationen mit der Möglichkeit, Wert oder Unwert zu sein, zum Beispiel wahr oder unwahr, stark oder schwach, recht oder unrecht, schön oder häßlich. Daraus entsteht ein spezifizierter Selektionsdruck“ (ders. 1975b: 175), der dort, wo sich die Kommunikation an einer entsprechenden binären Codierung orientiert, besondere Kommunikationserfolge ermöglicht.

Die symbolisch generalisierten Kommunikationsmedien werden damit zu einer Erscheinungsform von Technik: „Das Technische an der Struktur der Kommunikationsmedien gründet in den Eigenschaften binärer Codes, beliebig anfangende Prozesse zu schematisieren, als Operationsabfolgen zu regulieren und in ihrer Selektivität in Kettenbildungen zu verstärken über das hinaus, was einzelne Teilnehmer überblicken und verantworten können. (...) Wo hinreichend spezifizierte Medien-Codes mit diesen Funktionen vorausgesetzt werden können, wird der Aufbau komplexer Systeme durch eine Art Autokatalyse beschleunigt. Die Orientierung an relativ *einfachen* und im sozialen Verkehr zumutbaren Regeln führt dann bei einer für das System *zufällig* variierenden Umwelt zum Aufbau von zunehmend komplexen Strukturen.“ (Luhmann 1975a: 72)

Die Theorie symbolisch generalisierter Kommunikationsmedien bringt genau den Zusammenhang zum Ausdruck, über den ich im vorigen Kapitel im Zusammenhang mit Spezialsprachen gesprochen habe. Wir sehen nun, dass Interaktionszusammenhänge, die auf der Grundlage restringierter Sprachcodes (und korrespondierender Verhaltensweisen) Handlungsketten organisieren, deren Ablauf man mit einiger Erwartungssicherheit prognostizieren kann, Formen der Technisierung des Sozialen repräsentieren. In diesem Kapitel wird es darum gehen genauer zu untersuchen, auf welche Weise die VKI auf solche technischen Formen sozialer Verhaltensabstimmung zugreift. Dazu betrachte ich die vier Formen symbolisch generalisierter Interaktionsmedien, die Parsons (1980) unterscheidet: Geld, Macht, Einfluss und Wertbindung.

## **4.1 Markttausch**

Bereits in den Überlegungen des vorangegangenen Kapitels über restringierte Kommunikationsprotokolle hatte sich der Markttausch als besonders profilierter Mechanismus der Verhaltensabstimmung zwischen Agenten erwiesen. Das Kontraktnetz-Protokoll nimmt in der VKI eine Schlüsselstellung ein und dürfte der inzwischen am meisten verwendete Koordinationsmechanismus in Multiagentensystemen sein:

„Der Punkt bei dem Kontraktnetz ist, das sagt dir im Prinzip: Gut, ich habe irgendeinen Job zu tun, ich frage die Leute (d.h. die Agenten, Anm. d. Verf.). Und die Leute geben mir eine Zahl zurück und der mit der höchsten Zahl gewinnt. Und die Tatsache, dass das eigentlich nach wie vor das ist, was alle Leute (d.h. die Informatiker, Anm. d. Verf.), wenn der Name VKI fällt, zuerst sagen: das ist es. Das ist schon irgendwie bezeichnend. Das eigentliche Problem ist doch: Ja, wie wird denn diese Zahl berechnet? Und da (...) kriegt man eigentlich viel zu wenig Hilfe. (...) Ja. Das Kontraktnetz ist schön und gut, aber es löst dir allein nicht die Probleme. Gibt dir nur einen allgemeinen Mechanismus, wie du eben Aufgaben in einer verteilten Umgebung an die beste Partie [verteilen kannst].“ (UW5, 14: 39 - 15: 3)

Das Zitat bestätigt den zentralen Stellenwert von Kontraktnetzen in der VKI, betont aber zugleich, dass es sich hier zunächst erst um einen abstrakten Mechanismus handelt, der, um im konkreten Fall zu funktionieren, mit Inhalt gefüllt werden muss. Auf die darauf zielende Frage „Wie wird denn diese Zahl berechnet?“ gibt es zwei unterschiedliche Antworten. Bevor ich mich diesem Punkt zuwende, möchte ich jedoch zuerst einige Anwendungsfelder des Kontraktnetzes aufführen, die die breite Verwendung dieses Koordinationsmechanismus in der Multiagententechnologie dokumentieren:

- *Fertigungssteuerung*: Hier geht es darum, die Verteilung von Aufgaben auf Maschinen zu planen und zu steuern. Die einzelnen Maschinen werden dabei durch Agenten repräsentiert, die um die Aufgaben konkurrieren. Die Aufgaben werden ausgeschrieben, die Maschinen-Agenten, die in der Lage sind, die Aufgabe durchzuführen, geben ein Gebot ab, worauf die Maschine, deren Agent das beste Angebot macht, die Aufgabe übertragen bekommt. Als wesentlicher Vorteil gegenüber zentraler Planung wird die größere Robustheit und Flexibilität dieses Verfahrens angeführt. Wie wir in informellen Äußerungen erfahren haben, wird ein entsprechendes Fertigungssteuerungssystem, das am Daimler-Benz-Forschungszentrum in Berlin entwickelt wird, vermutlich in Deutschland das erste Multiagentensystem sein, das im industriellen Realbetrieb eingesetzt wird.

„*Contract net* ist ein Koordinationsverfahren, wo Sie im Prinzip eine Aufgabe auf andere Agenten verteilen. Wirklich einfach gesprochen, könnte man so was einsetzen, um die Verarbeitung auf Maschinen zu verteilen. Beim *contract net* ist es im Prinzip so, es wird ja als Verhandlungsmechanismus bezeichnet, ist aber letztendlich keiner, weil ich im Prinzip nicht diesen Verhandlungsprozess habe, dass zwei aufeinander zugehen. Es ist letztendlich so: Einer sagt: ‚Ich habe ein Problem.‘ Der schreibt das an bestimmte Leute aus, die geben ihm halt ein Angebot ab. D.h. die sagen, wie sie das Problem lösen, und er wählt dann halt das beste aus. Das ist letztendlich nichts anderes als ein *task distribution*-Verfahren. Irgendwie eine Aufgabe an andere verteilen. Da habe ich Kriterien, nach denen ich das mache, das sind dann halt meine Optimierungskriterien, die wende ich darauf an. Aber im Prinzip verteile ich sie nur. Das ist so ein typisches Koordinationsverfahren.“ (IW1, 9: 40 - 10: 4)

- *Transportlogistik*: Die Nutzung des Kontraktnetzes für verteilte Logistikanwendungen ist insbesondere am Beispiel des Speditionsszenarios von einem Team am Deutschen Forschungszentrum für Künstliche Intelligenz (DFKI) in Saarbrücken untersucht worden. Hier geht es darum, Speditionsaufträge so auf die einzelnen Transporteinheiten (z.B. LKWs) zu verteilen, dass bei möglichst hoher Auslastung die möglichst kürzeste Wegstrecke zurückgelegt wird. Das Multiagentensystem „MARS“ modelliert Software-Agenten, die je einer Transporteinheit zugeordnet sind (LKW-Agenten). Die eingehenden Speditionsaufträge werden von einem Manager-Agenten ausgeschrieben. Die LKW-Agenten bewerten auf Grund der Kenntnis ihrer eigenen Situation (gegen-

wärtiger Standort, freie Kapazitäten usw.) zu welchen Kosten sie den Auftrag annehmen können und senden ein Gebot. Der Manager-Agent wählt das günstigste Gebot aus und erteilt den Zuschlag (vgl. Fischer et al. 1995).

„[Bei] diesen Verkehrsapplikationen, wo es darum geht, Aufgaben zwischen verschiedenen Fahrzeugen zu verteilen, da läuft dann so ein *contract net*-Protokoll, wobei dann einer eine Ausschreibung macht und sagt: ‚Hier ist die *task*, sagt mir wieviel es kostet.‘ Und dann sagen die einzelnen *trucks* halt, die geben halt ein *bid* ab, das heißt sie schicken: ‚Ich biete so und soviel.‘ Und oben der entscheidet und sagt: ‚*grant*‘ oder ‚*reject*‘ zu den anderen.“ (FW3, 14: 18-22)

- *Agentenunterstützte Abwicklung von electronic commerce*: Benutzer-Agenten gehen im Internet auf die Suche nach vom Benutzer nachgefragten Diensten oder Gütern, einer bestimmten Information etwa, einer CD oder Hotelübernachtungen und Beförderungsangeboten für eine geplante Reise. Sie suchen entweder selbst oder beauftragen einen Broker-Agenten, der einen Marktüberblick für die entsprechende Domäne besitzt. Die Koordination von Nachfrage und Angebot läuft über ein Kontraktnetz, wobei es entweder darum geht, das billigste oder das am besten passende Angebot (oder eine Kombination aus beidem) herauszufinden:

„Also die grobe Architektur ist, dass der Benutzer einen Benutzer-Agenten hat, dieser Benutzer-Agent bekommt den Auftrag von dem Benutzer, also quasi: ‚Plane mir eine Reise von dort bis dort und dahin bis dahin!‘ Und zuerstmal muss er Kontakt aufnehmen mit einem sogenannten *broker agent* oder *travel agent*, der weiß, wie diese Anfrage zu dekomponieren ist, der Benutzer-Agent gibt (...) dem Broker-Agent (...) die sonstigen Rahmenbedingungen, also: nur mit Flugzeug oder mit Auto und Flugzeug oder mit Bahn oder was immer, darf so und so viel kosten, wie auch immer, also diese Rahmenbedingungen. Der Broker-Agent ist dann dafür zuständig, diese Aufgabe zum einen zu dekomponieren (...) und ist dafür auch zuständig dann, die entsprechenden Aufgaben, also die entsprechenden Dienste-Agenten aufzusuchen und die diesen auch dann zu stellen. (...) Der Broker-Agent koordiniert diese Dienste, also verteilt da die Aufgaben, schickt es an die Dienste weiter und die geben ihre Rückmeldungen und der sammelt dann die Meldungen und stellt es zusammen und gibt es an den Benutzer-Agenten weiter.“ (IW2, 3: 44- 4: 5)

- *Terminplanung*: Die einzelnen Agenten repräsentieren das Zeitbudget ihrer Benutzer und handeln mittels Kontraktnetz gemeinsame Termine aus:

„Und das haben wir zum Beispiel auch gesehen, im Rahmen von FIPA, dass so ein *contract net* auch gut für die Terminvereinbarung zum Beispiel passt, also diese verteilte Terminplanung, also was bei uns und vielen anderen so angeschaut wird, dass wenn man *contract net* so definiert, dann kann man sagen: Terminvereinbarung ist einfach *contract net*, da bekommt man Vorschläge zurück und man wertet diese Vorschläge aus und entweder man akzeptiert dann alle, mit einem bestimmten konkretisierten Vorschlag, der diesen verschiedenen Vorschlägen entgegen gekommen ist oder aber man muss die alle ablehnen, weil es keinen überlappenden Termin gibt.“ (IW2, 8: 36-41)

- *Mensch-Maschine-Schnittstelle*: In diesem bereits oben angesprochenen Anwendungsfall geht es um die Adaption von Computersystemen an persönliche Gepflogenheiten eines Benutzers. Beispielsweise gibt der Benutzer den Befehl, einen auf dem Bildschirm repräsentierten Gegenstand nach links zu bewegen. Dies kann nun – je nach Eigenart des betreffenden Benutzers – entweder bedeuten „vom Benutzer aus gesehen nach links“ oder aber „aus der auf dem Bildschirm dargestellten Perspektive nach links“. Verschiedene Agenten repräsentieren diese unterschiedlichen Deutungsmöglichkeiten und konkurrieren mittels Kontraktnetz um die zutreffende Deutung des Befehls.



Diese Beispiele illustrieren die breite Verwendung des Markttausches als Koordinationsmechanismus zwischen Agenten. Kontraktnetze werden nicht nur für Anwendungen im Bereich des *electronic commerce* vorgesehen, für die „Marktplatz logischerweise die beste Metapher“ (UW12, 6: 5) ist. Auch für Abstimmungsprozesse innerhalb einzelner Organisationen werden entsprechende Multiagentensysteme modelliert – wie in den Beispielen der Fertigungssteuerung und des Speditionsszenarios – und reproduzieren hier gleichsam das Vorgehen der Einrichtung unternehmensinterner Profit-Center. Schließlich finden sich aber auch Anwendungen, in denen es um Abstimmungsprozesse geht, die man üblicherweise nicht als Gegenstand von Markttausch thematisiert: den „Tausch“ von Terminvorschlägen oder Deutungen von Benutzerabsichten gegen Zustimmung. Diese Vielfältigkeit der Verwendung des Kontraktnetzes unterstreicht dessen Leistungsfähigkeit als allgemeiner Koordinationsmechanismus, der auf sehr unterschiedliche Gegenstände bezogen werden kann. Ein Koordinationsmechanismus, auf den zuzugreifen für die VKI offensichtlich so nahe liegt, dass er auch für Problemstellungen gewählt wird, für die er möglicherweise ungeeignet ist:

„Also, z.B. einer von uns, der schreibt über Management von Netzwerkkapazitäten. Und der hat auch die Marktplatzmetapher und da gibt's immer Verwirrungen: Wo ist jetzt da der Marktplatz, ist der am Server, ist der am einzelnen Client, ist der an irgendwelchen Knoten, ist der am Rooter, (...) wo ist der eigentlich? (...) Weil: Da passt die Metapher eigentlich überhaupt nicht. Und der hatte in den Diskussionen immer größere Schwierigkeiten als andere, seine wissenschaftliche Idee darzulegen. (...) Und wenn er vielleicht eine bessere oder geeignetere Metapher gewählt hätte, hätte er vielleicht weniger Schwierigkeiten, das zu erklären, was er erreichen will.“ (UW12, 6: 25-37)

Ich komme damit zu der Frage, wie die Inhalte spezifiziert werden, also zu der Frage des eingangs zitierten Gesprächspartners: Wie wird die Zahl berechnet, auf deren Grundlage die Angebote der Bieter verglichen werden? Bei genauerer Betrachtung hat diese Frage zwei Aspekte: Erstens: Nach welchen Kriterien berechnen die Bieter die Zahl? und zweitens: Auf welche Weise wird die Zahl zum Kriterium der Auswahl zwischen Angeboten? Wie wir gleich sehen werden, hängen beide Fragen eng miteinander zusammen.

Eine Antwort auf diese Fragen lautet, dass sich die Berechnung des Angebotes direkt aus bestimmten festgelegten Größen des Bereichs ergibt, die der jeweilige Bieter-Agent repräsentiert: So ist beispielsweise durch die Auslegung einer Maschine festgelegt, wie lange sie braucht, um ein bestimmtes Werkstück zu bearbeiten. Im Speditionsszenario bieten Informationen wie „Ich muss so und so viele zusätzliche Kilometer fahren und erreiche eine Auslastung von so und so viel Prozent, wenn ich den Auftrag übernehme“ die Grundlage für die Berechnung von Angeboten. In anderen Fällen setzt der Anbieter von vornherein bestimmte Preise (für Hotelübernachtungen, CDs, Informationen usw.) fest. Die Zahlen, die die Bieter als Angebot abgeben, repräsentieren also ein bestimmtes zur Erledigung der fraglichen Aufgabe erforderliches Zeitintervall, eine Messgröße für Auslastung und Verbrauch von Ressourcen, eine bestimmte Geldsumme usw. Dadurch werden sie in vielen Fällen gleichsam von selbst zum Kriterium für die Auswahl zwischen Angeboten: nämlich, wenn das Ziel ist, die kürzeste Zeit, die beste Ressourcenausnutzung oder den billigsten Preis zu realisieren. In diesen Fällen wird der Koordinationsmechanismus des Markttausches deshalb zu einer hochgra-

dig technisierten Form der Verhaltensabstimmung, weil sich die Erstellung der Angebote und damit zugleich die Auswahl zwischen ihnen an berechenbaren Größen orientiert. Die Abstimmung von Angebot und Nachfrage erfolgt gleichsam automatisch, ohne zusätzlichen Verhandlungsbedarf und kommt damit dem ingenieurtechnischen Problemlösungsinteresse sehr weit entgegen:

„Aber da (im Fall der Fertigungssteuerung, Anm. d. Verf.) würde ich nun wieder meinerseits das Ziel haben: Ich möchte es also so einfach machen, dass die [Agenten] gar nicht mehr viel miteinander verhandeln. Eigentlich sind die Kriterien ja klar, nach denen sie entscheiden sollen. Und wenn ich jetzt genau weiß, nach welchen Kriterien sozusagen in der Verhandlung zum Schluss [ein Agent] obsiegt, dann kann ich gleich ein Protokoll machen, was so arbeitet. Dann brauche ich eigentlich nicht lange erst Daten hin und her zu schieben, sondern ich könnte vielleicht gleich solche *tools* da einbauen, die von vornherein abwägen, die Interessen, die jetzt da anstehen.“ (UP1, 26: 43 - 27: 3)

Es kann allerdings eine ganze Reihe von Gründen geben, die einen derartig hohen Grad an Technisierung marktförmiger Abstimmungsprozesse verhindert. Dies ist etwa der Fall, wenn es nicht das eine passende Angebot für eine Nachfrage gibt, sondern verschiedene nach unterschiedlichen Kriterien mehr oder weniger passende, oder wenn die Nachfrage selbst präzisierungsbedürftig ist. Mit solchen Komplizierungen hat man es etwa bei der Reiseplanung sehr schnell zu tun:

„Wenn im Reisebüro irgendjemand sitzt, der würde jetzt (nachdem der Kunde den Wunsch geäußert hat, am 24. Februar für eine Woche auf Zypern Urlaub zu machen, Anm. d. Verf.) nachschlagen in seiner Datenbank. Und würde sagen: ‚Tut mir leid, ich habe das nicht für Sie.‘ Dann wäre die nächste Frage: ‚Naja, vielleicht versuchen wir es mal Anfang März.‘ (...) Das nützt mir nichts, wenn ich bloß laufend negative Antworten kriege. Ich möchte sozusagen mal wissen, was er mir jetzt als nächste Alternativen bieten würde dabei. (Der Gesprächspartner führt uns seinen Prototyp eines Reiseplanungssystems vor. Es gibt auf die obige Anfrage ein Angebot für eine dreiwöchige Reise nach Zypern ab dem 13. März aus.) Das Nächste wäre, dass ich jetzt vielleicht sagen würde: ‚Also tut mir jetzt leid, der 13. März geht auf keinen Fall. Ich möchte in jedem Falle was am 24. Februar haben. Und für sieben Tage. Und wenn es eben dann partout nicht Zypern ist, dann kannst du mich auch nach, was weiß ich, nach den Kanaren schicken.‘ Dass ich also solche Sachen (...) auch mit dem Computer verhandeln kann, je nachdem, wo ich hin will. (...) man könnte den Computer natürlich auch in die Lage versetzen, dass er seinerseits wieder sagt, also was weiß ich: ‚Ich könnte jetzt ein Angebot machen‘ (...) also: ‚Ich hätte sofort was auf den Kanaren.‘ Wie ein guter Reisebüroangestellter das auch machen würde.“ (UP1, 26: 9 - 32) „Wenn ich also jetzt in die Verhandlung reingehe und eigentlich gar nicht genau weiß, was möchte ich denn nun eigentlich. Ich möchte eine schöne Reise machen. Aber ich möchte eigentlich erstmal wissen, was der mir eben anbieten kann. Insofern könnte dann natürlich eine Verhandlung länger dauern. Dass man einfach sagt: ‚Ich gebe erstmal ein paar vage Stichpunkte vor und mit der Zeit wird es sozusagen von beiden Seiten her präziser.“ (UP1, 28: 25-31)

Ein anderer Grund kann darin bestehen, dass Anbieter und/oder Nachfrager ihre Gewinnmarge durch strategisches Handeln zu vergrößern suchen, also etwa zunächst zu niedrigeren Preisen nachfragen oder zu höheren anbieten als sie letztlich zu realisieren bereit sind. Sie legen also die berechnete Zahl, was eine Dienstleistung oder eine Ware ihnen Wert ist bzw. sie kostet, nicht offen, sondern versuchen „zu pokern“ (so UP1, 27: 17).

„Das ist eine andere Sache, wenn ich jetzt hier im Finanzsektor sowas machen würde (d.h. einen den Marktmechanismus verwendendes Multiagentensystem nutzen, Anm. d. Verf.). Dann muss ich natürlich mit verdeckten Karten spielen, die nicht offen liegen lassen. Aber wenn es um Maschinenpläne geht, dann möchte ich eigentlich möglichst schnell zum Resultat kommen. Und eigentlich sind diese Kriterien klar, aber ich brauche nichts zu verstecken. Also muss ich versuchen, möglichst gleich von

Anfang an möglichst viel Information rüber zu schicken, damit mein Partner, mit dem ich verhandle, möglichst gut entscheiden kann.“ (UP1, 28: 12-19)

Für die Weiterentwicklung des Reiseplanungssystem, an dem der im Folgenden zitierte Gesprächspartners mitarbeitet, sind solche realweltlichen Preisverhandlungen denn auch vorgesehen:

„Es sind wirkliche Verhandlungen auch angedacht, also man stellt sich vor, dass man dann entsprechend mit dem Hotel-Agent über den Zimmerpreis verhandeln kann oder mit dem Parkraum-Agent über den Preis vom Parkplatz,“ (IW2, 4: 11-13)

Die zweite Antwort auf die Frage, wie denn die Kriterien festgelegt werden, auf deren Grundlage im Kontraktnetz die Abstimmung von Angeboten und Nachfragen erfolgt, lautet also: durch Verhandlung. Nun kann man sich durchaus vorstellen, dass man Agenten verhandlungsfähig machen kann, indem man sie mit Regeln versorgt, nach denen sie bereit sind, vom ersten Angebot bzw. der ersten Nachfrage abzugehen. Beispielsweise könnte man einem Hotel-Agent die Regel mitgeben, einen bestimmten Preisnachlass in Abhängigkeit von der Länge des Kundenaufenthaltes des Hotels zu gewähren, regelmäßige Kunden zu begünstigen oder Ähnliches, also Vorgehensweisen zu realisieren wie sie auch der menschliche Hotelmanager befolgt. Mit einer solchen Implementation von Verhandlungsspielräumen verbinden sich jedoch zwei Probleme: Zum einen kann die erforderliche Wissensbasis entsprechender Agenten sehr schnell sehr umfangreich werden, so dass es unter Umständen weniger aufwendig ist, entsprechende Aufgaben bei den menschlichen Aufgabenträgern zu belassen. Zum anderen setzt sie eine Übertragung von Entscheidungsverantwortung an technische Agenten voraus. Und dies ist etwas, vor dem die meisten unserer Gesprächspartner zurückschrecken. Ich komme auf diesen Punkt weiter unten (auf S. 41ff.) zurück.

Einige unserer Gesprächspartner streichen besonders deutlich heraus, dass man im Markttausch zwischen Agenten nicht unbedingt davon ausgehen kann (bzw. sollte), dass die Abstimmung von Angeboten und Nachfragen gleichsam von selbst erfolgt, sondern Verhandlungen erfordert. Auffällig ist, dass diese Gesprächspartner zur Begründung gerne auf den Markttausch zwischen menschlichen Akteuren verweisen:

„,Informations-Agenten werden bezahlt, müssen dem anderen was geben‘ – diese Transaktionskosten[betrachtung] ist langweilig. Vor allem wenn sie autonom sind, werden die ja wohl nicht alles akzeptieren. Machen wir ja auch nicht, wenn wir (...) miteinander handeln würden, auf dem Marktplatz geht das ja genauso. So und da ist man schon schwupp bei Mechanismen oder Methoden: Wie machen es denn die Menschen?“ (UW2, 21: 46-54)

„Also wenn man irgendwie so forscht und vorwärts denkt, und sich fragt: ‚Was sollen denn Agenten können, die was einkaufen?‘ Nicht, dann ist es so am nahe liegendsten zu fragen: ‚Was würde denn ein Mensch machen, der in ein Kaufhaus geht und verhandelt, oder der irgendwelche Verhandlungen führt?‘“ (UW12, 6: 1-4)

Allerdings ist dieser Verweis auf den Markttausch zwischen Menschen nicht durchgängig überzeugend. Wer beispielsweise im Kaufhaus versuchen würde, mit dem Verkäufer in Preisverhandlungen zu treten, würde auf einiges Unverständnis treffen. Auch hier erfolgt das Zusammenfügen von Angeboten und Nachfragen gleichsam von selbst, also ohne zusätzlichen Verhandlungsbedarf. Denn die Angebotspreise sind Festpreise, was die Austauschsituation dahingehend vereinfacht, dass dem Kunden nur die Alter-

native bleibt, zu dem Preis zu kaufen oder nicht zu kaufen. Oder wie es einer unserer Gesprächspartner formuliert:

„Also wenn jetzt Situationen entstehen, wo ich zwischen sehr vielen verschiedenen Alternativen zu entscheiden habe, [wo] nicht von vornherein klar ist, welche Alternative ich nehmen werde, da wird die Sache dann komplex. Die Kunst für mich ist gerade, das so runterzubrechen, dass man das vermeidet. Also einfacherere Situationen hat. Und (...) ich denke mal, dass auch im menschlichen Leben viel so funktioniert. Wir verhandeln ja nicht wirklich. In der Regel gehen wir in einen Laden und wissen, dass der Verkäufer uns diesen Gegenstand zu verkaufen hat, weil der daran interessiert ist, seinen Geschäftsumsatz zu erhöhen und sind höchstens frustriert, wenn er sich also nicht entsprechend verhält. Wenn wir aber auch nicht mit ihm in Verhandlung treten (...), dann gehen wir einfach raus und gehen in den nächsten Laden. Ja? Also im Prinzip, unser ganzes menschliches Zusammenleben ist ja auch darauf abgestimmt, dass wir also möglichst einfache Mechanismen haben.“ (UP1, 31: 25-39)

In anderen Fällen dagegen, auf dem Flohmarkt etwa, gehört das Feilschen zu den Spielregeln und derjenige, der darauf verzichtet, gibt sich als Unkundiger zu erkennen. Aber auch hier ist der Verhandlungsbedarf sehr begrenzt, solange die Verhandlungspartner ihre Vorstellungen weiterhin in Form von Geldsummen ausdrücken können. Erst wo dies nicht mehr umstandslos möglich ist, also den potenziellen Tauschpartner die Anhaltspunkte für die Abschätzung fehlen, was der Tauschgegenstand ihnen und ihrem Gegenüber Wert ist, werden Verhandlungen wirklich voraussetzungsreich. Dies ist insbesondere bei Unikaten der Fall, dort also, wo keine Vergleichsmöglichkeiten mit äquivalenten Gütern bzw. Dienstleistungen bestehen bzw. dies nur eingeschränkt möglich ist. Der durchschnittliche Konsument kommt jedoch eher selten in solche Tauschsituationen.

Viele unserer Gesprächspartner diskutieren das sich gleichsam automatisch koordinierende Kontraktnetz und die verhandlungsbasierte Abstimmung von Angeboten und Nachfragen als einander gegenüberstehende Lösungswege. Aus der sozialtheoretischen Perspektive ist dagegen der viel entscheidendere Gesichtspunkt, ob ein generalisiertes Austauschmedium – das allgemeine Warenäquivalent Geld – benutzt wird. Ist dies der Fall, dann sind in den meisten Fällen auch diejenigen Tauschakte hochgradig vorstrukturiert, in denen dennoch ein gewisser Verhandlungsbedarf besteht. Dass diese Ähnlichkeit in der VKI häufig übersehen zu werden scheint, hängt sicherlich wieder mit dem umgekehrten Komplexitätsgefälle der agententechnologischen und der sozialtheoretischen Betrachtung technisierter Interaktionszusammenhänge zusammen: Aus der Perspektive der VKI führt jede Abweichung von der technischen „Ideallinie“ des verhandlungsfreien Kontraktnetzes zu einer enormen Steigerung des Modellierungsaufwandes entsprechender Multiagentensysteme. Aus der Perspektive der Soziologie kann man dagegen – vor dem Hintergrund der Kontingenz menschlichen Handelns – schon sehr viel früher von hochgradig erwartungssicheren und in diesem Sinne technisierten Interaktionszusammenhängen sprechen.

Hier wie dort ist dabei vorausgesetzt, dass die beteiligten Akteure bereit sind, sich in eine Wettbewerbssituation zu begeben und nach bestimmten Regeln „fair zu spielen“. Wie bereits angesprochen, lassen sich solche Regeln in geschlossenen Multiagentensystemen direkt vorgeben. In offenen Multiagentensystemen dagegen sind institutionelle Rahmenbedingungen erforderlich, die sich nicht unproblematisch auf der

Ebene des technischen Systems selbst implementieren lassen. Aus diesem Grund sind die derzeit erfolgreichsten Anwendungen des Kontraktnetzes solche, die Marktmechanismen innerhalb einer organisationalen Einheit realisieren, welche ihre Mitglieder zugleich auch kontrollieren kann. Die Realisierung des offenen agentengestützten Markttausches erweist sich dagegen als sehr viel schwieriger. Und dies nicht in erster Linie wegen des Problems technischer Inkompatibilitäten zwischen Agenten (die, wie das Problem der Standardisierung der Agentenkommunikation gezeigt hat, allerdings auch nicht unterschätzt werden sollten), sondern wegen des Problems der Verletzung der Spielregeln durch einzelne Agenten bzw. deren Benutzer:

„Du hast ein System, in dem gibt's irgendwo Leute, die CDs – oder Agenten, die CDs anbieten, es gibt irgendwelche anderen Leute, die CDs kaufen. Jetzt, was ist die beste Strategie für deinen Agenten? Du kannst nicht davon ausgehen, dass die Leute, die die Agenten bauen, die CDs verkaufen wollen, dass die alle fair spielen, sie wollen natürlich auch ihr Geld optimieren. Also diese ganzen spieltheoretischen Erkenntnisse.“ (IW5, 13: 41-45)

„Risikoabschätzung ist ein ganz wichtiger Faktor, den ich auch bei den Informations-Agenten einführen möchte, um halt eben Lügen im Vorhinein ein bisschen zu reduzieren. Ähm, das ist auch ein weites Thema, was die Agenten, nicht nur die Informations-Agenten, sondern Agenten allgemein vielleicht doch betrifft. Also nicht unbedingt gutwillige, also auch möglicherweise böswillige, wie kann ich das in den Griff kriegen? Und da spielt natürlich sofort soziales Verhalten mit rein: Können sich andere Agenten zum Beispiel gegenüber einem böswilligen Agenten wehren? Können sie ihn ausschließen, können sie es erkennen, können sie es denn selbstständig erkennen, können sie es im Team erkennen? Wenn sie es nur in einem Team erkennen können, dann ist es natürlich für sie selbstverständlich, nach sozialen Regularien dieses Team zu bilden. Wovon sie natürlich selber wieder einen individuellen Nutzen haben, weil sie nämlich gegenüber dem [böswilligen Agenten] in einem Team geschützt werden usw. Das sind aber alles menschliche Verhaltensmuster.“ (UW2, 20: 10-24)

Der in der letzten Äußerung enthaltene Vorschlag, soziale Regularien einzuführen, um Vorsorge gegen ein missliebiges Verhalten einzelner Agenten zu treffen, leitet direkt über zu dem zweiten Koordinationsmechanismus, den ich in diesem Kapitel besprechen möchte: der Etablierung und Durchsetzung kollektiv verbindlicher Entscheidungen durch direkte Machtausübung oder durch verbindliche Gesetze. Ich schließe diesen Abschnitt mit einem Zitat ab, der die besondere Eignung des Kontraktnetzes als eines technischen Mittels der Agentenkoordination noch einmal unterstreicht. Bezeichnend an diesem Zitat ist, dass die Herkunft des Kontraktnetzes aus der Welt menschlicher Handlungskoordination zwar erwähnt wird, der Punkt, um den es mir in meiner Analyse hier besonders ging, aber unterschlagen wird: dass die technische Eignung des Kontraktnetzes darauf beruht, dass bereits der über Geld abgewickelte Markttausch zwischen Menschen eine hochgradig technisierte Form von Interaktion ist.

„KI ist motiviert dadurch, dass man Intelligenz abbilden kann. Dasselbe gilt ja auch irgendwie für die VKI. Nur am Ende kommt ja bei der KI und bei der VKI immer eine Software-Technik raus. Software-Technik heißt: Ich habe ein Problem, ich habe bestimmte Algorithmen als Vorgehensweisen und Ansätze und mit denen entwickle ich eine Lösung. Sicher auch dieses *contract net* ist sicher auch mal motiviert, wie es halt in Organisationen funktioniert oder wie Ausschreibungsverfahren funktionieren. Aber was rausgekommen ist, ist eine Software-Technik.“ (IW1, 10: 14-22)

## 4.2 Macht und Gesetz

In Interaktionszusammenhängen, in denen sich die angestrebte Handlungskoordination nicht auf der Grundlage eines wie immer gearteten Eigeninteresses der Beteiligten an

dem erforderlichen Austausch von Leistungen organisieren lässt, wird im Bereich menschlicher Sozialität nicht selten auf den Mechanismus machtförmiger Verhaltensabstimmung zurückgegriffen. Macht kann man mit Weber definieren als „Chance, innerhalb einer sozialen Beziehung den eigenen Willen auch gegen Widerstreben durchzusetzen, gleichviel, worauf diese Chance beruht“ (Weber 1922/1972: 28), oder spezifischer als „Chance, für einen Befehl bestimmten Inhalts bei angebbaren Personen Gehorsam zu finden“ (ebd., Weber nennt dies Herrschaft). Macht kann wie Geld als ein symbolisch generalisiertes Kommunikationsmedium verstanden werden: Auch hier geht es um einen abstrakten Mechanismus, der sich für verschiedenste Koordinationsbemühungen in Anspruch nehmen lässt: „Die Form der Macht ist nichts anderes als (...) die Differenz zwischen der Ausführung der Weisung und der zu vermeidenden Alternative.“ (Luhmann 1997: 356) Solange „die Vermeidungsalternative“ (ebd.) für die Adressaten einer Weisung weniger anstrebenswert ist als der Gehorsam und solange die Sanktionsdrohung glaubhaft ist, bestehen gute Chancen tatsächlich Gehorsam zu finden. Ohne diese symbolische Generalisierung wäre Macht auf physischen Zwang reduziert und verlöre damit den größten Teil ihrer Koordinationskraft. Machtförmige Handlungskoordination in diesem Sinne ist nicht nur die Interaktionssequenz von direkt geäußelter Weisung und deren Befolgung. Auch Anweisungen, die in den Verfahrensregeln einer Organisation oder den Gesetzen des politischen Gemeinwesens niedergelegt sind, sind Formen machtgestützter Verhaltensabstimmung.

Angesichts der Bedeutung, die Macht im Gefüge menschlicher Gesellschaften besitzt, ist es zunächst einigermaßen überraschend, dass sie als Koordinationsmechanismus in der VKI vergleichsweise wenig thematisiert wird. Die Gründe hierfür sind wiederum in dem bereits mehrfach angesprochenen Umkehrungsverhältnis von Einfachheit und Komplexität zu suchen: Die unter dem Gesichtspunkt ihrer technischen Nutzbarkeit einfachste Form machtförmiger Koordination ist zweifellos die strikt gekoppelte Abfolge von Befehl und Gehorsam in streng hierarchisch strukturierten Interaktionszusammenhängen. Sie repräsentiert die am weitesten gehende Technisierung von Machtbeziehungen und damit die Form, die sich am unproblematischsten technisch nachbilden lässt. Aus Gründen, die wir gleich kennen lernen werden, ist die VKI an ihr jedoch nicht sonderlich interessiert. Ein größeres Interesse besteht an Formen machtförmiger Koordination, die von dieser technischen Ideallinie abweichen. Zugleich ergeben sich dann aber auch ernsthafte Umsetzungsschwierigkeiten.

Woraus resultiert das Desinteresse der VKI an hierarchischen Machtbeziehungen? Um diese Frage zu beantworten, empfiehlt es sich auf eine Unterscheidung zu rekurrieren, die wie wenige andere zum festen Bestand der subdisziplinären Selbstbeschreibung der VKI zählt und in kaum einem Einführungstext fehlt:

„People usually distinguish two main areas of research in DAI (Bond and Gasser, 1988): distributed problem solving and multiagent systems. *Distributed problem solving* (DPS) considers how the task of solving a particular problem can be divided among a number of modules (or ‚nodes‘) that cooperate in dividing and sharing knowledge about the problem and about its evolving solution(s). In a pure DPS system, all interaction (cooperation, coordination if any) strategies are incorporated as an integral part of the system. Research in *multiagent systems* (MAS) is concerned with the behavior of a collection of (possibly preexisting) autonomous agents aiming at solving a given problem. A MAS

can be defined as ‚a loosely-coupled network of problem solvers that work together to solve problems that are beyond their individual capabilities‘ (Durfee et al., 1989). These problem solvers, often called agents, are autonomous and may be heterogeneous in nature (characterized by various degrees of problem solving capabilities).“ (Moulin/Chaib-draa 1996: 4f.)

An die Nennung dieser Alternative schließt sich in der Regel ein Plädoyer zu Gunsten der Beschäftigung mit Multiagentensystemen an:

„A MAS has significant advantages over a single, monolithic, centralized problem solver: *faster problem solving (...); decreased communication (...); more flexibility (...); and increased reliability*“ (ebd.: 5).

„Task and resource allocation are the problems of assigning responsibilities and resources for a sub-problem to a problem solver. On the one extreme the designer may make all the task assignment in advance, thus creating a non-adaptive problem-solving organization. (...) In an adaptive task allocation system where tasks are assigned dynamically, it is possible to do load balancing and bottleneck-avoidance“ (Gasser 1992: 12).

„The next choice point follows the well-known partition of DAI into *Distributed problem solving* and *Multi Agent Systems*. Here I argue that the benefits of agent oriented programming could become more obvious for the programmers community in the case of collaborating autonomous agents. Here AOP could contribute to a lot of open technical problems (while in distributed problem solving there is a comparison to other already established approaches like object oriented programming and concurrent programming – using agent oriented techniques may be of good use too, but it might be difficult to convince programmers of a new paradigm if they have useful ones).“ (Burkhard 1995: 294)

Betrachtet man die neueren Sammelbände der VKI-Community, so wird deutlich, dass das Betätigungsfeld der VKI heute nahezu exklusiv die Beschäftigung mit Agenten und Multiagentensystemen ist.<sup>11</sup> Geradezu programmatisch kommt dies im Titel eines von Gerhard Weiß 1999 herausgegebenen Bandes zum Ausdruck: „Multiagent Systems. A Modern Approach to Distributed Artificial Intelligence“. Die genannte Unterscheidung repräsentiert also keine tatsächliche Zweiteilung der VKI. *Distributed problem solving*, ein Ansatz, der in der Frühzeit der VKI einige Bedeutung hatte (vgl. Lesser/Corkhill 1987), dient vielmehr als Hintergrund, in Abgrenzung von dem es leicher fällt auszudrücken, was man eigentlich will. Auch in unseren Interviews wird einzig auf diese Weise darauf Bezug genommen:

„Und auch da (bei der Betrachtung, die über den einzelnen Agenten hinausgeht, Anm. d. Verf.) gibt es eigentlich, denke ich, wieder zwei Ansätze. Einer ist eben dieser problemlösungszentrierte Ansatz. Du hast irgendeine Anwendung und versuchst einen verteilten Ansatz dafür zu kriegen. Also *cooperative distributed problem solving*, die Ecke. Und es hat sich zum Beispiel gezeigt, dass man mit dem Ansatz ja relativ gute Lösungen für Scheduling-Probleme rauskriegt, für dynamische Scheduling-Probleme - Sachen, die auch am DFKI zum Beispiel gelaufen sind. *Workflow* ist eine andere Geschichte, *routing* von *tasks*. Da ist es natürlich dann immer so, dass wie immer bei verteilten Anwendungen, dass der zentrale Ansatz natürlich immer besser ist, so er denn realisierbar ist. Also der Grund, warum man versucht, verteilte Problemlösungen [zu machen], also diese Sicht auf Dinge zu haben, ist nicht der, dass man daraus bessere Lösungen für Probleme rauskriegt, die man anders auch lösen kann. Sondern (...) dass man das Wissen eben nicht zentral verfügbar hat. Und sich eben überlegen muss, wie kriegt man das raus (...) Eine zweite Sache ist dann eben – weniger die problemlösungsorientierte Sicht, sondern einfach die systemorientierte Sicht. Man hat eben ein verteiltes System mit mehreren autonomen Einheiten. Das System ist offen, das heißt, man hat keinen globalen

---

<sup>11</sup> In einem Gespräch am Rande der AAAI Fall Symposium Series 1997 in Boston hat Carl Hewitt die Einschätzung, dass VKI mehr und mehr identisch wird mit Multiagentensystem-Forschung, gleichfalls geäußert.

Blick auf das System, sondern man hat so was wie das Internet und man hat mehrere menschliche oder maschinelle Einheiten. Und dann hat man eben solche Sachen wie Kooperation, Verhandlung zwischen autonomen selbstinteressierten Einheiten. Und das ist eben das Interessante, denke ich: Wie weit lassen sich da Analogien aus den menschlichen Organisationsstrukturen übertragen?“ (IW4, 4: 40 - 5: 18)

Bei näherer Betrachtung zeigt sich, dass die Abgrenzung zwischen den beiden Herangehensweisen nicht eben trennscharf ist und *distributed problem solving* – das klingt in der zuletzt angeführten Äußerung an – offensichtlich doch einiges zur Entwicklung der VKI beigetragen hat. Beides kommt etwa darin zum Ausdruck, dass das Kontraktnetz, das, wie wir gesehen haben, heute zum Kernbestand der Multiagententechnologie zählt, zunächst im Rahmen des *distributed problem solving* entwickelt wurde (vgl. Davis/Smith 1983). Wie sich das Verhältnis der beiden Ansätze in der genaueren inhaltlichen Analyse darstellt, muss uns hier jedoch nicht interessieren. Für den Punkt, um den es hier geht, kommt es darauf an, wie der Ansatz des *distributed problem solving* heute wahrgenommen wird und aus welchen Gründen er so wahrgenommen wird.

Es fällt auf, dass der Ansatz des *distributed problem solving* deutlich in die Nähe zentralistischer Problemlösungsansätze gerückt wird. Die Wahrnehmung in den hier angeführten Belegstellen ist die, dass dort nur so viel Verteilung zugelassen wird, wie es das zu lösende Problem notwendig erfordert. Die Art und Weise der Aufgabenverteilung zwischen den einzelnen Problemlösern ist weitgehend vorstrukturiert oder vom Programmierer vollständig vorgegeben, mit dem Ergebnis, dass sich die entsprechenden Systeme nicht so sehr von früheren Programmarchitekturen, etwa denen der objektorientierten Programmierung, unterscheiden. Abgelehnt wird der Ansatz, weil er noch zu sehr dem Vorgehen einer globalen Planung von Problemlösungen verhaftet ist und deshalb die Autonomie der einzelnen Agenten zu sehr beschneidet. Als negative Folge wird vor allem auf die geringe Adaptivität und Flexibilität entsprechender Systeme verwiesen.

Zentrale Planung des Arbeitsablaufes, Zuweisung von Aufgaben an weisungsgebundene Stellen, Festlegung von Zuständigkeiten, Arbeitspflichten und Weisungsbefugnissen und als Kehrseite dieser Form der Arbeitsorganisation: geringe Flexibilität und Reaktionsfähigkeit bei unvorhergesehenen, und das heißt in die Planung nicht einbezogenen Ereignissen – das sind genau auch die Merkmale, die in besonderem Maße für hierarchische Organisationen prägend sind. Wir sehen nun, warum diese, unter dem Gesichtspunkt technischer Nutzbarkeit einfachste Form machtförmiger Koordination in der VKI weitgehend ausgespart bleibt: Weil sie genau diejenige Art und Weise der Organisation einer Mehrzahl von Problemlösern darstellt, in Abgrenzung von der die VKI heutzutage einen wesentlichen Teil ihrer Identität als eigenständiges Forschungsfeld bestimmt.

Nun könnte man einwenden, dass disziplinäre Identität zwar schön und gut ist, aber eine Abgrenzung töricht ist, die in diesem Fall dazu führt, dass ein wichtiger, und noch dazu technisch hochgradig verfügbarer Mechanismus menschlicher Verhaltensabstimmung unbeachtet bleibt. So einfach ist die Sache jedoch nicht. Denn die Abgrenzung der VKI von dem Mechanismus hierarchischer Organisation ist zugleich eine Abgren-



zung von dem zentralen Strukturierungsprinzip herkömmlicher technischer Maschinen. Pointiert formuliert kann man sagen, dass alle maschinelle Technik sich an der Sozialvorstellung der strikten Hierarchie orientiert. Entwicklungsgeschichtlich, so argumentiert der Technikhistoriker Mumford, ist die hierarchische Organisation menschlicher Arbeit, wie sie etwa beim Bau der ägyptischen Pyramiden erfolgte, Vorläufer und Vorbild aller späteren gegenständlich-technischen Maschinerie. Er bezeichnet sie deshalb als die große Arbeitsmaschine:

„Diese kollektiven Einheiten heute als Maschinen zu bezeichnen, ist kein leeres Spiel mit Worten. Bestimmt man eine Maschine mehr oder weniger auf Grund der klassischen Definition von Franz Reuleaux als Kombination resistenter Teile, deren jeder eine spezielle Funktion hat, unter menschlicher Kontrolle operierend, um Energie zu nutzen und Arbeit zu verrichten, dann war die große Arbeitsmaschine in jeder Hinsicht eine echte Maschine: um so mehr, als ihre Komponenten, obgleich aus menschlichen Knochen, Nerven und Muskeln bestehend, auf ihre rein mechanischen Elemente reduziert und streng auf die Ausführung begrenzter Aufgaben zugeschnitten war. (...) Die Schwierigkeit bestand darin, eine zufällige Ansammlung von Menschen, losgelöst von Familie und Gemeinschaft sowie von ihren vertrauten Beschäftigungen, jeder mit eigenem Willen oder zumindest der Erinnerung an einen solchen, in eine mechanisierte Gruppe zu verwandeln, die auf Befehl manipuliert werden konnte. Das Geheimnis der mechanischen Kontrolle bestand darin, daß ein einziger Kopf mit genau bestimmtem Ziel an der Spitze der Organisation war und daß es eine Methode gab, Anweisungen über eine Reihe von Funktionären weiterzugeben, bis sie die kleinste Einheit erreichten. Exakte Weitergabe der Anweisung und absolute Unterwerfung waren gleichermaßen von wesentlicher Bedeutung.“ (Mumford 1977: 222f.)

Das hierarchische Strukturierungsprinzip von Befehl und Befehlsausführung ist auch den technischen Produkten der Informatik unübersehbar eigen. Die imperativen (!) Programmiersprachen geben dies bereits in ihrer Benennung zu erkennen und ebenso deutlich im Sprachcode selbst. Mit der Modularisierung der Software-Architekturen verzweigt sich die zunächst lineare Kette von Anweisungen in Richtung einer zunehmend arbeitsteiligen Organisation, in der die einzelnen Teilprogramme für je unterschiedliche Teilaufgaben vorgesehen sind. Mit dem Siegeszug der Objektorientierung wird die Entkopplung dieser Module gegenüber dem Umfeld des Programms, in das sie eingebettet sind, noch ein wesentliches Stück vorangetrieben. Sie repräsentieren jetzt eigenständige Einheiten, ausgestattet mit bestimmten Daten, über die nur sie verfügen, und bestimmten ihnen eigenen Funktionen und Prozeduren. Der einzige substantielle Unterschied der Objekte des objektorientierten Programmierens gegenüber Software-Agenten ist, dass sie nicht von sich aus aktiv werden können, sondern aufgerufen werden müssen, also nur im Rahmen der Anweisungsstruktur eines Hauptprogramms funktionieren, in das sie eingebunden sind. Diesen Unterschied würde die Multiagentensystem-Forschung wieder aufheben, wenn sie auf den Koordinationsmechanismus der hierarchischen Organisation zugreifen würde. Und in diesem Sinne ist der Verzicht auf diese Form der Verhaltensabstimmung für das Forschungsprogramm der VKI in der Tat konstitutiv.

Die Nützlichkeit von *Maser-slave*-Strukturen – oder etwas freundlicher formuliert: *Client-server*-Strukturen – für bestimmte, insbesondere hochgradig routinisiert abwickelbare Abstimmungsprozesse zwischen technischen Einheiten wird von unseren Gesprächspartnern zwar anerkannt und im Einzelfall werden solche Strukturen verwendet:

„Also bei bestimmten Problemen, da muss man nicht großartige Verhandlungen und so weiter haben, sondern da braucht man einfach, will mal sagen, da reicht *master-slave* aus, so mit diesen *client-server frameworks*, wobei es gibt da ab und zu mal schon ein paar Beiträge der Agententechnologie, die dabei helfen können. Also zum Beispiel die interaktive, standardisierte aber erweiterbare Schnittstelle" (IW2, 6: 49-52).

Nichtsdestotrotz wird das Merkmal des eigenständigen Aktivwerdens der Agenten, ihre sogenannte „pro-activeness“ (Wooldridge/Jennings 1995: 2), als charakteristischer Unterschied der Multiagententechnologie gegenüber solchen Formen der hierarchischen Organisation technischer Einheiten herausgestellt:

„Also, wo ist der Unterschied zwischen dem, was die EDV heute macht und dem, was die DAI und MAS macht? Denn heute, wenn man Betriebs-EDV anschaut, da sind überall *client-server*. *Client-server*-Architekturen sind heute *state of the art* und das, was man heute macht. Aber bei *client-server*-Architekturen ist es eben so, dass der eigentliche *server* meist ein rein passives System ist. Die *clients* sind die aktiven in dem Sinne, dass sie Informationen ..., die fordern irgendwelche Dienstleistungen an, weil sie meistens von irgendeinem Menschen halt dazu angehalten wurden. Und der *server* erbringt dann eine Dienstleistung und dann ist es wieder zu Ende. Und ein Multiagentensystem ist eben anders. Da sind die einzelnen ..., da ist dann auch der *server* aktiv und wird von sich aus informiert. Und das ist auch die ganze Transformation, die auch im WWW zum Beispiel stattfindet (...), dass ich anfangs, von den Informations-Providern aktiv Dinge wegzuschicken. Und nicht nur passiv irgendwelche Dinge anzubieten und zu sagen: ‚Wenn es denn einer wirklich findet und wirklich haben will, dann gebe ich es ihm auch.‘ Und je mehr diese aktive Sache da hereinkommt, umso mehr kommen diese Aspekte der Multiagentensysteme da rein." (FW1, 11: 36-53)

Allerdings begegnet die VKI dem Merkmal der *pro-activeness* nicht ohne Vorbehalte. Wie oben bereits angedeutet, wollen die meisten unserer Gesprächspartner die Autonomie der Agenten nicht auf Bereiche menschlicher Entscheidungsverantwortung ausdehnen. In diesem Zusammenhang dient der Verweis darauf, dass die hierarchische Überordnung des Menschen gegenüber den Agenten gewahrt bleiben muss, als Korrektiv von Vorstellungen einer – aus Sicht dieser Gesprächspartner – überzogenen Ausstattung von Agenten mit Entscheidungsbefugnissen:

„Bisher habe ich das (die Agenten, Anm. d. Verf.) immer sozusagen als Knechte gesehen, als Sklaven, und zwar in Softwareform. So wie es ein Transaktionsmonitor ist oder halt eben ein SQL-Server oder wie auch immer. Meinetwegen intelligent, ein bisschen, dass man ein bisschen benutzerfreundlich ist und in dem Sinne auch autonom: Der urteilt da auch vor sich hin. Aber in dem Moment, wo jetzt über Rollen und Verhalten gesprochen wird und so Eigenschaften zugebilligt werden wie ... Adaptivität mag ja auch noch gehen, das haben wir bei verteilten Systemen ja auch. Nicht, Balance, ja lass Balance finden, wenn ein Netz überlastet ist, aber müssen sich schnell anpassen an die Umgebung und so weiter. Oder bei Expertensystemen, wenn ein Teil zusammenbricht, muss die Wissensbasis rekonfiguriert werden. Dass das nicht alles per Hand gemacht wird, dass die tools das dann merken. Ja, aber in dem Moment, wo sie wirklich in den Alltagsbereich des Menschen eingreifen – wie zum Beispiel mail lesen, wie zum Beispiel Informationen suchen für jemanden, wie zum Beispiel Raketen abfeuern in einem Auftrag eines anderen Agenten, der wiederum in Hongkong von einem Server-Agenten gesteuert wird, der wiederum in Peking und so weiter ...“ (UW2, 7: 20-43).

... dann, so ist zu ergänzen, geht es mit der Autonomie der Agenten doch entschieden zu weit. Ein gewisses Maß an hierarchischer Strukturierung von Multiagentensystemen kann, sofern es das Merkmal der *pro-activeness* der Agenten nicht vollständig aushebelt, im Interesse der Sicherstellung des Primats der menschlichen Nutzer eines solchen Systems also durchaus sinnvoll sein. Sucht man in der Welt menschlicher Sozialität nach Vorbildern solcher, weniger strikt hierarchischen Formen der Handlungskoordination, so liegt es nahe zu fragen, wie einer der eben zitierten Gesprächspartner

vorschlägt: „Wie weit lassen sich da Analogien aus den menschlichen Organisationsstrukturen übertragen?“ (UW4, 4: 5: 17f.)

Dies liegt vor allem nahe, nachdem es ein wesentliches Ergebnis der Organisationsforschung der letzten vierzig Jahre gewesen ist, das Bild von der Organisation als einer strikt zweckrational vorgeplanten, weisungsgesteuerten Form der Handlungskoordination zur Erreichung vorgegebene Ziele zu revidieren. Zumindest „einen Teil der Aktivitäten fast jeder Organisation“, so Cohen, March und Olsen (1972/1990: 331), kann man zutreffender als „organisierte Anarchien“ beschreiben. Dieser Begriff soll zum Ausdruck bringen, dass empirisch vorfindliche Organisationen zu nicht unerheblichen Anteilen durch unklare Präferenzen, durch konkurrierende Interessen und Handlungsrationaltäten einzelner Mitglieder oder Organisationsabteilungen, durch Vorgehen nach Versuch und Irrtum, durch Produktion überschüssiger Lösungen, für die erst noch nach Problemen gesucht werden muss und Ähnliches charakterisiert sind. Die Entscheidungsfindung erfolgt deshalb vielfach an den formalen Hierarchien geregelter Anweisungsbefugnisse vorbei durch informelle Kommunikationskanäle oder durch mikropolitische Aushandlung divergierender Interessen (vgl. March 1962/1990; Crozier/Friedberg 1979). Die tatsächlichen Interaktionsstrukturen eröffnen den Mitgliedern einer Organisation in diesem Sinne deutlich höhere Handlungsspielräume als die Organigramme der formalen Organisationsstruktur abbilden. Dies erhöht zugleich die Flexibilität und Adaptivität von Organisationen.

Der Vorschlag, „ideas about human organizations as a source of inspiration in the design of computer systems“ (Malone 1990: 56) zu nutzen, ist in der VKI schon sehr früh formuliert worden (vgl. Fox 1981: 70). Aus dieser Perspektive geht es darum, unter Rückgriff auf die Organisationsforschung unterschiedliche Typen von Organisationen zu unterscheiden, ihre Vor- und Nachteile mit Blick auf unterschiedliche Problemstellungen zu identifizieren und die für die jeweilige Problemstellung angemessenste Organisationsstruktur im Multiagentensystem nachzubilden. Entsprechende Organisationstypologien für Multiagentensysteme finden sich etwa bei Fox (1981) und bei Malone (1990). Ansätze einer dynamischeren Betrachtung von Organisationsstrukturen für Multiagentensysteme sind unter Rückgriff auf soziologische Konzeptionen des symbolischen Interaktionismus insbesondere von Gasser et al. (1989) und Bond (1990) vorgestellt worden.

Die Idee der Nutzung von Organisationsstrukturen als Modell der Agentenkoordination ist in der VKI-Community allerdings ohne sonderliche Resonanz geblieben. Nach Meinung eines, mit dieser Thematik besonders vertrauten Gesprächspartners, sind die Gründe dafür zu einem Teil in der Art und Weise zu suchen, in der die drei wichtigsten Protagonisten diesen Vorschlag unterbreitet haben:

„Wir haben in dem Umfeld glaub ich drei Namen, über die es sich lohnt zu reden. Das ist der Gasser und das ist der Melone und das ist der Fox. Ja, und bei dem Gasser z.B. ist sichtbar, dass er sehr informale Organisationsbegriffe präferiert, und sich nicht so gerne einlassen möchte auf die Arbeit von dem Mintzberg (vgl. Mintzberg 1979, Anm. d. Verf.) beispielsweise. (...) Wobei er ja selbst derzeit in einer Behörde arbeitet, wo es sicherlich festgelegte Regeln gibt, und so weiter. Okay, aber da wollte er also wirklich nicht dran. Ne, da hat er sich also dagegen gewehrt. Und der Gasser ist sicherlich einer, der tief in der VKI drinnensteckt. Das ist der Melone, jedenfalls zu dem Zeitpunkt, als

er diesen Aufsatz verfasst hat (gemeint ist wahrscheinlich Malone 1987, der erste Aufsatz des Autors zu diesem Thema, Anm. d. Verf.), nicht gewesen. Inwieweit er da heute drinnen ist, kann ich nicht beurteilen, und der hat den Transfer von der Mintzberg'schen Seite her gemacht. Der Melone ist natürlich auch betriebswirtschaftlich geprägt, deswegen lag das auch nahe an der Stelle irgendwo, aber mir würde da schon ein Stück weit das Eingehen auf VKI-Belange fehlen. Das ist ein Aufsatz, wie ich ihn in der Betriebswirtschaftslehre ruck, zuck produzieren kann, mit irgendeiner betriebswirtschaftlichen Theorie im Hintergrund, und einem, naja, einem einfachen Grundverständnis von so einem technischen System. Ja, ich kann als Betriebswirt relativ schnell einen Aufsatz schreiben über Produktionsplanung und Steuerung, realisiert durch intelligente Agenten. Aber dann steht da nix über die Umsetzung drinnen, dann ist das halt ein BWL-Aufsatz und das bleibt ein BWL-Aufsatz. (...) Und bei dem Fox ist es sicherlich deutlich. Ja, der hat (...) zu einem sehr frühen Zeitpunkt in der VKI das Thema Organisation belegt. Er hat natürlich relativ schnell gelernt, dass das, was er formuliert hat, nicht umsetzbar ist, und er hat sich dann auf andere Themen irgendwie rausentwickelt. Er wird aber heute immer noch als Beleg herangezogen (...) Aber das, was er unter Organisation versteht, was Melone unter Organisation versteht, das findet sich in der VKI-Diskussion so im Organisationsbegriff nicht wieder. Wenn wir Chaib-draa z.B. nehmen, im O'Hare/Jennings-Buch dieser Einführungsartikel. Da schreibt er zwei Seiten über Organisation. Da finde ich das, was der Fox und was auch der Melone gemeint haben, inhaltlich nicht wieder. Ja, und ich könnte mir vorstellen, dass Chaib-draa da einfach auch zu weit [weg] ist von den betriebswirtschaftlichen Sachverhalten, um zum Verständnis beizutragen.“ (UP3, 4: 1-36)

Gasser als derjenige der genannten drei Autoren, der von der VKI herkommt, lässt sich nicht genügend auf die strukturtheoretischen Bemühungen der Organisationsforschung ein, um diese für die VKI nutzbar machen zu können – und will dies auch gar nicht –, die anderen beiden, Fox und Malone, die dort genau ansetzen, sind nicht genügend in der VKI verwurzelt, um die tatsächliche Brauchbarkeit der Organisationsmodelle für die VKI zu demonstrieren. Der Transfer scheitert an dem für interdisziplinäre Forschung typischen Problem der wechselseitigen Anschlussfähigkeit heterogener Wissenskulturen.

Es gibt aber auch einen eher inhaltlichen Grund, der das Vorhaben der Modellierung von Multiagentensystemen nach dem Vorbild menschlicher Organisationen nach Auskunft des eben zitierten Gesprächspartners erschwert – jedenfalls dann, wenn man sich nicht für den Idealtyp der strikten Hierarchie interessiert. Wir stoßen hier wieder auf das Problem des umgekehrten Komplexitätsgefälles in menschlichen Gesellschaften und in Gesellschaften technischer Agenten. Es begegnet uns im Folgenden als Problem, dass das Konzept der Rolle angewendet auf menschliche Positionsinhaber in Organisationen ganz andere Implikationen hat als dann, wenn man es zur Modellierung technischer Agenten verwendet:

„Und wenn man jetzt Organisationsmetaphern beispielsweise nimmt (...): Da gibt es schlicht Inkompatibilitäten zwischen den Begriffen. Nehmen wir den Begriff der Rolle zum Beispiel, der in der VKI einfach anders verwendet wird, als man das als Betriebswirt gelernt hat. Als Betriebswirt habe ich halt die Vorstellung, in einem Betrieb eine Rolle zu definieren, die jemand übernimmt per Eintritt in das organisatorische System und sagt: ‚Jawoll, ich übernehme jetzt die Rolle Abteilungsleiter‘. Dazu gehört das und das und das, die ganzen Rollenerwartungen. Und mit diesen Rollenerwartungen wird die entsprechende Person, das Mitglied dieser Organisation dann irgendwie auch verbunden. Ja, wenn man also eine Position anspricht, und sagt: ‚Wer ist zuständig im Betrieb?‘, sagt man: ‚Das ist die und die Stelle und der Stelleninhaber, der muss dann diese Rollenerwartung erfüllen.‘“ (UP3, 1: 10-22)

Die Übertragung dieser Form der Strukturierung von Arbeitszusammenhängen auf Multiagentensysteme

„würde sofort zwei Probleme nach sich ziehen. Das eine ist die Flexibilität des Multiagentensystems, die geht dann kaputt, ja, und die wird gebraucht. Ich habe ja ein Multiagentensystem deshalb, weil ich in einer schwer überschaubaren, schlecht verstandenen Domäne Probleme lösen lassen will, und das heißt: Ich kenne die Suchräume nicht. Das ist ein technisches Problem der VKI. Wenn ich mir also vorstelle, so ein Suchraum ist als Baum zum Beispiel repräsentiert, dann ist das Problem in der VKI oder in der KI generell, dass ich diesen Baum nicht kenne und ich muss ihn durchsuchen lassen. Und wenn ich jetzt eine stabile Rollenerwartung formuliere, heißt das, dass ich einen Agenten nur auf eine bestimmte Art und Weise suchen lassen kann. Damit gehe ich von vornherein die Gefahr ein, dass ich Teile des Baumes unerreichbar mache. Ja, das ist ein ganz anderes Problem, als wenn ich eine Rollenerwartung in einem Betrieb formuliere, weil ich dem Menschen da bestimmte Freiheiten noch lassen kann und weil es dort auch sinnvoll ist, aus allen möglichen Gründen, die mit dem Sozialsystem zusammenhängen, für eine gewisse Stabilität zu sorgen. Damit dieses System als solches stabil bleibt. Ja, und das andere Problem – also das eine war Flexibilität –, das andere ist Autonomie. Das würde die Autonomie der Agenten einschränken. Und die VKI tut sich schwer damit, die Autonomie von einem Agenten einzuschränken, weil, wie gesagt, wir brauchen das gerade für diese flexiblen Suchprozesse. Und wenn man jetzt den Analogieschluss zieht und sagt: ‚Als Mensch trete ich in eine Organisation ein, unterschreibe einen Arbeitsvertrag, trete einen Teil meiner Autonomie dann auch für bestimmte abgegrenzte Zeiten z.B. ab, ja, dann hat das einen Sinn. Ich habe eine Motivationstheorie zum Beispiel, mit der ich sagen kann, gut, wenn mein *benefit* größer ist als das, was ich reinstecke, dann lasse ich das so. Ja, das sind Dinge, die sich so nicht übertragen lassen auf die Agentenszenarien, weil der Agent ja grade als autonomes ..., ich sag mal als autonome Suchmaschine, irgendwie konzipiert worden ist und ich den wirklichen technischen Nutzen von dem System nur habe, wenn ich diese Autonomie auch ausschöpfe.“ (UP3, 1: 50 - 2: 15)

Ich habe diese Äußerung so ausführlich wiedergegeben, weil in ihr der Umkehrungszusammenhang von Autonomie und Strukturbildung besonders prägnant zum Ausdruck kommt: In menschlichen Organisationen ist die Festlegung von Berufsrollen sinnvoll, weil das Hauptproblem bei der Koordination der Akteure darin besteht, die Kontingenz des Handelns zu reduzieren und dadurch eine gewisse Stabilität des Handlungszusammenhanges sicherzustellen. Ausgangspunkt ist hier der Umstand, dass die Akteure beim Eintritt in die Organisation ein beträchtliches Maß eigenständiger Handlungsfähigkeit mitbringen. Die Festlegung einer Berufsrolle schränkt diese Autonomie nur partiell ein, sodass den Organisationsmitgliedern trotz der berufsrollenbezogenen Handlungsrestriktionen genügend Freiräume gelassen werden können, um flexibel auf unerwartete Ereignisse reagieren zu können. Agenten dagegen bringen die Eigenschaft eigenständiger Handlungsfähigkeit nicht von vornherein mit. Werden ihre Handlungsprogramme als Berufsrollen programmiert, dann können sie zunächst nichts anderes, als die in der Rollenbeschreibung festgelegten Handlungsmuster zu aktivieren. Sie können sich dann aber nur bei solchen Problemstellungen angemessen verhalten, hinsichtlich derer die erforderlichen Handlungsbeiträge dieses Agenten in vorgeplanter Weise festgelegt worden sind. Der flexible Umgang mit unvorhergesehenen oder unvorhersehbaren Problemen ist auf diese Weise nicht möglich, die Autonomie der Agenten ist dementsprechend gering.

Ausgehend davon, dass es die Aufgabe von KI und VKI ist, technische Lösungen für schlecht strukturierte Probleme zu erarbeiten, was in der VKI insbesondere durch das Merkmal der Autonomie der Agenten möglich wird, markiert der hier zitierte Gesprächspartner das Vorgehen bei der betrieblichen Organisation von Arbeitszusammenhängen und bei der Konstruktion von Multiagentensystemen als einander geradezu entgegengesetzt:

„Wollen wir aus der Gruppe heraus so einen Formenmechanismus haben, der den einzelnen Agenten dann deformiert in irgendeiner Weise? (...) Natürlich, als Organisationsmann würde ich sagen, klar, muss ich haben. Natürlich, ich will ja, wenn ich nun einen Etat an die Investition binde, dann will ich auch sicher sein, dass der das macht, was diese Organisation interessiert. Und wenn die Organisation ihre Ziele verändert, ja bitte sehr, dann hat der Agent Pech gehabt, dann wird er sich darauf einstellen müssen. Und aus einer VKI-agentenorientierten Perspektive würde ich sagen: Moment, das ist ein autonomes System und ob er das lernen will, das muss ihm selber überlassen bleiben. Ja, und da denke ich, sind die Ansprüche auch sehr unterschiedlich. (...) Aus betriebswirtschaftlicher Sicht würde es dann vielleicht auch so sein, dass wir sagen, die Diskussion wollen wir gar nicht haben, weil wir ja eine bestimmte betriebliche Funktion realisieren wollen, die muss dieser Agent erfüllen, und wenn es die Funktion nicht mehr gibt, dann kommt er ins Archiv. (...) Man muss da wirklich sagen, aus welcher Perspektive man sich das anguckt.“ (UP3, 16: 8-22)

Die Notwendigkeit einer derart konfrontativen Diskussion der Organisation menschlicher und agentengestützter Arbeitszusammenhänge leuchtet allerdings nicht recht ein, jedenfalls dann nicht, wenn man auch von Multiagentensystemen erwartet, dass sie sich nicht in beliebiger Weise verhalten, sondern zumindest mit Blick auf bestimmte übergeordnete Problemstellungen, die vorgegeben sind, ein vorhersehbares Verhalten aufweisen – wie selbstständig auch immer der Weg der Problemlösung innerhalb des Multiagentensystems zu Stande kommt. Die Vorgabe von Strukturmerkmalen und die Eröffnung von Handlungsspielräumen stellt sich dann nicht als exklusive Alternative dar. Erforderlich ist es vielmehr, ein geeignetes Mischungsverhältnis zu finden:

„Ich möchte natürlich eine gewisse Toleranz zulassen, aber ich bin auch immer bestrebt, diese Toleranz irgendwie zu minimieren. Also zulassen und trotzdem einschränken, das ist so die Gradwanderung. Und, ja, letztlich kann ich das ganz hervorragend, indem ich einfach Modelle aus der Soziologie, Gesellschaftsorganisation, übernehme und versuche, sie auf meine Agenten draufzupropfen, um sie ein bisschen mehr vorhersehbarer agieren zu lassen. Aber eben trotzdem nicht sämtliche Freiheiten zu nehmen. Also nicht zu sagen: ‚Hier habe ich eine hierarchische Organisationsform, bei dieser oder jener Aufgabe funktioniert die perfekt und deshalb übertrage ich die auf meine Agenten und dort muss das genauso perfekt ablaufen.‘“ (UW4, 12: 47-58)

In der Tat kann man davon ausgehen, dass ein gewisses Maß an vorgegebener Arbeitsteilung und Funktionsspezialisierung ein wesentlicher Bestandteil sehr vieler Multiagentensysteme ist. Dass damit ein Vorgehen gewählt wird, das in bestimmten Aspekten dem der formalen Strukturierung von Organisationen ähnelt, wird allerdings offensichtlich nur von wenigen VKI-Forschern ausdrücklich wahrgenommen. Und dort, wo diese Strukturähnlichkeit thematisiert wird, geschieht dies häufiger unter Rekurs auf die eigene Alltagserfahrung mit der Organisation von Arbeitszusammenhängen als unter Bezugnahme auf organisationswissenschaftliche Forschungen, sodass die Organisationsforschung als Ideengeber für die Strukturierung von Multiagentensystemen im Ergebnis eine erstaunlich geringe Rolle spielt:

„Das ist, glaube ich, einer der jedenfalls für uns gegebenen Vorteile, dass man viele Überlegungen aus der täglichen Arbeitsorganisation mindestens auf ein solches Agentensystem übertragen kann, bei dem nicht die Funktionsspezialisierung einzelner Agenten emergiert, sondern die sozusagen so ein bisschen per Design vorgegeben ist, also mit Zuständigkeiten für spezielle Aufgaben. Und dann habe ich mir überlegt, ich habe mich schon mein Leben lang mit Wissensorganisation beschäftigt, ich versuche, meinen Bereich hier gut zu organisieren, was ja eher so eine Arbeitsorganisation, das Miteinander, ist. Und man weiß auch wie schön das ist, dass man Leute hat, auf die man sich einfach blind verlassen kann, das heißt eine übertragene Aufgabe wird erledigt. Man weiß auch, dass man sich in manchen Fällen vielleicht Alternativen von Lösungen vorlegen lässt und in manchen Fällen, dass man besser nachhakt und sich die Sachen zurückgeben lässt. Und alle diese verschiedenen Ideen, wie man

eigentlich in einem Konzert mit mehreren Beteiligten zusammenarbeiten kann, kommen jetzt bei unseren Agentensystemen auch vor, in der Organisation der Problemlösung." (UP2, 1: 46 - 2: 4)

Der Modus machtförmiger Handlungskoordination erschöpft sich, wie bereits angemerkt, nicht in der Strukturierung von Interaktionssequenzen direkt geäußerter Weisung und deren Befolgung. Kodifizierte Anweisungen, etwa in Form von Gesetzen, fallen ebenfalls unter diese Form der Verhaltensabstimmung. Mit dem Vorschlag, Multiagentensysteme mittels sozialer Gesetze zu strukturieren, verhält es sich ähnlich wie mit demjenigen, auf Organisationsstrukturen zu rekurrieren: Es liegt ein einschlägiger Artikel vor, in dem dieser Vorschlag unterbreitet wird und der in diesem Zusammenhang stets zitiert wird (Shoham/Tennenholtz 1992); der Vorschlag wird eher skeptisch beurteilt; aber in der Praxis der Konstruktion von Multiagentensystemen werden eine Vielzahl von Festlegungen getroffen, die sich im Sinne von Gesetzen interpretieren lassen. Und der Grund für die Skepsis ist der gleiche wie gegenüber organisationaler Strukturierung: die Befürchtung einer zu weitgehenden Einschränkung der Autonomie der Agenten:

„Man kann (...) schnell und rudimentär erste Ergebnisse erreichen, wenn man nämlich solche sozialen Gesetze, wie sie auch von Shoham definiert wurden, einfach knallhart in die Systeme reinprogrammiert und damit Verhaltensweisen direkt erzwingt. Damit schränkt man natürlich diese Autonomie und diese *pro-activeness* ganz stark ein.“ (UW4, 10: 45-49)

Wiederum bleibt bei dieser Form von Kritik unerkannt, dass die Festlegung kollektiv verbindlicher Regeln, also ein Vorgehen, das mit der Etablierung sozialer Gesetze in menschlichen Zusammenhängen eine gewisse Strukturähnlichkeit hat, durchaus gängige Praxis bei der Konstruktion von Multiagentensystemen ist. So betont einer unserer Gesprächspartner:

„Kooperation zwischen Software-Agenten: Also erstmal unabhängig jetzt davon, dass wir die technischen Details klären müssen. Wissensaustauschformat, Wissenskommunikationssprache. Wir brauchen, ja, hatte ich vorhin schon gesagt, soziale Regularien. Oder streichen wir das Wort sozial: Wir brauchen Verhandlungsprotokolle, Regularien. Und das wird sehr stark von der Anwendung abhängen, ja, denn wir werden kaum ein Multiagentensystem einfach nur so im luftleeren Raum stehen haben. Das wird immer irgendwie eingesetzt werden.“ (UW2, 11: 52- 12: 6)

Ein interessanter Punkt in dieser Äußerung ist, dass der Gesprächspartner die Strukturähnlichkeit mit sozialen Gesetzen einerseits zu sehen scheint, dann aber das Attribut „sozial“ zurücknimmt und lieber neutraler von Verhandlungsprotokollen oder Regularien spricht. Dass es erforderlich ist, Rahmenbedingungen zu setzen, an denen sich die Agenten orientieren müssen, steht für ihn außer Frage. Dies ist sicherlich ein verallgemeinerbarer Tatbestand. Man wird kein Multiagentensystem finden, das keine impliziten oder expliziten Regelungen vorsieht, durch die bestimmte Restriktionen der Verhaltensmöglichkeiten der Agenten eingeführt werden. In diesem Sinne geht die Kritik an der Nachbildung sozialer Gesetze, die mit der damit verbundenen Einschränkung der Agenten argumentiert, an der Praxis der Konstruktion von Multiagentensystemen vorbei.

Zu Recht kann man allerdings fragen, ob solche, für die Agenten eines Multiagentensysteme verbindlichen Regelungen tatsächlich die Analogie zu Gesetzen in menschlichen Zusammenhängen bilden. Versteht man Gesetze als Weisungen, die befolgt

werden, weil der betreffende Akteur den Gehorsam für weniger vermeidenswert hält als die im Fall des Ungehorsams angedrohte Strafe, so wird man in den meisten Fällen eher verneinen müssen, dass Regularien in Multiagentensystemen mit Gesetzen vergleichbar sind. Solange entsprechende Regelungen in Multiagentensystemen gleichsam automatisch zu ihrer Befolgung führen und die Agenten nicht die Möglichkeit haben, die Vor- und Nachteile ihres Gehorsams oder Ungehorsams gegeneinander abzuwägen, liegt keine mit menschlichen Handlungsvollzügen vergleichbare soziale Situation vor. Dem eben zitierten Gesprächspartner ist mit seinem Vorbehalt gegen das Attribut „sozial“ also zuzustimmen.

Die entscheidende Differenz ist hier wieder die zwischen geschlossenen und offenen Multiagentensystemen: Solange die Konstruktion des fraglichen Multiagentensystems in den Händen eines Entwicklers bzw. eines Entwicklerteams liegt, gibt es zunächst wenig Grund, den Agenten die Möglichkeit des Ungehorsams mitzugeben, um dann anschließend eine Wahrnehmung dieser Möglichkeit durch Implementation von Sanktionsmechanismen wieder zu verhindern. Vielmehr kann von vornherein sichergestellt werden, dass die beteiligten Agenten nicht anders agieren können, als die erforderlichen Rahmenbedingungen zu befolgen. Erst in offenen Multiagentensystemen, deren Agenten von denjenigen, die ein entsprechendes Interesse an der Durchsetzung kollektiv verbindlicher Regeln haben, nicht qua Programmierung kontrolliert werden können, ergibt die Einrichtung von Sanktionsinstanzen, die ein bestimmtes Fehlverhalten bestrafen (oder ein bestimmtes Wohlverhalten prämiieren), einen Sinn.

Im Zusammenhang mit dem Problem der Ausschaltung opportunistischen Verhaltens in offenen Kontraktnetz-Architekturen sind wir auf dieses Problem bereits gestoßen. Es steht zu erwarten, dass die agententechnische Realisierung sozialer Gesetze mit der Zunahme der Bedeutung offener Multiagentensysteme an Gewicht gewinnen wird. Die soziale Komplexität von Multiagentensystemen nimmt dadurch in zweifacher Hinsicht zu: Zum einen werden, wie schon angesprochen, solche Formen der Durchsetzung kollektiv verbindlicher Entscheidungen nur dann den gewünschten Effekt haben, wenn die Sanktionen auf die Benutzer der bestrafte oder belohnte Agenten durchschlagen. Dies wird vielfach aber erst dann der Fall sein, wenn die entsprechenden Zusammenhänge als hybride Systeme organisiert sind, die Regeln des Zusammenlebens also die Agenten und deren Benutzer betreffen. Zum anderen wird auch die soziale Situation der Agenten selbst komplexer, weil sie sich dann gleichzeitig an mehr als einem Koordinationsmechanismus orientieren müssen, also etwa an den Regeln des Markttausches und zugleich an den Gesetzen, die die Rahmenbedingungen festlegen, innerhalb dessen Tauschhandlungen zulässig sind (vgl. Schulz-Schaeffer 1998).

### **4.3 Einfluss**

Als Einfluss kann man mit Parsons (1980) dasjenige Interaktionsmedium bezeichnen, in dem die Interaktion vermittelt wird durch die Erwartung, dass bestimmte Akteure bezogen auf bestimmte Problemstellungen über das bessere Wissen verfügen oder aus anderen Gründen die kompetenteren Akteure sind, sodass es vorteilhaft ist,



ihrer Sicht der Dinge zu folgen. So wie Macht nicht identisch ist mit Gewalt, sondern zur Erlangung von Gehorsam zumeist die glaubhafte Androhung von Zwangsmaßnahmen ausreicht, ist Einfluss nicht identisch mit dem konkreten Aufweis überlegener Fertigkeiten, sondern beruht auf der glaubhaften Behauptung von Expertise, wie sie sich in der Reputation der entsprechenden Person ausdrückt. Auf Grund von Einfluss die Vorschläge eines anderen Akteurs zu übernehmen, bedeutet, auf seine überlegenen Fertigkeiten zu vertrauen. Genau hierin besteht die Entlastungsfunktion des Mediums Einfluss. Denn anderenfalls müsste jeder, der eine bestimmte Expertise benötigt, diese selbst erwerben, um das fragliche Problem entweder selbst zu lösen oder um herauszufinden, welche der Personen, die sich als Experten geben, tatsächlich Experten sind.

Für die Nutzung von Einfluss als Mechanismus der Koordination in Multiagentensystemen gilt Ähnliches wie für machtförmige Koordination: Weniger noch als dort findet sich ein expliziter Rekurs auf diese Form sozialer Verhaltensabstimmung. Zugleich aber beruht eines der zentralen Argumente für die Vorzüge der Agententechnologie auf einer Form der Interaktion von Agenten, die, fände sie im Bereich menschlicher Sozialität statt, als durch das Medium Einfluss vermittelt thematisiert werden könnte. Aber auch hier kann wieder gefragt werden, ob die Komplexität der sozialen Situation der Agenten in entsprechenden Zusammenhängen nicht zu gering ist, um berechtigterweise von einer Analogie zu einflussvermittelter Koordination zu sprechen.

Das angesprochene Argument identifiziert einen der Vorzüge von Multiagentensystemen gegenüber zentralisierten Problemlösern in dem Merkmal der „decreased communication by transmitting only highlevel partial solutions to other agents rather than raw data to a central site“ (Moulin/Chaib-draa 1996: 5). Hierbei kommen zwei Gesichtspunkte zur Geltung, die zusammengenommen den Vorzug des Problemlösens auf der Grundlage je spezialisierten Expertentums ausmachen.

Viele Probleme, so ein häufig vorgebrachtes Argument der VKI, sind gleichsam „natürlich“ verteilt.

„Attempting to construct solutions in a distributed manner is often a natural extension of the problems themselves“ (Decker 1987: 729; vgl. auch Davis 1980: 42; Lesser/Corkhill 1983a: 9; Shoham/Tennenholtz 1992: 278).

„Bei vielen Problembereichen liegt bereits eine natürlich gegebene Verteilung vor und legt daher eine Anwendung der VKI nahe.“ (Martial 1992a: 8).

Dies gelte insbesondere für räumliche und funktionale Verteilung (vgl. Davis/Smith 1983: 67). Ein Beispiel für räumliche Verteilung ist das Problem der Integration und Interpretation von Daten an verschiedenen Orten aufgestellter Sensoren, etwa bei Verkehrsüberwachungssystemen (vgl. Lesser/Corkhill 1983b). Expertensysteme für medizinische Diagnose sind ein Beispiel für funktionale Verteilung. Hier besteht das Problem darin, auf der Basis unterschiedlichen Expertenwissens, das in einer Anzahl verschiedener, spezialisierter Expertensysteme implementiert ist, zu einer Diagnose zu gelangen (vgl. Durfee et al. 1989: 64).

Unter diesen Bedingungen – das ist der eine Punkt – ist es von Vorteil, Teilprobleme des zu lösenden Gesamtproblems dort zu lösen, wo sie anfallen, also etwa den

räumlich verteilten Sensoren eines Verkehrsüberwachungssystems die Aufbereitung der von ihnen gemessenen Daten zu überlassen oder den Röntgenarzt (bzw. sein Expertensystem) mit der Auswertung der Röntgenaufnahmen zu betrauen. Die Devise ist dabei die, die Probleme jeweils dort zu lösen, wo die Kompetenz am größten ist, und das heißt unter anderem auch: dort, wo der direkteste Zugang zu den entsprechenden Daten besteht. Zugleich kann auf diese Weise der Aufwand an Datenübermittlung und die Anzahl der Übermittlungsfehler reduziert werden. Der andere Punkt ist, dass sich auf diese Weise das Wissen, das jeder einzelne Agent besitzen muss, um einen sinnvollen Beitrag zur Problemlösung leisten zu können, wesentlich verringern lässt: „(E)in Agent muss nicht über alles Kontrolle haben, sondern nur über seinen lokalen Bereich.“ (IW1, 11: 36f.) Jeder der beteiligten Agenten ist mithin nur bezogen auf einen kleinen Teilbereich Experte und in allen anderen Bereichen Laie und muss sich als solcher auf die Expertise anderer Agenten verlassen. Dies entlastet auf der einen Seite jeden einzelnen Agenten und ermöglicht auf der anderen Seite die Lösung von Problemstellungen, die sehr viel komplexer sind als „was einzelne Teilnehmer überblicken und verantworten können“ (Luhmann 1975a: 72).

Es ist offensichtlich, dass diese Form koordinierten Problemlösens in menschlichen, und insbesondere in modernen Gesellschaften von erheblicher Bedeutung ist. In der Welt, in der wir leben, gibt es kaum einen Handlungsbereich, in dem wir uns nicht beständig auf je spezialisiertes Expertentum verlassen, über das wir nicht selbst verfügen. Giddens bezeichnet solche Kontexte gesonderten Expertentums als Expertensysteme (er meint damit nicht die Expertensysteme der KI-Forschung) und definiert sie als „Systeme technischer Leistungsfähigkeit oder professioneller Sachkenntnis, die weite Bereiche der materiellen und gesellschaftlichen Umfeldler, in denen wir heute leben, prägen.“ (Giddens 1995: 40f.) Erläuternd heißt es:

„Schon allein dadurch, daß ich jetzt in meinem Haus sitze, bin ich in ein Expertensystem oder in eine Reihe von Expertensystemen verstrickt, auf die ich mich verlasse. Ich habe keine sonderliche Angst davor, in meiner Wohnung die Treppe zu benutzen, obwohl ich weiß, daß ein Zusammenstürzen des Gebäudes grundsätzlich möglich ist. Über die vom Architekten und vom Bauunternehmer bei Entwurf und Bauausführung benutzten Wissensbestände bin ich kaum informiert, aber dennoch ‚glaube ich an‘ das, was sie ausgeführt haben. Auf die Tüchtigkeit dieser Personen muß ich zwar vertrauen, doch mein ‚Glaube‘ gilt eigentlich nicht ihnen selbst, sondern der Triftigkeit des von ihnen angewandten Expertenwissens, und das ist etwas, was ich im Regelfall nicht vollständig überprüfen kann.“ (ebd.: 41)

Der entscheidende Gesichtspunkt, der die Differenz von Experten und Laien zu einem Faktor werden lässt, durch den die Koordination von Handlungen in vereinfachender Weise vorstrukturiert wird, besteht in dem Umstand, dass man sich als Laie auf eine von anderer Stelle erbrachte Leistung verlassen kann (und muss), ohne das dort verwendete Wissen und die dort vorhandene Kompetenz selbst beurteilen zu können. Gäbe es keine Anhaltspunkte dafür, dass solches Vertrauen berechtigt ist, dann wäre dies eine riskante Kooperationsstrategie. Einfluss ist das Interaktionsmedium, dessen Aufgabe es ist, solche Anhaltspunkte zu vermitteln: Man sucht beispielsweise am neuen Wohnort einen Zahnarzt, indem man dort ansässige Bekannte fragt, mit welchem Arzt sie gute Erfahrungen gemacht haben; oder man möchte ein neues Auto kaufen und

schreckt instinktiv vor dem Hersteller zurück, dessen Rückrufaktion zuletzt durch die Medien ging; oder aber man sucht einen einschlägigen wissenschaftlichen Artikel zu einem bestimmten Thema und liest zuerst den Text, dessen Autor die höchste Reputation hat.

In allen diesen Beispielen ist der Ruf, den die infrage kommenden Experten oder Expertensysteme genießen, das Kriterium (oder wenigstens eines der Kriterien), das die Wahl zwischen ihnen beeinflusst. Wie im Fall von Geld und Macht wirkt auch hier das Interaktionsmedium – der durch seinen Ruf begründete Einfluss, der einem (individueller oder kollektiver) Akteur zugebilligt wird – in rational berechenbarer Weise als Selektionsvorschlag, der die am Interaktionsgeschehen Beteiligten motiviert, sich in einer bestimmten, tendenziell vorhersehbaren Weise zu verhalten: Den betreffenden Experten muss, sofern sie daran interessiert sind, dass ihre Expertise auch zukünftig nachgefragt wird, daran gelegen sein, eine entsprechende Reputation aufzubauen und zu erhalten. Dies ermöglicht es umgekehrt den Laien, den Ruf der Experten zum Auswahlkriterium zu machen und auf diese Weise hinreichend treffsicher die passende Expertise für ihr Problem finden zu können, ohne die Kompetenz der Experten selbst überprüfen können zu müssen.

Fragt man, in welchem Sinne Einfluss als Mechanismus der Koordination technischer Agenten genutzt wird oder genutzt werden könnte, so stößt man unweigerlich wieder auf die inzwischen altbekannte Differenz zwischen geschlossenen und offenen Multiagentensystemen. Solange die Programmierung der Eigenschaften und Fertigkeiten der Agenten in den Händen eines Entwicklers oder Entwicklerteams liegt, ist Einfluss als Koordinationsmechanismus in den meisten Fällen zu komplex. Man kann den gewünschten Effekt auch einfacher haben, nämlich, indem man den Agenten wechselseitig bekannt gibt, welcher der anderen Agenten als Spezialist für welche Teilaufgabe vorgesehen ist. Die Notwendigkeit der jeweils nicht fachkundigen Agenten, sich ein eigenes Bild über die für sie inhaltlich nicht beurteilbare Kompetenz anderer Agenten zu machen, erübrigt sich damit, das Interaktionsmedium Einfluss ist nicht erforderlich.

Anders ist die Situation auch hier im Fall offener Multiagentensysteme, wobei zwei Formen der Öffnung von Belang sind: Öffnung für neue Mitglieder und Öffnung mit Blick auf Veränderbarkeit der Kompetenzen der beteiligten Agenten. Ich betrachte zunächst den ersten Fall: Anders als in geschlossenen kann man in mitgliedsoffenen Multiagentensystemen nicht voraussetzen, dass es ein globales Wissen darüber gibt, welches jeweils die kompetenten Agenten für die Bearbeitung einer bestimmter Aufgabe sind. Damit entsteht ein Problem, wie es auch für die Handlungsfähigkeit von Menschen in modernen Gesellschaften von einiger Bedeutung ist: Als Laie, der darauf angewiesen ist, auf Expertenleistungen zugreifen zu können, muss man, wie schon angesprochen, Anhaltspunkte für die Wahl eines geeigneten Experten besitzen, in gewisser Weise also Experte der Suche nach Experten sein oder aber es muss „Experten für Ex-

perten“ geben,<sup>12</sup> deren Rat man vertraut (vgl. Berger/Luckmann 1969: 47f.). Agentenkommunikationssprachen wie KQML oder wie der Standardisierungsvorschlag der FIPA versuchen diesem Problem Rechnung zu tragen, indem sie die Möglichkeit der über *broker* vermittelten Suche nach kompetenten Kooperationspartnern, also die Suche mittels Experten für Experten, vorsehen.

Dies verschiebt das Problem allerdings nur, weil man sich nun natürlich fragt, wie diese *broker* zu dem entsprechenden Wissen gelangen. Gegenwärtig scheint die zu meist gewählte Lösung dieses Problems, wie etwa in dem oben besprochenen Reiseplanungssystem, darin zu bestehen, den Broker-Agenten dieses Wissen vorzugeben, also die entsprechende menschliche Expertise in deren Wissensbasis einzugeben. Im Interesse einer weiter gehenden softwaretechnischen Automatisierung von Koordinationsprozessen auf der Grundlage der Differenz von Laien und Experten wäre es allerdings erstrebenswert, die Agenten selbst – oder auch nur einige von ihnen – in die Lage zu versetzen, die Kompetenz als Experten für Experten eigenständig zu entwickeln. In diesem Zusammenhang wäre es nun in der Tat sinnvoll, auf das Interaktionsmedium Einfluss zu rekurrieren. Überlegungen, die in diese Richtung gehen, haben wir in unseren Interviews wie auch in der gesichteten Literatur jedoch bestenfalls ansatzweise gefunden (etwa im Zusammenhang mit der Vertrauensproblematik, über die noch zu reden sein wird; vgl. unten S. 70ff.). An diesem Punkt kann man also nur über die mögliche zukünftige Relevanz eines im Bereich der menschlichen Sozialität einflussreichen Koordinationsmechanismus spekulieren, ohne seine tatsächliche Nutzung in der Agententechnologie bereits konstatieren zu können.

Ein wenig, wenn auch nicht sehr viel konkreter ist die Sache im Fall einer Öffnung von Multiagentensystemen mit Blick auf eine Veränderbarkeit der Kompetenzen von Agenten, in Zusammenhängen also, in denen vorgesehen ist, dass Agenten in der Lage sind zu lernen. Auch hier ergibt sich, sofern die Kooperation darauf beruht, die passende Expertise für das fragliche Teilproblem zu finden, unweigerlich die Notwendigkeit einer Expertise des Auffindens von Experten. Denn angesichts lernbedingter Veränderung von Kompetenzen kann es hier ebenfalls kein global vorgegbares Wissen darüber geben, welcher Agent für welches Problem der richtige Ansprechpartner ist.

Einschränkend muss allerdings berücksichtigt werden, dass die VKI bei der Konzeption und Realisierung lernfähiger Agenten noch sehr in den Anfängen steckt. Die prinzipielle Bedeutung des Merkmals der Lernfähigkeit wird zwar von nahezu allen unseren Gesprächspartnern herausgestrichen, wobei sich dessen besondere Relevanz aus der Verbindung mit dem Merkmal der Autonomie der Agenten ergibt:

„Die Fähigkeit zu lernen und so weiter - ganz, ganz wichtige Sache bei Agenten. Wenn man das Autonome in den Vordergrund stellt, einen Agent als autonomes System [betrachtet], der in einer offenen Umgebung lebt, das heißt, in keinem geschlossenen System, sondern in einem System, wo sich kontinuierlich Sachen ändern, wo neue Systeme dazukommen, wo, ja, sich irgendwie über die Zeit auch Organisationsstrukturen ändern, wo sich die Spielregeln ändern, nach denen die Agenten funktionieren.“ (IW5, 4: 23-28)

---

<sup>12</sup> Ein Beispiel eines solchen Experten für Experten ist der Hausarzt, der einen im Bedarfsfall an den zuständigen Facharzt überweist.

Zugleich aber teilt die Mehrheit der von uns Befragten die Einschätzung, dass die praktische Bedeutung des Lernens in Multiagentensystemen beim gegenwärtigen Stand der VKI eher als gering einzuschätzen ist:

„Also Lernen spielt schon eine Rolle. Es gibt sicherlich genügend Situationen, wo wir es entweder sinnvoll einsetzen können und wo wir es einsetzen würden. Es ist nur so, dass die Lerntechnik, die zur Verfügung steht, meistens nicht ausreicht. Nicht anwendbar. Also dass wir meistens immer Lösungen finden, wie wir das umgehen. Ein Beispiel ist, dass der Benutzer das dann einstellen kann. Es wäre sehr oft einfacher, wenn das System sich selbst einstellen könnte. Nur die Techniken, die mir bekannt sind, sind in der Regel nicht einsetzbar.“ (IW1 13: 35-42)

Die Sicht eines Gesprächspartners, der anmerkt, er wisse „noch keine Anwendung, in der das (d.h. lernfähige Agenten, Anm. d. Verf.) jetzt schon eine Rolle spielt“ (IL1, 23: 33) erscheint nichtsdestotrotz als zu pessimistisch. Immerhin hat die VKI in einem Bereich, dem des Benutzer-Profiling, bereits Einiges in Richtung auf lernfähige Agenten erreicht, auch wenn dieser Bereich sicherlich nicht zu den Kerngebieten der VKI zählt. Benutzer-Profiling zielt darauf, Agenten dazu zu befähigen, zunehmend besser auf die spezifischen Interessen, Wünsche und Eigenarten eines bestimmten Benutzers einzugehen. Im Vordergrund stehen hier die sogenannten persönlichen Assistenz-Agenten („personal assistant agents“), also Agenten, die einem Benutzer zugeordnet sind und – ähnlich einer Sekretärin oder einem Sekretär – in seinem Auftrag bestimmte Aufgaben übernehmen soll: im Internet nach Informationen suchen, den Terminkalender verwalten, als virtueller Ansprechpartner seines Benutzers auf dessen Homepage auftreten u.ä. Benutzer-Profiling erfolgt durch *feedback* und Verstärkung (Lernen durch *reinforcement*), durch Imitation des Benutzers und/oder durch Schlussfolgerungen auf dem für den Agenten sichtbaren Verhalten des Benutzers. Beim Lernen durch *feedback* gibt der Benutzer seinem Agenten jeweils Rückmeldungen über seine Zufriedenheit mit dessen Performanz, aus denen sich der Agent im Laufe der Zeit ein Bild davon zusammensetzt, was sein Benutzer goutiert und was nicht:

„Also beim Lernen in der KI ist es sehr wichtig, dass die Mechanismen *feedback* bekommen müssen. Sie müssen *reinforcement* bekommen, also sie müssen Informationen darüber bekommen: ‚Wie gut war jetzt die letzte Aktion, die ich gemacht habe?‘“ (FW3, 11: 19-21)

Beim Lernen durch Beobachtung des Benutzers übernimmt der elektronische Assistent eine aktivere Rolle. Die Agentin „Letizia“ etwa beobachtet, für welche Internet-Seiten ihr Benutzer auf seinem Rechner einen *link* abspeichert, auf welchen Seiten er sich wie lange aufhält usw., entwickelt daraus Hypothesen, welche Informationen ihn interessieren könnten und verwendet dieses Wissen bei eigenständigen Internetrecherchen (vgl. Lieberman 1995).

Weiter oben hatte ich einen Anwendungsfall des Benutzer-Profiling angesprochen, in dem unterschiedliche Agenten um die zutreffende Interpretation des deiktischen Sprachgebrauchs eines Benutzers konkurrieren (also etwa um richtige Beantwortung der Frage, welche Richtung der Benutzer meint, wenn er befiehlt, einen auf dem Bildschirm repräsentierten Gegenstand nach links zu bewegen). Von den mir bekannten Multiagentensystemen ist dies, zumindest innerhalb der deutschen VKI, der Anwendungsfall, der am deutlichsten in Richtung einer über das Medium des Einflusses ge-

steuerten Form der Verhaltensabstimmung geht. Dieses Multiagentensystem läuft auf der Grundlage der Kontraktnetz-Architektur, der Gegenwert, den ein Agent für seine Leistung erhält, ist in diesem Fall aber nicht eine Art virtuelles Geld, sondern ein „credit value“ (Lenzmann/Wachsmuth 1997: 209), also ein Wert, der das kumulierte Maß der Zufriedenheit und/oder Unzufriedenheit der Nutzer dieses Agenten mit seinen Expertenleistungen in vergangenen Kooperationen ausdrückt. Dieser „credit value“ erhöht oder verringert die Chance, dass der betreffende Agent bei der zukünftigen Suche nach einem Experten ausgewählt wird, er repräsentiert also nichts anderes als die Reputation des Agenten.

Hier finden wir mithin einen softwaretechnisch implementierten Mechanismus der Etablierung von Kooperationen auf der Grundlage der Differenz von Experten und Laien, der in analoger Weise zur Nutzung des Interaktionsmediums Einfluss in menschlichen Sozialbeziehungen strukturiert ist. Die Autoren merken an, ihre Vorgehensweise sei „motivated by the way tasks would normally be assigned in a company“ (ebd.: 208f.). Nun würden Organisationsforscher sicherlich bestreiten, dass Einfluss der dominante Modus der Verhaltensabstimmung in Unternehmen ist und auf den zumindest formal zumeist vorgeordneten Modus der Macht verweisen. Immerhin aber gibt es durchaus Organisationstypen, in denen Handlungsabstimmung mittels Reputation eine bedeutende Rolle spielt (vgl. Mintzberg 1989: 183ff., 205ff.), wobei das Paradebeispiel dann regelmäßig die Universität ist. Man kann vermuten, dass auch die Autoren primär die Organisation der Universität im Blick hatten. Worauf ich mit dem Zitat hinweisen wollte, ist, dass sie ausdrücklich beanspruchen, auf einen Koordinationsmechanismus zwischen menschlichen Akteuren zu reflektieren. Sie nehmen ihn jedoch nicht in seiner allgemeinen Form wahr (nämlich als Interaktionsmedium Einfluss), sondern in seiner speziellen Form als Koordinationsmechanismus innerhalb organisationaler Zusammenhänge.

Betrachtet man Einfluss allein als Modus der Verhaltensabstimmung innerhalb organisierter Sozialbeziehungen, dann gehen bestimmte Aspekte der Leistungsfähigkeit dieses Interaktionsmediums verloren, die insbesondere aus der Möglichkeit der Generalisierung von Einschätzungen über die Reichweite konkreter Interaktionsbeziehungen hinaus resultieren. In dem eben skizzierten Multiagentensystem ist es so, dass der jeweilige Ruf eines Agenten ausschließlich auf der Ebene dyadischer Interaktionsbeziehungen organisiert ist: „Contractors store different credits for different contractor-manager relationships to distinguish between managers.“ (Lenzmann/Wachsmuth 1997: 208) Die Annahme die hier zu Grunde liegt, ist, dass die nachgefragte Expertise eines Agenten (des „contractors“) aus der Sicht unterschiedlicher Nachfrager (der „manager“) von unterschiedlichem Wert sein kann. So legt beispielsweise der eine Benutzer, wenn er die Anweisung gibt: „Rücke diesen Gegenstand nach links!“ seine eigene räumliche Perspektive zu Grunde, ein anderer dagegen die auf dem Bildschirm repräsentierte. Sofern das Expertentum der Agenten darin besteht, jeweils eine der beiden Interpretationen anzubieten, sind es also unterschiedliche Agenten, die die zutreffende Expertise vorhalten.

Der „credit value“ ist mithin das Maß an Wertschätzung der Expertise eines Agenten seitens eines anderen Agenten bzw. Akteurs. Es ist klar, dass die Reichweite des Interaktionsmediums Einfluss, wird es in dieser Weise implementiert, in zweifacher Hinsicht begrenzt bleibt: Zum einen nimmt mit zunehmender Zahl der Agenten, die sich über diesen Mechanismus koordinieren, die Anzahl der erforderlichen Reputationswerte stark anwachsend zu, das Verfahren ist also nur bei geringer Teilnehmerzahl praktikabel. Zum anderen fehlt die Möglichkeit einer Objektivierung von Reputations einschätzungen, also die Möglichkeit, deren Verlässlichkeit dadurch zu erhöhen, dass die Einschätzungen mehrerer und möglicherweise sehr vieler Agenten zusammenfließen.

In gewissem Sinne repräsentiert diese Form der Nutzung von Reputation als Koordinationsmechanismus also eher den Modus der Vergemeinschaftung als den der Vergessenschaftung. Sie gründet wesentlich auf der zunehmenden Bekanntschaft zwischen Agenten, das Merkmal der symbolischen Generalisierung von Verhaltenserwartungen über die je spezifische Historie einer dyadischen Interaktionsbeziehung hinaus fehlt dagegen. Genau dieser Aspekt der Generalisierung von Erwartungen hinsichtlich der Kompetenz eines Experten und ihre symbolische Repräsentation als Reputation machen dagegen die besondere Leistungsfähigkeit des Interaktionsmediums Einfluss im Bereich menschlicher Sozialität aus.

Wenn ich weiter oben (auf S. 27) beispielsweise von einer „bekannten Unterscheidung“ spreche, „die Weber ... aufgestellt hat“, so impliziere ich damit, dass alle (soziologischen) Leser dieses Texte erstens wissen, dass nur einer der vielen soziologischen Autoren namens Weber gemeint sein kann, nämlich derjenige mit dem Vornamen Max, und dass sie zweitens akzeptieren, dass die angesprochene Unterscheidung bedenkenswert ist, und zwar zunächst einzig aus dem Grund, weil sie jener Max Weber aufgestellt hat. Ich recurriere mit anderen Worten auf die besondere Reputation von Max Weber als eines der Begründer der Soziologie als wissenschaftlicher Disziplin, indem ich zugleich voraussetze, dass diese Reputation mehr ist als eine persönliche Wertschätzung der soziologischen Kompetenz Webers durch je einzelne Rezipienten seiner Überlegungen, nämlich Ausdruck einer Einstellung, die alle ausgebildeten Soziologen zumindest im Grundsatz teilen. Nicht viel anders ist die Situation, wenn ich mir einen Zahnarzt empfehlen lasse: Wertvoll wird die mir mitgeteilte Wertschätzung eines bestimmten Zahnarztes erst unter der Voraussetzung, dass sie eine verallgemeinerbare Kompetenz zum Ausdruck bringt, der Zahnarzt also nicht nur die Zähne der ihn Empfehlenden erfolgreich repariert, sondern erwartbar auch mit den meinen so verfahren wird.

Eine Multiagenten-Simulation von Castelfranchi et al. (1998) illustriert diesen Zusammenhang zwischen dem Ausmaß erforderlicher Bekanntschaft und der Leistungsfähigkeit von Reputation als Koordinationsmechanismus recht deutlich. In dieser Studie wird Reputation als Mechanismus der wechselseitigen Einschätzung einer bestimmten Normkonformität von Agenten genutzt. Vorausgesetzt ist ein Handlungsfeld, in dem Agenten unter bestimmten Umständen interagieren, wobei sie sich in der Inter-

aktion dann mit Blick auf eine vorgegebene Handlungsnorm konform oder nicht konform verhalten können. Dabei ist normkonformes Verhalten nur dann von Vorteil, wenn sich auch der betreffende andere Agent normkonform verhält. Die vermutete Haltung der anderen Agenten wird wiederum mittels eines Reputationswertes abgeschätzt. Verglichen werden zwei Formen des Erwerbs von Reputationswissen. Im einen Fall ist die Reputationseinschätzung, die jeder Agent bezogen auf andere Agenten trifft, ausschließlich ein Resultat direkter Interaktion mit den fraglichen anderen Agenten. Im anderen Fall erfolgt die Reputationsbildung unter Einbeziehung von Auskünften von Agenten über Agenten.

Der interessante, wenn auch nicht überraschende Befund der Multiagenten-Simulation lautet: Solange jeder Agent seine Einschätzung der Reputation der anderen Agenten ausschließlich aus eigenen Interaktionserfahrungen ableitet, ist die Wirkung des Reputationsmechanismus als Verfahren der Abwehr opportunistischen Verhaltens vergleichsweise gering. Denn im ungünstigsten Fall hat sich ein Agent bei der Einschätzung aller anderen Agenten je einmal geirrt – und die Kosten des Irrtums tragen müssen – bevor er deren Reputation zutreffend einschätzen kann. Dagegen greift der Reputationsmechanismus sehr viel schneller und wirkungsvoller, wenn Einschätzungen kommuniziert werden können. Nun eilt den Agenten ihr guter oder schlechter Ruf voraus. Es muss also nicht jeder Agent selbst erst die entsprechenden guten oder schlechten Erfahrungen sammeln, um seinen Interaktionspartner einschätzen zu können.

Diese Multiagenten-Simulation bestätigt nur einen Umstand, der uns im alltäglichen Leben allen geläufig ist: Unter Zweckmäßigkeitgesichtspunkten ist die Wahl eines Kooperationspartners auf der Grundlage eigener Erfahrungen mit jedem potentiellen Kandidaten vielfach zu kostenträchtig. Man bezieht deshalb häufig die Erfahrungen anderer mit ein, was den Effekt hat, dass sich die Kosten des Erfahrungen Sammelns auf eine größere Zahl von Akteuren verteilen. Dort, wo sich Akteure in dieser Weise auf Reputation beziehen, setzt in doppelter Weise eine Bewegung ein, die weg vom vergemeinschafteten und hin zum vergesellschafteten Handeln geht. Die eine Konsequenz ist, dass Reputationseinschätzungen nun auch unabhängig von der persönlichen Bekanntschaft mit dem Träger der Reputation verwendet werden können (auch wenn weiterhin aus persönlicher Bekanntschaft resultieren). Damit zusammenhängend wird zweitens die Reputationseinschätzung zu einer zunehmend verallgemeinerten Aussage über den eingeschätzten Akteur.

Zur Erläuterung des zweiten Aspektes ist darauf zu verweisen, dass die Wirksamkeit des Reputationsmechanismus bei der Handlungskoordination stets auf einer Verallgemeinerung beruht. Das Mindestmaß an Verallgemeinerung lautet: „Weil ich den betreffenden Akteur in der Vergangenheit als einen kennen gelernt habe, der die und die Verhaltenseigenschaften gezeigt hat, vermute ich, dass er sich auch in Zukunft auf diese Weise verhalten wird.“ Es liegt auf der Hand, dass die auf Koordinationsprozesse bezogene Leistungsfähigkeit entsprechender Einschätzungen in dem Maße zunimmt, in dem die Erwartungen, die sie ausdrücken übertragbar werden. Ich habe dies am Beispiel der Wahl des Zahnarztes eben erläutert. Der Grad der Kumulation von Einzeler-



fahrungen dürfte ein wesentlicher Faktor zunehmender Übertragbarkeit von Reputationseinschätzungen sein: Wenn man sich bei der Wahl des Zahnarztes auf den Rat eines Bekannten verlässt, hat man das Risiko, dass in dessen Einschätzung individuelle Vorlieben oder Aversionen eingeflossen sind, die man nicht teilt. Verfügt man dagegen über den übereinstimmenden Rat mehrerer, nimmt dieses Risiko ab. Bereits auf der hier erreichten Stufe der Verallgemeinerung von Reputationseinschätzungen – und damit komme ich zu dem zuerst genannten Aspekt – wird es möglich, Kooperationsbeziehungen erfolgreich einzugehen, ohne dass ein vorgängiges Bekanntschaftsverhältnis vorausgesetzt ist. Dies ist allerdings insofern nur die halbe Wahrheit, weil zugleich ein Netz von Bekanntschaftsbeziehungen vorausgesetzt ist, mittels dessen Einschätzungen des persönlich unbekanntem Anderen (einseitig oder wechselseitig) verfügbar werden.

Von Reputation als vergesellschafteter Form der Handlungskoordination kann im Vollsinn dagegen erst gesprochen werden, wenn Reputationseinschätzungen über Netzwerke von Bekannten hinaus verallgemeinert sind. Dies setzt voraus, dass sich eigenständige gesellschaftliche Institutionen ausprägen, über die Reputation verteilt und entzogen wird. Beispiele dafür sind: die Verbraucherschutzorganisation, die Untersuchungen über die Qualität konkurrierender Waren und Dienstleistungen veröffentlicht, oder die wissenschaftliche Fachzeitschrift, in der zu veröffentlichen die Reputation der Wissenschaftler steigert, denen dies gelingt. Der entscheidende Unterschied zur Reputationsbildung über Bekanntschaftsnetzwerke besteht darin, dass die reputationsbildenden Institutionen selbst eine generalisierte Reputation besitzen, die bei der Vergabe von Reputation eingesetzt wird und von der abhängig ist, in welchem Maße Reputation vergeben werden kann: die Reputation der Verbraucherschutzorganisation etwa, zutreffende Produktbewertungen anzufertigen, die sich allein am Interesse der Verbraucher orientieren, oder die Reputation als führende Fachzeitschrift einer wissenschaftlichen Disziplin, die die Reputierlichkeit der Tatsache, in ihr publiziert zu haben, bestimmt. Der Grad der Generalisierung von Reputationswerten, die innerhalb von Bekanntschaftsnetzen zirkulieren, dürfte dagegen in der Regel geringer sein, wenn auch in Einzelfällen „Flüsterpropaganda“ eine Breitenwirkung haben kann, die sie eher als ein Phänomen der Vergesellschaftung denn als eines der Vergemeinschaftung erscheinen lässt (man denke etwa an die Verbreitung des Wissens über neu aufkommende wissenschaftliche Moden und über „angesagte“ Autoren, an die Verbreitung jugendkultureller Stile u.ä.).

Insgesamt sieht man, dass die besondere Leistungsfähigkeit des Interaktionsmediums Einfluss – wie die der anderen Interaktionsmedien – in seiner Eignung zur Strukturierung vergesellschafteter Austauschbeziehungen besteht, solcher also, die auf der Grundlage rationaler Berechnung zwischen Akteuren aufgebaut werden, welche wenig anderes voneinander wissen müssen als dies, dass sie bestimmte Rationalitätsannahmen teilen und sich ihnen entsprechend verhalten. In diesem Fall also der Annahme, dass die Reputation eines Experten ein hinreichend verlässliches Kriterium seiner Auswahl ist, weil sie eine objektivierte Einschätzung der beobachteten Performanz dieses Experten repräsentiert, in der die erfolgreiche Ausübung von Expertise positiv

und vorgegaukelte Expertise (Scharlatanerie) negativ zu Buche schlägt. Eine besonders interessante Beobachtung ist dabei, dass einflussvermittelte Interaktion offensichtlich über einen Selbstschutzmechanismus gegen opportunistisches Verhalten verfügt, den es im Fall geldvermittelter Interaktion beispielsweise nicht gibt. Während im Fall des Markttausches opportunistisches Handeln (also etwa das des Betrügers) nur unter Rückgriff auf machtbasierte Strukturen begrenzt werden kann, bestraft sich der Scharlatan gleichsam selbst, indem er die eigene Reputation untergräbt.<sup>13</sup>

Vor dem Hintergrund der bereits angesprochenen Schwierigkeit dafür zu sorgen, dass Sanktionsmechanismen nicht wirkungslos verpuffen, weil sie die Benutzer opportunistischer Agenten nicht treffen, ist die Beobachtung des Selbstschutzmechanismus des Mediums Einfluss von erheblicher Bedeutung. Als eine Form der Einrichtung von Sanktionsmechanismen auf der Ebene des Multiagentensystems, die gleichsam automatisch auf die Benutzer durchschlägt, hatte ich oben (auf S. 20f.) die Möglichkeit angeführt, eine wechselseitige Bewertung der Vertrauenswürdigkeit von Agenten in diesem Sinne zu verwenden. Wie in den eben besprochenen Multiagentensystemen beruhen aber auch solche Bestrebungen wesentlich auf kumulierter Bekanntschaft, was deren Reichweite als Koordinationsmechanismus begrenzt. Verallgemeinerte Reputation bietet hier einen interessanten Ausweg, indem sie einerseits in genau dem erforderlichen Sinne selbstsanktionierend wirkt, andererseits aber im Modus der Vergesellschaftung eingesetzt werden kann. Meines Wissens ist diese Option der Nutzung des Interaktionsmediums Einfluss in der VKI jedoch bislang nicht wahrgenommen worden.

#### **4.4 Wertbindung**

Schließlich wirken auch gemeinsam geteilte Werte als Medium der Strukturierung von Interaktionszusammenhängen zu weitgehend vorhersehbaren Handlungsketten. Werte symbolisieren die Erwartung, dass es bestimmte Einstellungen oder Handlungen gibt, die die meisten Mitglieder der Gesellschaft befürworten bzw. ablehnen (sollten), weil es normativ richtig bzw. falsch ist so zu handeln. Anders als bei den drei anderen genannten Interaktionsmedium ist die Orientierung an durch Werte vorgegebenen Selektionsvorschlägen zunächst nicht durch die Mitnahme positiver oder die Vermeidung negativer Sanktionen motiviert, sondern durch den Eigenwert des Wertes als Ausdruck dessen, was richtig und was falsch ist. Ohne diese Erwartung wäre es sehr viel schwieriger, Handlungen zu koordinieren, die für die Beteiligten nicht unmittelbar mit der Realisierung von Vorteilen oder der Vermeidung von Nachteilen verbunden sind.

Mittelbar greift allerdings auch hier ein Sanktionsmechanismus: Wie Gesetze zielen auch Werte auf die Festlegung kollektiv verbindlichen Handelns, im Unterschied zu machtbasierter Strukturierung besteht der korrespondierende Sanktionsmechanismus hier jedoch nicht in monetärer oder physischer Bestrafung bzw. Belohnung, sondern in der sozialen Missbilligung des Fehlverhaltens und der sozialen Anerkennung des

---

<sup>13</sup> Es versteht sich von selbst, dass dies eine idealtypische Betrachtung ist. Im empirischen Einzelfall können eine Vielzahl von Faktoren bewirken, dass der Rückkopplungszusammenhang von erfolgreich ausgeübter Expertise und Reputationserwerb gestört ist.

Wohlverhaltens. Sofern Missbilligung zum Ausschluss aus und Anerkennung zur Aufnahme in soziale Zusammenhänge(n) führt, deren Mitgliedschaft für einen wichtig oder erstrebenswert ist, erweisen sich Anerkennung und Missbilligung als wirkungsvolle Sanktionen (ein wunderschönes literarisches Beispiel hierfür ist Sinclair Lewis' *Babbitt*).

Die VKI hat, insbesondere in ihren Anfängen, einen Koordinationsmechanismus eingesetzt, der sich als eine Form von Wertbindung interpretieren lässt und der, bevor er in den letzten Jahren ein wenig in Verruf gekommen ist, von ähnlicher Bedeutung war wie die Kontraktnetz-Architektur: die Gutwilligkeitsannahme („benevolent agent assumption“). Die frühesten VKI-Ansätze versuchen das Koordinationsproblem zu lösen, indem sie von der Situation der Kooperation zwischen bereitwilligen Agenten ausgehen (vgl. Davis 1980: 42). Die kooperierenden Einheiten werden konzipiert als „friendly agents, who wish to do what they are asked to do“ (Morgenstern 1987, zitiert nach Martial 1992b: 41). Die zunehmende Unzufriedenheit mit der Gutwilligkeitsannahme resultiert aus zwei Beobachtungen: der Beobachtung, dass menschliche Akteure sich einander gegenüber keineswegs durchgängig gutwillig verhalten, und der Beobachtung, „that cooperation does not assume benevolence between nodes“ (Durfee et al. 1989: 67). Rosenschein und Genesereth, deren Kritik die Skepsis gegenüber der Gutwilligkeitsannahme ausgelöst hat (vgl. Decker 1987: 731), argumentieren:

„In the real world, agents are not necessarily benevolent in their dealings with one another. Each agent has its own set of desires and goals, and will not necessarily help another agent with information or with actions. Of course, while conflict among agents exist, it is not total. There is often potential for compromise and mutually beneficial activity. Previous work in distributed artificial intelligence, bound to the benevolent agent assumption, has generally been incapable of handling these types of interaction. Intelligent agents capable of interacting even when their goals are not identical would have many uses.“ (Rosenschein/Genesereth 1988: 227)

Die Einschätzung, dass sich Mechanismen der Koordination auf der Grundlage konfligierender oder konkurrierender Interessen und Ziele der beteiligten Agenten für die VKI als besonders wertvoll erweisen könnten, ist durch die Kontraktnetz-Architektur und deren breite Anwendbarkeit eindrucksvoll bestätigt worden. Selbstverständlich ergibt sich daraus nicht der Umkehrschluss, dass die Gutwilligkeitsannahme in allen Fällen eine untaugliche Lösung des fraglichen Koordinationsproblems ist. Dennoch scheint sie in den Ruch des Unstatthaften gekommen zu sein. Es ist jedenfalls auffällig, dass nahezu alle unsere Gesprächspartner darauf verzichten, das Verhalten ihrer Agenten als gutwillig zu beschreiben, obwohl Gutwilligkeit nach wie vor offensichtlich ein wichtiger Koordinationsmechanismus ist:

„Also ich möchte mal so sagen (...), dass man eigentlich immer von so einer Art gutmütigem Agenten, sagen wir mal so, ausgeht. Also nicht wie in der menschlichen Gesellschaft, dass manche Leute nicht arbeiten wollen.“ (UW3, 14: 26-29)

„Man kann aber einfach sagen: ‚Okay Leute, das (Verhaltensabstimmung mittels ausgehandelter Kompromisse, Anm. d. Verf.) ist uns ein bisschen zu abgedreht, weil das einfach eine [unakzeptable] Laufzeit bringt.‘ Ich meine, wenn die [Mitglieder eines Operationsteams im Krankenhaus mit Blick auf jede Handreichung] versuchen, irgendwie zu einem Kompromiss zu kommen, dann ist der Patient schon längst tot. Also muss man sich überlegen, dass man sagt, man geht davon aus, dass einfach alle

Agenten, die immer wieder daran beteiligt sind, von sich aus konstruktiv sind. Oder einfach alles, was sie wissen bereitstellen.“ (UW7, 8: 26-33)

Mit einiger Berechtigung kann man davon ausgehen, dass die Agenten der Mehrzahl der gegenwärtig vorliegenden Multiagentensysteme sich in vielerlei Hinsichten in einer zumeist nicht weiter explizierten Weise gutwillig verhalten, also beispielsweise nicht in der Lage sind, Informationen vorzuhalten, um die sie gebeten werden, Aussagen wider besseres Wissen zu treffen, ihre Problemlösungsfähigkeit strategisch zurückzuhalten o.Ä.

Kann man dies voraussetzen, so fragt sich, ob diese Gutwilligkeit in analoger Weise als Koordinationsmechanismus wirkt wie die mittels Wertbindung erreichten Formen der Verhaltensabstimmung zwischen menschlichen Akteuren. Ein weiteres Mal hängt die Beantwortung dieser Frage davon ab, ob man es mit geschlossenen oder mit offenen Multiagentensystemen zu tun hat. Mit Blick auf geschlossene Multiagentensysteme muss diese Frage mit den gleichen Argumenten verneint werden, die auch im Fall von Macht und Einfluss dazu führten, dass von einer Analogie nicht wirklich die Rede sein kann. Dies lässt sich anhand des folgenden Beispiels illustrieren:

In einem Multiagentensystem von Connah und Wavish (1990) geht es um Kooperationsprozesse zwischen einfachen reaktiven Agenten innerhalb eines simulierten experimentellen Szenarios: Das Szenario besteht aus einem Raum, in dem Möbelstücke stehen, und aus Agenten, denen die Aufgabe vorgegeben ist, je eines der Möbel aus dem Raum zu transportieren. Kooperationserfordernisse ergeben sich dann, wenn ein Möbelstück für einen Agenten allein zu schwer ist. Die Regeln, nach denen sich das Verhalten jedes der Agenten richtet, lauten: (1) Gehe zu dem richtigen Möbelstück. (2) Hebe es an. (3) Wenn es sich anheben lässt, trage es hinaus. (4) Wenn es zu schwer ist, überprüfe, ob ein anderer Agent vorhanden ist, der sich zur Zeit im „Hilfe-Modus“ befindet, d.h. die Bereitschaft äußert, anderen Agenten zu helfen. (5) Ist das der Fall, dann rufe ihn zur Hilfe, trage das Möbelstück mit ihm gemeinsam hinaus und gehe anschließend selbst in den Hilfe-Modus über. (6) Ist dies nicht der Fall, dann gehe sofort in den Hilfe-Modus über. (7) Ist eine Hilfe geleistet worden, aber die eigene Aufgabe noch nicht erfüllt, suche erneut nach einem hilfsbereiten Agenten. Ein Durchlauf der Simulation mit zwei Agenten und zwei schweren Möbelstücken sieht dann etwa folgendermaßen aus: Agent A erreicht sein Möbelstück zuerst, erkennt es als zu schwer, findet keinen hilfsbereiten anderen Agenten und geht in den Hilfe-Modus über. Wenn Agent B sein Möbelstück erreicht, ist Agent A bereits im Hilfe-Modus, sodass Agent B ihn herbeirufen und mit ihm das Möbelstück hinaustragen kann. Anschließend geht Agent A erneut auf die Suche nach Hilfe, und löst seine Aufgabe zusammen mit Agent B, der sich nun im Hilfe-Modus befindet.

Die verwendete Begrifflichkeit legt nahe, dass diese Agenten in ähnlicher Weise hilfsbereit sind, wie menschliche Akteure, die sich etwa dem Wert verpflichtet fühlen, bei kleineren Hilfeleistungen keine Umstände zu machen. Allerdings könnte man den beschriebenen Handlungszusammenhang ebenso gut als machtförmigen Abstimmungsprozess beschreiben, ohne dass sich in der Sache irgendetwas ändern würde. Statt „ein

Agent sucht Hilfe bzw. leistet Hilfe“ hieße es dann „ein Agent äußert einen Befehl bzw. befolgt einen Befehl“, der „Hilfe-Modus“ wäre dann der „Befehlsempfänger-Modus“ usw. Der Grund für die Austauschbarkeit der Begriffe besteht darin, dass die Art des Sanktionsmechanismus für den fraglichen Zusammenhang keine Rolle spielt, weil die beteiligten Agenten gar nicht anders können, als sich in der beabsichtigten Weise zu verhalten und deshalb auch kein Sanktionsmechanismus implementiert ist. Wiederum ist zu konstatieren: Von einer Analogie zu dem jeweiligen Interaktionsmedium (hier Wertbindung) kann man nur dann sprechen, wenn es darum geht, die Agenten unter der Voraussetzung, dass sie auch anders handeln könnten, dazu zu bringen, sich in einer bestimmten Weise zu verhalten. In geschlossenen Multiagentensystemen ist es dagegen in der Regel viel einfacher, das erwünschte Verhalten direkt zu erzwingen. Auch wenn es für das intuitive Verständnis entsprechender Multiagentensysteme hilfreich sein mag, ein Vokabular zu verwenden, das eine Analogie zu menschlicher Verhaltensabstimmung nahe legt, inhaltlich ist diese Analogie in diesen Fällen durch nichts gedeckt.

Anders ist die Lage auch hier im Fall offener Multiagentensysteme. Unter der Bedingung fehlender zentraler Kontrolle über die Eigenschaften und Verhaltensweisen der beteiligten Agenten könnte es sich durchaus als sinnvoll erweisen, den Mechanismus der über Wertbindung zugeteilten Anerkennung oder Missbilligung zu nutzen, um ein erwünschtes Verhalten zu erreichen. Entsprechende Konzepte lassen sich derzeit in der VKI jedoch noch nicht beobachten und es ist auch nicht sehr wahrscheinlich, dass sich daran schnell etwas ändern wird. Denn solange bei der Implementierung sozialer Gesetze in offenen Multiagentensystemen noch keine größeren Erfolge zu verzeichnen sind, ist kaum damit zu rechnen, dass der gegenüber manifester Belohnung/Bestrafung weniger kalkulierbare Sanktionsmechanismus der Anerkennung/Missbilligung zum Zuge kommen wird.

## **4.5 Resümee**

In der vergleichenden Betrachtung weisen die hier besprochenen Formen der Technisierung sozialer Verhaltensabstimmung mit Blick auf ihre Nutzbarkeit und Nutzung als Mechanismen der Koordination von technischen Agenten einige Spezifika auf, von denen zwei besonders hervorgehoben werden sollen. Eine Auffälligkeit ist, dass es hinsichtlich der Frage, ob man tatsächlich von einer Analogie zwischen den jeweiligen softwaretechnischen Realisierungen und dem korrespondierenden sozialen Interaktionsmedium sprechen kann, in den drei letztgenannten Fällen einen wesentlichen Unterschied macht, ob man geschlossene oder offene Multiagentensysteme betrachtet, im Fall des Markttauses dagegen offensichtlich nicht. Der Grund hierfür war, dass man in geschlossenen Multiagentensystemen in diesen drei Fällen, eine technisch einfachere Lösung wählen kann, als es der Rekurs auf das jeweilige Interaktionsmedium erlaubt: Man kann sich die Evaluation von Expertise über Reputation sparen, wenn man die Möglichkeit hat, den Agenten direkt bekannt zu geben, welcher Agent für welches Problem der Spezialist ist; und man kann auf die Implementation der Sanktionsmecha-

nismen Belohnung/Bestrafung oder Anerkennung/Missbilligung verzichten, wenn es möglich ist, die Befolgung von Befehlen, Gesetzen oder Normen direkt zu erzwingen.

Warum fehlt beim Markttausch diese Differenz? Die Antwort lautet: Der wesentliche Unterschied bei der Implementation des Kontraktnetz-Protokolls in geschlossenen oder in offenen Multiagentensystemen betrifft die Vermeidung opportunistischen Verhaltens von Agenten (bzw. deren Benutzern). Sie aber kann weder in menschlichen noch in Agentengesellschaften durch den Mechanismus des Markttauschs selbst sichergestellt werden, sondern erst durch ergänzende rechtliche oder normative Sanktionierung bzw., in geschlossenen Multiagentensystemen, durch direkte Erzwingung konformen Verhaltens. Mit Blick auf den Markttausch selbst dagegen gibt es für geschlossene Multiagentensysteme ganz offensichtlich keine gegenüber den Anforderungen offener Multiagentensysteme technisch einfachere Lösung. Warum also besteht in den drei anderen Fällen unter bestimmten Umständen die Möglichkeit, das betreffende Interaktionsmedium durch eine technisch einfachere Lösung zu substituieren und hier nicht? Folgende Antwortoptionen bieten sich an: (1) Der Koordinationsmechanismus des Markttausches ist bereits selbst eine derartig hochgradig technisierte Form der Verhaltensabstimmung, dass sie nicht mehr überboten werden kann, oder (2) die Lösung der Probleme, die der Markttausch löst, ist grundsätzlich weniger technisierbar als in den anderen Fällen, weshalb es keine einfachere Lösung als diese gibt.

Im Sinne der zweiten Antwortoption könnte man folgendermaßen argumentieren: Aus der ingenieurtechnischen Perspektive gilt, „dass der zentrale Ansatz natürlich immer besser ist, so er denn realisierbar ist“ (IW4, 4: 47f., ausführlicher zitiert oben auf S. 38). Die Substitute der drei anderen Interaktionsmedien in geschlossenen Multiagentensystemen rekurrieren in diesem Sinne sämtlich auf eine globale Vorgabe bestimmter Parameter oder Verhaltensweisen. Sofern es zutreffend ist, dass diese Möglichkeit im Fall des Markttausches nicht besteht, muss man annehmen, dass die Verteiltheit der Probleme, die der Markttausch löst, unhintergebar ist als in den anderen Fällen. Sie verwehren sich also dem Bemühen um einen zentralen Problemlösungsansatz und sind mithin weniger gut technisierbar. Nun kann man allerdings fragen, ob es denn tatsächlich keinen zentralen Problemlösungsansatz gibt, der den Markttausch unter bestimmten Umständen substituiert. Bezogen auf menschliche Gesellschaften ist die Antwort: Dies gibt es sehr wohl! Ein mögliches Substitut ist die zentralistische Planwirtschaft, die sich als volkswirtschaftlicher Koordinationsmechanismus zwar als deutlich weniger leistungsfähig erwiesen hat als der Markttausch, auf der Ebene der einzelnen Organisation als bürokratische Planung der internen Ressourcenallokation aber selbstverständlich eine bedeutende Rolle spielt.

Wenn es aber nicht so ist, dass die Probleme, die der Markttausch löst, in einer inhärenteren Weise verteilt sind und für zentrale Problemlösungstechniken grundsätzlich unzugänglicher sind als in den anderen drei Fällen, dann muss es an bestimmten Eigenschaften der vier benannten Koordinationsmechanismen selbst liegen, dass im Fall des Markttausches offensichtlich keine Notwendigkeit besteht, so möglich, auf einfachere

technische Lösungswege auszuweichen. Die Antwort auf die Ausgangsfrage ist mithin in Richtung auf die erste Antwortoption zu suchen.

Ein Unterschied, der in dieser Hinsicht von Bedeutung ist, besteht darin, dass die erwünschte Performanz des Markttauschs unmittelbar aus einem eigeninteressierten Handeln der Beteiligten resultiert, dies in den anderen drei Fällen aber nur mittelbar der Fall ist. Im idealtypischen Fall des vollständig transparenten Marktes ist das individuell nutzenmaximierende Handeln zugleich dasjenige Handeln, das die bestmögliche Allokation von Gütern und Dienstleistungen sicherstellt. In Handlungszusammenhängen, die über Macht, Einfluss und Wertbindung koordiniert werden, kann dagegen nicht mit der gleichen Selbstverständlichkeit vorausgesetzt werden, dass ein eigeninteressiertes Handeln der Individuen den entsprechenden Effekt hat. Es sind vielmehr letztlich immer erst die korrespondierenden Sanktionsmechanismen, die mittelbar bewirken, dass ein eigeninteressiertes Handeln zur Grundlage des Funktionierens der entsprechenden Koordinationsmechanismen wird. Im Fall machtbasierter Koordination beispielsweise hat der Befehlsempfänger zunächst kein unmittelbares Eigeninteresse am Gehorsam. Erst durch die Androhung von Bestrafung und die rationale Abwägung zwischen den Nachteilen, die der Gehorsam, und den Nachteilen, die die Bestrafung mit sich bringt, wird die Situation zu einer solchen, die auf der Grundlage eigeninteressierten Handelns strukturiert ist. D.h. für die Beteiligten ist es zunächst durchaus attraktiv, Handlungsoptionen zu wählen, die für die Performanz des entsprechenden Koordinationsmechanismus insgesamt betrachtet dysfunktional sind. Erst die korrespondierenden Sanktionsmechanismen, sofern sie greifen, sorgen dafür, dass solche dysfunktionalen Handlungsoptionen nicht in größerem Umfang wahrgenommen werden. Eine vergleichbar dysfunktionale Handlungsoption beim Markttausch bestünde darin, teurere Waren zu kaufen an Stelle gleichwertiger und gleich gut zugänglicher preiswerterer Waren. Selbstverständlich kommt dies aus vielerlei Gründen im real vorfindlichen Markttausch beständig vor. Worauf es allerdings ankommt ist, dass es in der idealtypischen Situation transparenter Märkte kein rationales Motiv für Käufer gibt, sich so zu verhalten.

Da es diese Idealtypik ist, auf der die Nachbildung des Markttausches im Kontraktnetz beruht, ist die VKI hier in der komfortablen Situation, auf einen Koordinationsmechanismus zugreifen zu können, in der das individuell rationale Verhalten der Agenten gleichsam automatisch die Gesamtrationalität des Austauschprozesses sicherstellt. In den drei anderen Fällen muss dagegen beständig Vorsorge gegen ein potentiell dysfunktionales Verhalten der Beteiligten getroffen werden, um die erwünschte Performanz des Koordinationsmechanismus zu gewährleisten. Wo es also möglich ist, diese störenden Verhaltensoptionen von vornherein auszuschalten, in geschlossenen Multiagentensystemen also, tendieren mithin auch VKI-Forscher – jedenfalls handlungspraktisch – dazu, ihre grundsätzliche Präferenz für Verteiltheit und *pro-activeness* der Agenten zurückzustellen und ein konformes Verhalten der Agenten direkt zu erzwingen. Die Differenz zwischen der Nutzung des Markttausches und der drei anderen Koordinationsmechanismen in geschlossenen Multiagentensystemen reflektiert also in der

Tat den Umstand, dass der Markttausch eine weiter reichende Form der Technisierung von Interaktionen repräsentiert als jene anderen Formen.<sup>14</sup> Dies erklärt nicht nur die breite Verwendung des Kontraktnetzes in der VKI, sondern umgekehrt auch den Umstand, dass der Rekurs auf die drei anderen Interaktionsmedien auch dort noch sehr in den Anfängen steckt, wo der Einsatz entsprechender Koordinationsformen besonders sinnvoll wäre: in offenen Multiagentensystemen.

In der vergleichenden Betrachtung weist neben dem Markttausch auch der Koordinationsmechanismus des Einflusses ein spezifisches Merkmal auf, das für die Abstimmung zwischen Agenten hochinteressant ist: das Merkmal, mit Blick auf opportunistisches Verhalten selbstsanktionierend zu wirken. In der einflussvermittelten Interaktion bestraft sich der opportunistisch Handelnde selbst. Indem er eine Kompetenz vorspiegelt, die er nicht besitzt, untergräbt er seinen Ruf als Experte und verringert damit seine zukünftigen Chancen, dass seine Expertise nachgefragt wird. In den anderen drei Fällen muss eine externe Sanktionsinstanz für die Verhinderung opportunistischen Verhaltens sorgen.

Der Grund für diese Differenz besteht darin, dass Reputation ein Medium ist, das kumulierte vergangene Erfahrungen in zukunftsgerichtete Erwartungen ummünzt. Geld- und machtsvermittelte Interaktionen operieren dagegen sehr viel zeitpunktbezogener, d.h. für ihr Funktionieren kommt es zunächst wesentlich darauf an, dass die beteiligten Parteien zum Zeitpunkt der betreffenden Transaktion über die erforderlichen Geld- oder Machtmittel verfügen. Das heißt natürlich nicht, dass nicht auch in geld- oder machtsvermittelter Interaktion Erfahrungen mit vergangenen Transaktionen Entscheidungshilfen für zukünftige Interaktionen darstellen können. Auf solchen Prozessen beruht die Einschätzung der Kreditwürdigkeit eines Tauschpartners im Markttausch ebenso wie die Einschätzung der Glaubwürdigkeit einer Sanktionsdrohung in der machtsbasierten Interaktion. Allerdings hat man es in solchen Fällen dann mit einer Ergänzung (und gegebenenfalls: Überlagerung) dieser Interaktionsmechanismen durch den Reputationsmechanismus zu tun: Die Interaktion wird strukturiert durch die Reputation, die man etwa als zahlungsfähiger Kunde oder als durchsetzungskräftiger Inhaber einer Machtposition besitzt.

Der entscheidende Unterschied ist, dass die Interaktionsmedien Geld und Macht selbst diese Informationen über vergangenes Verhalten des betreffenden Interaktionspartners nicht transportieren, während der Aussagewert von Reputation in nichts anderem besteht als genau darin.<sup>15</sup> Die selbstsanktierende Eigenschaft der einflussvermittel-

---

<sup>14</sup> Diese Differenz des Grades der Technisierung der jeweiligen Interaktionsmedien hat in der soziologischen Diskussion zu der Frage geführt, ob sich die Merkmale symbolisch generalisierter Interaktion, die Parsons am Beispiel des geldvermittelten Markttausches gewonnen hat, tatsächlich auf Phänomenbereiche wie die durch Macht, Einfluss oder Wertbindung vermittelte Interaktion übertragen lassen. Ich gehe auf diesen Punkt hier nicht weiter ein. Kritisch äußern sich dazu z.B. Habermas (1987, Bd. 2: 400ff.) und Künzler (1989).

<sup>15</sup> Das Interaktionsmedium Wertbindung habe ich in dieser Gegenüberstellung ausgelassen, weil es in der fraglichen Hinsicht eine Mittelstellung einnimmt: Einerseits ist Wertbindung wie Geld und



ten Interaktion könnte sich für die VKI aus den genannten Gründen insbesondere in offenen Multiagentensystemen als nützlich erweisen. Wie bereits konstatiert, lassen sich Forschungsanstrengungen in diese Richtung derzeit jedoch bestenfalls in Ansätzen beobachten.

## **5. Vergemeinschaftung von Agenten: Koordination durch Bekanntschaftsstrukturen**

Weber zufolge bezeichnet Vergemeinschaftung eine Form der Strukturierung sozialen Handelns, die wesentlich auf dem subjektiven Gefühl der Zusammengehörigkeit der Beteiligten beruht. Da nun aber nicht ersichtlich ist, in welchem inhaltlich gehaltvollen Sinne davon gesprochen werden kann, dass Software-Agenten Zusammengehörigkeitsgefühle besitzen oder entwickeln können, sind vergemeinschaftete Formen sozialer Verhaltensabstimmung auf den ersten Blick ungeeignete Kandidaten als Modelle der Handlungskoordination zwischen Agenten. Allerdings lassen sich Phänomene der Vergemeinschaftung – zu bestimmten Anteilen jedenfalls – auch ohne Rückgriff auf die Kategorie subjektiv gefühlter Zusammengehörigkeit beschreiben. Will man eine überbordend metaphorische Analogisierung menschlicher und agententechnologischer Verhaltensrepertoires vermeiden, empfiehlt es sich bis auf weiteres, die Betrachtung auf solche Aspekte der Vergemeinschaftung von Agenten zu beschränken.

Im Folgenden werde ich mich deshalb auf einen Begriff der Vergemeinschaftung beschränken, der Formen der Strukturierung sozialen Handelns beschreibt, wie sie als Resultat der Interaktion zwischen einander Bekannten entstehen. Die Unterscheidung zwischen Vergemeinschaftung und Vergesellschaftung soll hier also mittels der Differenz zwischen Bekanntschaft und Anonymität vorgenommen werden: Vergesellschaftete Formen sozialer Interaktion ermöglichen eine Handlungskoordination zwischen einander weitgehend Unbekannten, sofern diese sich in bestimmten Hinsichten an generalisierten Erwartungen orientieren. Als vergemeinschaftete Formen des sozialen Handelns sollen im Gegensatz dazu diejenigen Interaktionsmuster gelten, die sich auf der Grundlage wechselseitiger Bekanntschaft ausbilden. Der Begriff der Bekanntschaft soll dabei ein Wissen über Verhaltenseigenschaften des jeweils Anderen bezeichnen, das die Interaktionspartner in der direkten persönlichen Interaktion ausbilden.

Die Differenz Vergesellschaftung/Vergemeinschaftung führt in dieser Fassung neben der Unterscheidung Anonymität/Bekanntschaft zwei weitere Unterscheidungen mit: die Unterscheidung zwischen (möglicher) Punktualität und notwendiger Kontinuität des Interaktionsprozesses und die Unterscheidung zwischen (potentieller) Unbegrenztheit und notwendiger Begrenztheit der Möglichkeit zur Teilnahme an entsprechenden Interaktionszusammenhängen. Die technisierten Formen vergesellschafteter

---

Macht in dem Sinne zeitpunktgebunden, als sie glaubhaft nur durch das je aktuelle Handeln in Übereinstimmung mit dem betreffenden Wert zum Ausdruck gebracht werden kann. Andererseits besteht die Glaubwürdigkeit des wertbezogen Handelnden aber auch darin, dass er seine Werthaltungen nicht beständig wechselt, was seine Anerkennung an die Kontinuität seines Handelns und damit an seine Reputation als rechtschaffene Person bindet.

Interaktion reduzieren Kommunikationserfordernisse in einer Weise, die es – als ein Leistungsmerkmal solcher Interaktionsstrukturen – möglich macht, punktuelle Austauschbeziehungen einzugehen, also solche, in denen sich die Interaktion weitgehend auf das Ereignis des fraglichen Austausches selbst beschränken kann (also z.B. Ware gegen Geld oder Gehorsam gegen Belohnung/Straflosigkeit). Bekanntschaft dagegen erfordert Zeit. Indem man das je gegenwärtige Verhalten des Gegenübers im Lichte der vorangegangenen Erfahrungen interpretiert, kristallisiert sich auf beiden Seiten ein Bild des Anderen und ein Bild der Beziehung zueinander heraus und entstehen Handlungs- und Reaktionsmuster des Umgangs mit dem jeweils Anderen, die sich, soweit sie wechselseitig zumindest toleriert werden, als Interaktionsmuster verfestigen können. Dies bedeutet zugleich, dass die Anzahl der eigenen Bekannten stets begrenzt ist. Die symbolisch generalisierten Medien vergesellschafteter Interaktion verringern den Aufwand des Aufbaus einer Interaktionsbeziehung unter bestimmten Bedingungen so weit, dass dieser Aufwand im einmaligen Austausch mit einem Interaktionspartner, mit dem man vorher nie zu vor zu tun hatte und mit dem man möglicherweise zukünftig nie wieder zu tun haben wird, nicht größer ist als im wiederholten Austausch. Der Aufbau und die Aufrechterhaltung von Bekanntschaft bindet dagegen Zeit und Aufmerksamkeit, was die Zahl der Bekanntschaftsbeziehungen zwangsläufig begrenzt, die der einzelne Akteur einzugehen und zu pflegen in der Lage ist.

## **5.1 Das Problem der Skalierbarkeit – umgekehrt betrachtet**

Die bislang entwickelten Multiagentensysteme beschränken sich in aller Regel auf wenige Agenten. Dies hat zu einem guten Teil forschungspragmatische Gründe: Die zu meist universitären VKI-Forscher sind primär an prototypischen Realisierungen ihrer Konzepte interessiert, setzten diese also nur so weit softwaretechnisch um wie unbedingt erforderlich ist, um die fraglichen Funktionseigenschaften testen zu können. Bei bestimmten Multiagentensystem-Architekturen wäre es dagegen hochgradig interessant, sie unter der Bedingung sehr großer Multiagentensysteme auszutesten. Dies gilt insbesondere für solche Architekturen, die vergesellschaftete Formen des sozialen Austauschs nachbilden, weil deren Leistungsfähigkeit – jedenfalls in menschlichen Gesellschaften – ja gerade in offenen Interaktionszusammenhängen mit potentiell beliebigen Partnern zum Tragen kommt. Eine solche Zielstellung wirft ein Skalierungsproblem auf:

„Gerade zum Beispiel Multiagentensysteme, da ist gar nicht klar: In kleinen Bereichen funktionieren die vielleicht wunderbar, aber nehme ich jetzt eben hunderte oder tausend Agenten: Wie entwickelt sich das ganze System, welche Strukturen müssen sich zusätzlich rausbilden, damit das überhaupt funktionieren kann?“ (FW3, 12: 27 - 30)

Worin aber besteht das Skalierungsproblem, d.h. was sind die Gründe dafür, dass sich die Anzahl der Agenten in den gegenwärtigen Multiagentensystemen nicht umstandslos vergrößern lässt? Malsch (1998: 256f.) gibt dem Skalierungsproblem aus der soziologischen Perspektive die folgende Fassung: „Wie ist es möglich, einen tiefgreifenden realgesellschaftlichen Wandlungsprozeß, den wir mit der griffigen Formel der Umstellung von ‚Gemeinschaft‘ auf ‚Gesellschaft‘ (Tönnies) nur näherungsweise umschrei-

ben können, im Medium der Multiagenten-Technologie darzustellen? Und inwieweit korrespondiert diese sozialtheoretische Frage mit der informatischen Frage, wie es möglich ist, ‚skalierbare‘ Multiagenten-Systeme zu bauen? ... Wie ist es möglich, sehr große und dynamische Multiagenten-Systeme mit Hunderttausenden oder Millionen von eigensinnigen Agenten entsprechend dem Vorbild moderner menschlicher Gesellschaften zu bauen? Wie ist es möglich, die bisherige Begrenzung der Multiagenten-Systeme auf wenige autonome Agenten zu überwinden?“

Diese Fragenkaskade legt die Interpretation nahe, dass das Skalierungsproblem der VKI im Wesentlichen identisch ist mit der einen Seite des Mikro-Makro-Problems in der Soziologie, nämlich der Frage, wie aus der direkten Interaktion zwischen Anwesenden über Formen der Vergemeinschaftung kleinerer Kollektive schließlich übergreifende gesellschaftliche Zusammenhänge emergieren, die das Zusammenleben sehr vieler Akteure strukturieren. Bei der Analyse menschlicher Gesellschaften hat diese Frage sicherlich ihre Berechtigung. Auf das Skalierungsproblem der VKI ist sie jedoch nicht übertragbar und dies aus einem einfachen Grund: Die Fragestellung impliziert, dass das Problem des Übergangs von kleinen zu großen Kollektiven primär ein Problem des Übergangs von vergemeinschafteten zu vergesellschafteten Formen sozialer Interaktion ist. Aus der Tatsache, dass sich die VKI bislang vor allem mit kleinen Kollektiven von Agenten befasst hat, folgt jedoch keineswegs, dass die dort verwendeten Interaktionsmechanismen soziologisch interpretiert Formen der Vergemeinschaftung sind. Wie wir im vorangegangenen Kapitel gesehen haben, ist das Gegenteil der Fall: Die erfolgreichsten und am häufigsten verwendeten Koordinationsmechanismen der VKI sind ohne Zweifel vergesellschaftete Formen der Interaktion.

Das Skalierungsproblem der VKI besteht also sehr viel weniger darin, den Übergang von vergemeinschafteten zu vergesellschafteten Formen der Verhaltenskoordination zu bewältigen, als vielmehr darin, von Formen vergesellschafteter Interaktion zwischen wenigen Agenten zu solchen Interaktionen zwischen vielen Agenten zu gelangen. Für den soziologischen Betrachter mutet diese Problemstellung höchst merkwürdig an, weil es im Bereich der menschlichen Sozialität unmittelbar evident ist, dass Formen der Vergesellschaftung im Zusammenhang der Vergrößerung der raumzeitlichen Abstände, über die hinweg Interaktionen nun abgewickelt werden konnten, entstanden sind. Es ist diese Selbstverständlichkeit des soziologischen Blickwinkels, die die hier zurückgewiesene Interpretation des Skalierungsproblems der VKI nahe legt, da sie eine Entwicklungsdynamik, die mit vergesellschafteter Interaktion zwischen wenigen Interaktionspartnern beginnt, nicht erwartet.

Dass sich dies aus der Perspektive der VKI ganz anders darstellt, hängt ein weiteres Mal mit dem umgekehrten Komplexitätsgefälle der sozialtheoretischen und der agententechnologischen Betrachtung der Strukturierung von Sozialbeziehungen zusammen, wonach die sozialtheoretisch hochgradig voraussetzungsreichen Formen vergesellschafteter Interaktion unter ingenieurtechnischen Gesichtspunkten – in dem Maße, in dem sie die Interaktion technisieren – die am einfachsten nachzubildenden Vorbilder aus dem Bereich menschlicher Verhaltensabstimmung bilden. Wenn aber bereits die

kleinen Multiagentensysteme der VKI zu überwiegenden Anteilen auf der Grundlage vergesellschafteter Interaktion operieren, stellt sich die Frage, weshalb dann überhaupt ein Skalierungsproblem besteht. Die Antwort haben wir bereits im vorigen Kapitel kennen gelernt: Die jeweils erforderlichen Rahmenbedingungen des Funktionierens der besprochenen Koordinationsmechanismen lassen sich in geschlossenen leichter als in offenen Multiagentensystemen sicherstellen. Das Skalierungsproblem der VKI ist dementsprechend im Wesentlichen ein Problem des Übergangs von geschlossenen zu offenen Systemen.

Wenn es in der VKI also so etwas gibt, wie ein Skalierungsproblem, das mit dem Übergang zwischen vergemeinschafteten und vergesellschafteten Formen der Interaktion zusammenhängt, so stellt es sich vorrangig in der umgekehrten Richtung: Wie ist es möglich, die Interaktionsfähigkeit der Agenten in einer Weise zu erweitern und anzureichern, die es ihnen erlaubt, sich nicht nur auf den vorgezeichneten Pfaden technisierter Koordinationsmechanismen auszutauschen, sondern wechselseitig Erfahrungen übereinander zu sammeln, auf deren Grundlage Interaktionsbeziehungen und Austauschstrukturen eigenständig entstehen können, wie dies in der vergemeinschafteten Interaktion zwischen menschlichen Akteuren stattfindet? Es fragt sich allerdings, ob dies für die VKI überhaupt ein Problem ist. Für die große Mehrheit der VKI-Forscher ist die agententechnologische Nachbildung menschlicher Sozialität kein Selbstzweck. Sie muss sich vielmehr an dem ingenieurtechnischen Maßstab messen lassen, ob sich auf diese Weise Techniken entwickelt lassen, die bestimmte Probleme lösen bzw. besser lösen als andere Techniken. Stellvertretend für die Haltung der meisten der von uns Befragten kommt dies in dem folgenden Zitat deutlich zum Ausdruck:

„Also es geht darum, wirklich ein effizientes System zu bauen, so ist die Ansatzweise. Und das, wie gesagt, nicht diesen Adäquatheitsanspruch erhebt. Also das ist der Hintergrund unserer Forschung, die wir machen. Aber du könntest das natürlich auch anders angehen. Du könntest sagen: ‚Okay, das soll ein Modell für kognitive Leistung sein.‘ Aber dann muss eben die Methodologie eine andere sein. Dann darfst du da nicht hingehen und sagen: ‚Okay, wir implementieren das und holen jetzt Applikationen ran‘ (...). Sondern da muss die Methodologie ganz anders sein, da musst du eben hingehen und sagen: ‚Okay, wir implementieren das jetzt vielleicht und gucken mal, wie funktioniert das für dieses Problem.‘ Und dann musst du auch in der Natur nachgucken, musst sagen: ‚Okay, wie funktioniert denn der Mensch in dem und dem Problem?‘ Und wenn du dann irgendwie signifikante Überschneidungen zwischen dem Ergebnis von deinem System und dem Ergebnis des Menschen kriegst, dann kannst du sagen: ‚Bezüglich bestimmter Eigenschaften ist das ein gutes kognitives Modell.‘ (...) Also wie gesagt, also wir evaluieren anhand von Optimalität und Applikation: Gibt es da ein anderes Modell, das einfach besser ist?“ (FW3, 19: 41-51, 21: 20f.)

Vor diesem Hintergrund muss es schon sehr gute Gründe geben, der direkt ausgehandelten Interaktion und den aus der direkten Interaktion emergierenden Interaktionsmustern den Vorzug zu geben gegenüber den vergesellschafteten Formen der Interaktion. Denn unter dem Gesichtspunkt angestrebter technischer Effizienz von Multiagentensystemen ist das Anknüpfen an bereits technisierte Formen sozialer Koordination zunächst viel versprechender als die Nachbildung der weniger stark technisierten und schlechter technisierbaren Muster vergemeinschafteter Zusammenhänge. Gründe dafür, dies dennoch anzustreben, finden sich, wie sich im Folgenden zeigen wird, insbesondere dort, wo der Wunsch den Ausschlag gibt, die typische Schwäche eines jeden

technischen Verfahrens zu überwinden bzw. zu kompensieren: die des Verlusts an flexibler Anpassung an veränderliche Gegebenheiten je lokaler Einsatzzusammenhänge.

Es gilt festzuhalten: Eine im Sozialisationsprozess gleichsam naturwüchsig erworbene Sprach- und Interaktionsfähigkeit ist etwas, das in menschlichen Gesellschaften als gegebene Eigenschaft ihrer durchschnittlich kompetenten Mitglieder behandelt werden kann. Dies evoziert die Frage, wie menschliche Kollektive von den Mustern direkt ausgehandelter Interaktion zu übergreifenden gesellschaftlichen Strukturen gelangen. Diese Frage hat, zumindest entwicklungsgeschichtlich, ihre Berechtigung (wenn sie auch für die Analyse gegenwärtiger Gesellschaften zu schlicht ist, da jene übergreifenden Strukturen für das werdende Gesellschaftsmitglied zumeist in genau dem gleichen Sinne sozialisatorisch vermittelte Gegebenheiten sind). Der Prozess der zunehmenden Befähigung von Agenten zur Interaktion verläuft aber mit einiger Prägnanz stets in umgekehrter Richtung. Der soziologische Beobachter der VKI kann sich einige Enttäuschungen ersparen, wenn er dies in Rechnung stellt. Er wird die in der VKI-Literatur häufig recht freizügig verwendeten anthropomorphisierenden Beschreibungen von Agenten, die eine menschlichen Akteuren vergleichbare Interaktionskompetenz suggerieren, dann mit einiger Zurückhaltung interpretieren und nicht erwarten, anderswo auf vergemeinschaftete Interaktion zwischen Agenten zu stoßen als dort, wo es um die Anreicherung und Flexibilisierung technisierter Interaktionsstrukturen geht.

## **5.2 Ausgehandelte Kooperation**

Wenn in unseren Interviews die Rede auf die Möglichkeit der in der Interaktionssituation selbst direkt ausgehandelten Kooperation kam, waren die Reaktionen unserer Gesprächspartner in der Regel zurückhaltend bis ablehnend. Der Tenor der Äußerungen lässt sich dahingehend zusammenfassen, dass das Ziel der VKI sein müsse, zunächst möglichst einfache Koordinationsstrukturen zu finden und softwaretechnisch in den Griff zu bekommen, und, solange dies möglich sei, auf die als eher langwierig und kompliziert eingeschätzte Strategie der direkten Aushandlung von Kooperationsbeiträgen zu verzichten. Einige Zitate, die diese Grundhaltung belegen, habe ich oben bereits angeführt (vgl. IW1, 4: 16-20, zit. auf S. 14 und UP1, 26: 43 - 27: 3, zit. auf S. 33f.)

Allerdings ist es unter bestimmten Umständen durchaus sinnvoll, technisierte Koordinationsverfahren, wie etwa das Kontraktnetz, um Aspekte freier Aushandlung zu erweitern. Gründe für ein solches Vorgehen sind oben (auf S. 33f.) bereits angesprochen worden: Wenn es, wie etwa im Beispiel des Reiseplanungssystems, nicht unbedingt das eine genau passende Angebot für eine Nachfrage gibt oder die Nachfrage selbst spezifizierungsbedürftig ist, sind Aushandlungsprozesse geeignet, um auf dem Weg des Vorschlags von näherungsweise passenden Angeboten, der Modifikation der Nachfrage usw. zu einer befriedigenden Lösung zu gelangen. Ein weiterer Grund liegt vor, wenn nicht von vornherein sichergestellt ist (durch Zielvorgaben, an denen sich alle Agenten orientieren müssen, durch vorgeplante Arbeitsteilung oder durch die unsichtbare Hand des Marktes), dass sich die Beiträge der beteiligten Agenten zu erwünschten Performanz des Systems aufsummieren. Dies wäre beispielsweise im An-

wendungsfeld der agententechnologischen Fertigungssteuerung dann der Fall, wenn ein Multiagentensystem nicht nur innerhalb eines Unternehmens, sondern unternehmensübergreifend eingesetzt würde:

„Der Grund, dass das so ist (d.h. dafür, dass bei Multiagentensystemen für die Fertigungssteuerung innerhalb eines Unternehmens nach Ansicht des Gesprächspartners auf Verhandlung verzichtet werden kann, Anm. d. Verf.), lässt sich letztendlich darauf zurückführen, dass ich in der Produktion keine völlig autonomen Systeme habe. Also ich habe immer ein übergeordnetes Ziel, das ich verfolge. Und ich muss das übergeordnete Ziel natürlich runterbrechen. Wenn ich das irgendwann einmal runtergebrochen habe, dann kann es passieren, dass der eine [Agent] ein anderes Ziel verfolgt als der andere, sodass sie dann schon mal in Situationen kommen, wo man eigentlich verhandeln müsste. Aber das ist ja nie so, dass sie ihre eigenen Interessen verfolgen, sondern ich kann immer sagen: ‚In Rücksichtnahme auf das übergeordnete Ziel will ich dort eine Einigung herbeiführen.‘ Deswegen ist es so, dass ich in den meisten Fällen zu Koordination komme. Es wäre etwas anderes, wenn ich jetzt *distributed manufacturing* machen würde. Ich hätte zwei unterschiedliche Firmen, die ihre eigenen Ziele verfolgen würden, und diese beiden würden miteinander kooperieren. Dann hätte ich die Situation: Ich hätte zwei unterschiedliche Ziele. Auf irgendeiner Ebene muss ich eine Einigung herbeiführen. Dann hätte ich eine Verhandlungssituation. Aber so weit sind wir noch nicht.“ (IW1, 6: 29-45)

Allerdings unterscheiden sich die hier teils in die Zukunft verschobenen, teils aber auch schon in Angriff genommenen Ideen über Verhandlungsprozesse in Multiagentensystemen doch noch deutlich von dem, was man mit dem folgenden Zitat als „Verhandlung im engeren Sinne“ bezeichnen könnte:

„Unter Verhandlung verstehen wir im engeren Sinne, dass ich zwei Agenten habe, die eine unterschiedliche Meinung haben, unterschiedliche Ziele verfolgen und einen unterschiedlichen Lösungsvorschlag machen. Und die erarbeiten zusammen einen gemeinsamen Lösungsvorschlag.“ (IW1: 6, 20-24)

Diese reine Form direkt ausgehandelter Kooperation wäre eine solche, bei der das Ergebnis des Aushandlungsprozesses allein durch die in der Interaktionssituation selbst vorgebrachten und akzeptierten (bzw. abgelehnten oder modifizierten) Wissensbestände, Interessen und Zielerreichungsstrategien zustande käme. Das setzt voraus, dass die fragliche Situation nicht (oder wenigstens nicht wesentlich) vorstrukturiert ist durch eine gemeinsame Weltsicht und gemeinsam geteilte Wissensbestände, durch kompatible Ziele und Interessen oder durch übereinstimmende Bewertungen von Mitteln der Zielerreichung. Bereits im Bereich menschlicher Interaktion stoßen wir selten auf Situationen, die in diesem Sinne unstrukturiert sind. In der Welt der Agenten werden wir nach solchen Situationen vollends vergeblich suchen. Die Möglichkeit der Verwendung von Verhandlungsprozeduren in Multiagentensystemen beruht vielmehr in allen Fällen, in denen dies angedacht wird, ganz wesentlich darauf, dass in einem insgesamt eher gut vorstrukturierten Interaktionszusammenhang einzelne Komponenten des Koordinationsprozesses zur Aushandlung freigegeben werden.

Bei denjenigen der von uns Befragten, die sich mit diesem Thema befassen, werden Verhandlungsmöglichkeiten zumeist in den Rahmen von Kontraktnetz-Architekturen eingebettet. Dies engt den Spielraum für Aushandlung, wie oben (auf S. 34f.) bereits angemerkt, deutlich ein. Es geht dann etwa nur noch darum auszuhandeln, ob die Partner bereit sind, von ursprünglichen Preisvorstellungen, Terminwünschen, Reiseplänen usw. so weit abzuweichen, dass eine für potentielle Vertragspartner befriedigende Abstimmung von Angeboten und Nachfragen möglich wird. Nichtsdestotrotz wird damit

die technische Striktheit (aber auch Einfachheit) des verhandlungsfreien Markttausches aufgeweicht und werden Anpassungsmöglichkeiten eröffnet, die ihm fehlen. Während dort davon ausgegangen wird, dass es stets von vornherein hinreichend viele zueinander passende Angebote und Nachfragen gibt, um die erforderliche Performanz des angestrebten Zusammenhanges sicherzustellen, ermöglicht die Einbeziehung von Verhandlungsaspekten es, Angebote und Nachfragen aufeinander abzustimmen, die dies nicht von vornherein sind.

Bedarf nach einer solchen Flexibilisierung des Interaktionsgeschehens im Markttausch besteht dort, wo es relativ zu der Zahl der aktuell auf dem Markt befindlichen Angebote bzw. Nachfragen eine hohe Zahl möglicher Ausprägungen von Angeboten bzw. Nachfragen gibt. Die Wahrscheinlichkeit, dass eine solche Situation eintritt, ist dann besonders groß, wenn die Angebote Unikate sind (vgl. oben S. 35) und/oder die Nachfragen spezifische Gegebenheiten des einzelnen Nachfragers reflektieren und/oder dann, wenn die Zahl der Anbieter und Nachfrager (samt ihrer Angebote und Nachfragen) vergleichsweise gering ist (wie im Fall der Terminplanung). Das Merkmal hoher Individualität von Angeboten oder Nachfragen impliziert jedoch für sich genommen ebenso wenig wie das einer geringen Zahl von Marktteilnehmern, dass entsprechende Verhandlungen zugleich auch schon den Tatbestand der Vergemeinschaftung erfüllen. In den entsprechenden Termin- oder Reiseplanungssystemen sind solche Aushandlungsprozesse bislang vorwiegend als punktuelle Ereignisse konzipiert, die keine Bekanntschaft der Interaktionspartner voraussetzen. Dies entspricht durchaus der Situation, in der entsprechende Verhandlungen vielfach im gesellschaftlichen Austausch menschlicher Akteure geführt werden, etwa der des Feilschens auf dem Flohmarkt.

In anderen Fällen kann es jedoch sehr wohl von Vorteil sein, Verhandlungen zwischen Tauschpartnern auf der Grundlage zunehmender Bekanntschaft zu führen. So zeigt beispielweise die Untersuchung sogenannter Innovationsnetzwerke (vgl. Freeman 1991; Kowol/Krohn 1995; Schulz-Schaeffer et al. 1997), dass es in der innovationsbezogenen Kooperation zwischen Organisationen günstig sein kann, marktförmige Beziehungen durch Formen des Austauschs anzureichern, die wesentlich auf Vertrauen und Reziprozität beruhen. Solche Kooperationsformen erweisen sich dem vergesellschafteten Austausch (in diesem Fall insbesondere: dem markt- und dem machtförmigen) vor allem dann als überlegen, wenn es nicht möglich ist, die Kooperationsbeiträge der Beteiligten vor Beginn der Kooperation hinreichend zu spezifizieren und in ihrem Wert zu bestimmen. In Ermangelung dieser Kriterien ist es für die Beteiligten deshalb rational, Kooperationspartner auszuwählen, in deren Kompetenz und Fairness man vertraut, wobei die Grundlage dieses Vertrauens in vielen Fällen vergangene Erfahrungen sind: Man hat den Partner kennen gelernt als einen, der Probleme der fraglichen Art bewältigen kann, der einen nicht zu übervorteilen sucht usw.

Ansätze in der VKI, die in ähnlicher Weise darauf zielen, Erfahrungen, die ein Agent mit anderen Agenten macht, zu einem Wert aufzusummieren, die dieser Agent seinem zukünftigen Verhalten gegenüber den von ihm dergestalt bewerteten anderen

Agenten zu Grunde legt, hatte ich oben (auf S. 20f. und 52f.) bereits angesprochen. Solche Werte repräsentieren dabei die Reputation, die Agenten – als Resultat ihres vergangenen Verhaltens – bei anderen Agenten (bzw. bei menschlichen Benutzern) angesammelt haben. Die Nutzung von Reputation als Kriterium der Partnerwahl kann sich auf eine ganze Reihe unterschiedlicher Verhaltenseigenschaften von Agenten richten, deren Prognose für Kooperationsentscheidungen unter Umständen hilfreich ist: In dem eben genannten Fall wird der Reputationswert zur Einschätzung der Kompetenz der Agenten bei der Problemlösung verwendet. Dagegen geht es bei Castelfranchi et al. (1998) um Reputation bezogen auf Normkonformität (vgl. oben S. 54f.). Und viele andere Anwendungsfelder sind denkbar.

In dem Ansatz eines weiteren Gesprächspartners dienen die kumulierten Werte dazu, die Ehrlichkeit anderer Agenten einschätzen zu können. Die Grundidee ist hier ähnlich wie bei Castelfranchi et al. (1998): Die Agenten sind daran interessiert, die Normkonformität anderer Agenten mit Blick auf die Norm der Gutwilligkeit, d.h. deren Kooperationsbereitschaft einzuschätzen. Dies geschieht über einen Reputationswert für Ehrlichkeit, der jeweils aus der Wahrnehmungsperspektive eines Agenten die Differenz zwischen geäußelter und vollzogener Kooperationsbereitschaft eines anderen Agenten misst. In diesem Sinne wird dann auch der Begriff des Vertrauens definiert:

„Ich benutze den Begriff Vertrauen nur für die Wahrscheinlichkeit, die ein spezieller Agent ermittelt hat, dass sich ein anderer an seine Aussagen halten wird.“ (FD1, 3: 52-54)

Zusätzlich wird von großen Multiagentensystemen ausgegangen, in denen es sinnvoll ist, sich auf die Auskünfte dritter Agenten über die mutmaßliche Kooperationsbereitschaft des fraglichen anderen Agenten zu verlassen. Indem nun des Weiteren vorgesehen ist, dass die Agenten nicht nur bei Auskünften über ihr eigenes Verhalten, sondern auch bei Auskünften über die Vertrauenswürdigkeit anderer Agenten falsche Aussagen treffen können, um strategische Vorteile zu erringen (etwa indem sie mögliche Konkurrenten um einen erwünschten Kooperationspartner diskreditieren), wird der Reputationswert für Ehrlichkeit noch in einer zweiten Hinsicht genutzt, nämlich um Einschätzungen der diesbezüglichen Vertrauenswürdigkeit des betreffenden Agenten zu erhalten:

„Ich will im Prinzip untersuchen, inwiefern so etwas funktioniert, wenn ich einem Agenten vertraue und der vertraut einem anderen. Und ich aber nichts über den dritten weiß. Inwiefern ich dem dann auch vertrauen kann. (...) Vertrauen sehe ich dann eben als so eine Art Kitt, der diese Gemeinschaft zusammenhält und über das ich dann eben auch andere Informationen weitertragen könnte. Also wenn ich zum Beispiel weiß, dass irgendein Agent sehr sehr zuverlässig ist, dann kann ich ja auch die Informationen, die ich von diesem Agenten bekomme, ganz anders bewerten. Oder wenn er mir Dienstleistungen anbietet oder wenn er mir ja, irgend etwas anbietet.“ (FD1, 2: 50 - 3: 9)

„Also im Prinzip wird es für den für den Agenten in zwei Schritten laufen, (...) zwei verschiedene Prozesse: Zum einen wird er immer seine Umgebung wahrnehmen, immer beobachten: Wie verhalten sich die anderen? Inwiefern sind sie wirklich ehrlich? (...) Und das andere, was halt reinkommt, ist ... Um das vielleicht besser zu erklären: (...) Bisher im Multiagentensystem geht man immer davon aus, dass die alle vertrauenswürdig sind. Oder aber überhaupt nicht. Und ich wollte jetzt die Information benutzen: Wenn ich (d.h. der fragliche Agent, Anm. d. Verf.) in einer Gruppe bin mit bestimmten Agenten und da kommen neue rein und andere gehen raus. Also ich habe eine Fluktuation in der Gruppe. Dass ich da eben diese Ideen benutzte, wenn ich jemandem vertraue, der jemandem anderem vertraut, dann weiß ich auch etwas Neues über diesen dritten. Und das kann eben dadurch entstehen,



dass die schon mal zusammen irgendwo in einer anderen Gruppe drin waren, wo ich nicht drin war.“ (FD1, 4: 28-39)

„Also, ich gebe den [Agenten] eigentlich keine Basislinie vor. Sondern, ich denke mal, die werden am Anfang recht vorsichtig sein, was Kooperation angeht. Und sich nach einigen Testläufen schon recht gut orientieren können in ihrer Gruppe, mit wem sie halt zusammen arbeiten können. Und diese Idee, dass ich eben rauskriegen kann oder damit berechnen kann, wie ehrlich andere sind, soll eben großen Multiagentensystemen die Möglichkeit geben, auch damit klarzukommen, dass in die eigene Gruppe neue Agenten eintreten, von denen ich nicht weiß, ob sie benevolent sind oder nicht. Also ob sie sich kooperativ verhalten oder nicht. Und eben die Möglichkeit zu haben, darüber Daten zu sammeln und dann zu entscheiden, mit wem ich nun rein aus Prinzip kooperiere und mit wem nicht.“ (FD1, 12: 9-16)

Wie bei den meisten anderen Versuchen, Kooperationsstrategien auf der Grundlage der Einschätzung einer wie immer gearteten Vertrauenswürdigkeit von Agenten zu bilden, handelt es sich auch hier um ein Multiagentensystem, dem eine Interaktionssituation zugrunde liegt, die zu Zwecken der Theoriebildung erdacht worden ist (und die, wie in den meisten Fällen, aus dem Umfeld von Spieltheorie und Rational-Choice-Theorie stammt). Ausgangspunkt ist die spieltheoretische Standardsituation des Gefangenendilemmas, die im Kern durch folgenden Zusammenhang des Handelns zweier Akteure charakterisiert ist: (1) Wenn die Beteiligten kooperieren, teilen sie sich den Gewinn; (2) wenn ein Akteur kooperiert und der andere egoistisch handelt, bekommt der egoistische Akteur den ganzen Gewinn; (3) wenn beide egoistisch handeln, ist der Gewinn, den die Akteure erzielen, kleiner als im ersten Fall, aber größer als der des kooperativen Akteurs im zweiten Fall (vgl. Axelrod 1984). In Handlungszusammenhängen, wie sie das Gefangenendilemma beschreibt, ist die Wahl der günstigsten Handlungsalternative für jeden der Beteiligten mithin abhängig von der Wahl des bzw. der anderen Beteiligten. Es ist also anstrebenswert, etwas über das vermutliche Verhalten des/der Anderen in Erfahrung zu bringen.

Möglich wird dies insbesondere dann, wenn der beschriebene Handlungszusammenhang kein einmaliges Ereignis ist, sondern wiederholt vorkommt. Das spieltheoretische Standardbeispiel ist hier das iterierte Gefangenendilemma. Es bietet sich nun an, die Einschätzung des vermuteten zukünftigen Verhaltens des fraglichen anderen Agenten aus seinem vergangenen Verhalten abzuleiten. Dieser Strategie folgt das Verfahren, kumulierte Werte zu berechnen, die das Ausmaß der vergangenen Kooperativität (hier mittels der Einschätzung ihrer Ehrlichkeit) anderer Agenten ausdrücken. Unter Nutzung solcher Reputationswerte wird der Agent, der kooperationswillig ist, bestrebt sein, einen Agenten mit einem möglichst hohen Kooperationswert zu finden. Der egoistische Agent dagegen wird es im Laufe der Zeit immer schwerer haben, einen kooperationswilligen Agenten zu finden, den er ausbeuten kann. Auf diese Weise kann die selbstsanktionierende Wirkung von Reputation als Koordinationsmechanismus genutzt werden.

Zusammenfassend kann festgehalten werden: Der Aspekt direkt ausgehandelter Kooperation zwischen Agenten gewinnt in der Multiagenten-Forschung offenkundig an Bedeutung. In keinem Fall aber ist das Ziel die Bewältigung reiner doppelter Kontingenz, bei der allein die Interaktionen in der fraglichen Situation selbst das Ergebnis des

Aushandlungsprozesses produzieren. Vielmehr geht es stets darum, entweder auf bestehende Kommunikationsprotokolle zurückzugreifen und in diesem Rahmen Spielräume für direkte Aushandlung zu eröffnen. Oder aber die kooperationsbezogenen Aushandlungsprozesse rekurren auf soziale Strukturen, die im Verlauf der Interaktionssequenz selbst aufgebaut werden. Im ersten Fall haben wir es mit einer Flexibilisierung vergesellschafteter Formen des Austausches zwischen Agenten zu tun, im zweiten Fall geht es um die Etablierung von Strukturen direkter und indirekter Bekanntschaft, in der hier vorgeschlagenen Terminologie also um Phänomene der Vergemeinschaftung. Dabei ist die zentrale Sozialvorstellung, die in der VKI in den letzten Jahren zunehmend Beachtung gefunden hat, die der vertrauensgesteuerten Verhaltensabstimmung:

„Trust is one of the most important concepts guiding decision making and contracting in human societies. In artificial societies, this concept has until recently been neglected. The inherent benevolence assumption implemented in many multi-agent systems can have hazardous consequences when dealing with deceit in open systems. Therefore we present a formalisation of trust (...) This approach (...) includes social behaviour for implementing punishment without empowered authorities.“ (Schillo et al. 1999)

Vertrauen wird hier als ein Koordinationsmechanismus vorgestellt, der sich insbesondere dann als nützlich erweisen könnte, wenn die Gutwilligkeit (oder, wie man ergänzen kann: eine wie auch immer geartete Berechenbarkeit des Verhaltens) der beteiligten Agenten nicht von vornherein vorausgesetzt werden kann. Besonders interessant an dem Zitat ist, dass die Autoren offensichtlich auch die besonderen Vorteile erkannt haben, die aus der selbstsanktionierenden Wirkung des Reputationsmechanismus resultieren.

### **5.3 Agentenbezogene Einstellungen und Erwartungen**

Wenn die symbolisch generalisierten Interaktionsmedien in der soziologischen Theorie als generalisierte Erwartungen kennzeichnet werden, so ist damit zugleich eine Hypothese über deren Genese impliziert, die man grob wie folgt skizzieren könnte: Um sich in der sozialen Interaktion auf das Handeln anderer einstellen zu können, bilden Akteure Erwartungen übereinander aus. In dem Maße, in dem sich bestimmte Interaktionssequenzen einspielen, verfestigen sich die darauf bezogenen Handlungserwartungen. Gewinnen solche Interaktionssequenzen den Status mehr oder weniger fest stehender Interaktionsmuster, die typischerweise stets in bestimmten Situationen ablaufen, werden die auf sie bezogenen Erwartungen verallgemeinerbar. Es entstehen generalisierte Erwartungen, die abgerufen werden, wenn die fragliche Interaktionssituation die entsprechende Typik aufweist. Dies bedeutet umgekehrt, dass diese generalisierten Erwartungen von den individuellen Erwartungen der Gesellschaftsmitglieder abhängig sind, zumindest in dem Sinne, dass sie nur dann Bestand haben, wenn sie von einem hinreichend großen Teil der betreffenden Population in hinreichend ähnlicher Weise geteilt, verwendet und dadurch reproduziert werden.

Im Gegensatz zu dieser Skizze ist die Verwendung vergesellschafteter Formen der Verhaltenskoordination in der VKI dadurch gekennzeichnet, dass es keinen Reproduk-

tionszusammenhang zwischen den Verhaltensdispositionen der einzelnen Agenten und den Interaktionsstrukturen der Multiagentensysteme gibt. Dort wo wir Kommunikationsprotokolle zwischen Agenten beobachtet haben, die den symbolisch generalisierten Interaktionsmedien auf den ersten Blick ähneln, musste bei genauerer Betrachtung jeweils die Einschränkung hinzugefügt werden: Es fehlt den Agenten hinsichtlich der verwendeten Interaktionsstrukturen die Möglichkeit auch anders handeln zu können, da diese in der Regel fest vorgegeben sind. Aus einer Perspektive, die sich für soziale Koordinationsmechanismen unter dem Gesichtspunkt ihrer technisch effizienten Nachbildung interessiert, ist diese Vorgehensweise, wie wir gesehen haben, gut begründet. Der Reproduktionszusammenhang zwischen individuellem Handeln und gesellschaftlicher Struktur, der für das soziologische Nachdenken über menschliche Sozialität von zentraler Bedeutung ist, wird damit jedoch durchtrennt. Dies ist im Auge zu behalten, wenn in diesem Abschnitt von individuellen Einstellungen als Ergänzung generalisierter Erwartungen – die in der VKI eben keine Erwartungsstrukturen sind, sondern Regeln, die die Agenten zu befolgen gezwungen sind – die Rede ist.

Überlegungen, die individuelle Verhaltensdispositionen von Agenten zum Ausgangspunkt für die Koordination zwischen Agenten machen, sind in der VKI eng verbunden mit dem Konzept intentionaler Agenten, das durch die sogenannten BDI-Architekturen („BDI“ steht für „belief, desire, intention“) repräsentiert wird. Für die Strukturierung sozialer Interaktion ist die Frage, in welchem Sinne das Verhalten einer Person als sein Handeln betrachtet werden kann, ohne Zweifel von zentraler Bedeutung. Im alltäglichen Umgang wird dieses Problem dadurch bearbeitbar, dass die Interaktionspartner sich und einander intentionale Eigenschaften (Wünsche, Absichten und Überzeugungen) zuschreiben:

„When explaining human activity, it is often useful to make statements such as the following: Janine took her umbrella because she believed it was going to rain. Michael worked hard because he wanted to possess a PhD. These statements make use of a folk psychology, by which human behavior is predicted and explained through the attribution of attitudes, such as believing and wanting (as in the above examples), hoping, fearing, and so on. This folk psychology is well established: most people reading the above statements would say they found their meaning entirely clear, and would not give them a second glance.“ (Wooldridge/Jennings 1995: 4)

Diese Art und Weise Handlungen zu erklären, ist uns mit Blick auf den größten Teil unserer Interaktionen mit anderen Menschen derart selbstverständlich, dass es uns, wie es der im folgenden zitierte Gesprächspartner ausdrückt, gleichsam als der natürliche Weg gilt, Handlungen zu beschreiben. Wenn es aber solche intentionalen Einstellungen sind, mit denen menschliche Akteure ihre eigenen Handlungen und diejenigen anderer Akteure im Alltag unproblematisch zu erklären vermögen und die sie der wechselseitigen Koordination ihrer Handlungen zu Grunde legen, so ist es naheliegend, Multiagentensysteme in ähnlicher Weise auf der Basis intentionaler Agenten zu konzipieren:

„Also fangen wir jetzt an zu überlegen: Wie geht das beim Menschen? Schon bin ich noch nicht in der Soziologie, sondern erstmal beim einzelnen Agenten in der Psychologie und der Philosophie und fange halt an: Was bestimmt das Handeln des Menschen? Erstmal ein bestimmtes Wissen: *belief*. Dann haben wir bestimmte Aufgaben, Aufträge oder auch Ziele, je nachdem. Das sind die berühmten *desires* oder *goals*. Und dann ist da noch irgendwo, wenn es komplexer wird, wenn man mehrere

solcher Ziele gleichzeitig verfolgen will ... Dann muss er sehen, dass er nicht in Konflikte kommt, dann muss er irgendwie Entscheidungen treffen: *commitments* oder *intentions*, Absichten. Das ist dann so die berühmte BDI-Architektur. (...) Ja, das ist sozusagen, das ist einfach ein natürlicher Weg, so etwas zu beschreiben.“ (UP1, 4: 21-32)

Der Reiz dieses Ansatzes besteht dabei darin, dass man eine Vorgehensweise wählt, die sich direkt am normalen Leben des menschlichen Miteinanders orientiert und deshalb in hohem Maße eine intuitive Nachvollziehbarkeit besitzt:

„[Eine Motivation für intentionale Agenten ist,] dass man die Agententechnologie begreift als eine Entwurfsmethode, die aus dem normalen Leben resultiert sozusagen. Man versucht also, noch stärker als in der objektorientierten Modellierung, eine viel stärkere Abbildung der menschlichen Welt auf dem Rechner tatsächlich hinzukriegen. Das heißt, dass man in Begrifflichkeiten wie *commitments*, also irgendwelchen Versprechen, denkt. Oder was der Agent glaubt, das im Augenblick richtig ist. Dass er nicht nur irgendwie mit Daten versorgt ist, sondern halt mit Wissen versorgt ist. Dass er weiß wie Daten in einem Verhaltenskontext interpretiert werden und interpretiert werden müssen. Und dass man bei der Entwicklung von komplexen verteilten Systemen eigentlich mit dieser menschenähnlichen Denke daran geht das System zu entwickeln.“ (UW4, 8: 50 - 9: 1)

Entsprechende Definitionen intentionaler Agenten im Rahmen von BDI-Architekturen, die sich innerhalb der VKI inzwischen in großer Zahl finden, lauten dann etwa folgendermaßen:

„An agent is an entity whose state is viewed as consisting of mental components such as beliefs, capabilities, choices, and commitments. These components are defined in a precise fashion and stand in rough correspondence to their common sense counterparts.“ (Shoham 1993: 52)

Ausgehend von diesen alltagspsychologischen Grundannahmen ist in der VKI inzwischen ein Korpus von Forschungsarbeiten entstanden, die sich darum bemühen, das alltagssprachliche Vokabular intentionaler Begriffe formallogisch zu reformulieren und konzeptionell zu klären, auf welche Weise Überzeugungen und Wünsche Handlungsabsichten und Absichten Handlungen strukturieren. Diese Bemühungen stützen sich stark auf die analytische Handlungsphilosophie, insbesondere auf die Arbeiten von Dennett (1987) und Bratman (1987; 1990). Die bekanntesten Ansätze stammen von Cohen und Levesque (1990), Rao und Georgeff (1991), Werner (1989) und Shoham (1993). Sie werden zusammenfassend unter dem Begriff der BDI-Architekturen gefasst (vgl. Haddadi/Sundermeyer 1996).

Zugleich wirft das Konzept intentionaler Agenten die Frage nach der Handlungskoordination mit Blick auf nur gemeinsam zu realisierende Problemlösungen neu auf, nämlich als Frage danach, wie man eigensinnige Agenten – Agenten mit ihnen eigenen Zielen und Absichten – dazu bringt, koordiniert handeln zu wollen. Dieses Problem ist in der Literatur unter Stichworten wie „social commitments“ (Castelfranchi 1995), „joint commitments“ (Jennings 1996), „joint intentions“, „mutual beliefs“, „joint goals“ (Rao et al. 1992) u.Ä. diskutiert worden. Das Bestreben richtet sich dabei im Wesentlichen darauf, die Handlungskoordination auf mentale Zustände der beteiligten Agenten zurückzuführen. Koordiniertes Handeln wird dann beispielsweise mittels gemeinsamer Intentionen erklärt, die ihrerseits „as definable in terms of the conjunction of individual intentions together with mutual beliefs about the intentions of other individuals“ (Rao et al. 1992: 58) betrachtet werden. Oder aber das koordinierte Handeln wird als Resultat der gemeinsamen Verpflichtungen von Agenten konzipiert, die wie-

derum auf Verpflichtungen der je einzelnen Agenten anderen Agenten gegenüber zurückgeführt werden, unter bestimmten Bedingungen bestimmte Handlungsbeiträge zu erbringen (vgl. Jennings 1996: 196).

Das entscheidende Problem aller dieser Bemühungen besteht darin, dass – auch wenn es in plausibler Weise gelingt Beschreibungen zu entwickeln, die das gemeinsame Handeln in Multiagentensystemen auf die mentalen Zustände der beteiligten Agenten zurückführen, und es erfolgreich gelingt Agenten zu konstruieren, deren Interaktion auf der Grundlage solcher mentaler Zustände stattfindet – damit noch kein Prozess oder Mechanismus angegeben ist, der dazu führt, dass in irgendeiner Weise tatsächlich gemeinsame Überzeugungen, Absichten, Ziele und Verpflichtungen entstehen. Dieses Dilemma des BDI-Konzeptes, einerseits eine intuitiv plausible Beschreibung zu liefern, wie Handlungen zu Stande kommen, damit andererseits aber noch keine inhaltlich gehaltvollen Kriterien der Handlungswahl angegeben zu haben, kommt in der folgenden Äußerung deutlich zum Ausdruck:

„Ich bin der Meinung, dass man, um in Anführungszeichen höhere Formen der Intelligenz zu modellieren, tatsächlich so eine symbolische Repräsentation [des Glaubens und des Wissens von Agenten] braucht. Und dort – und da geht natürlich dann das Glaubensbekenntnis irgendwann mal los – sind die für mich bisher überzeugendsten Theorien die BDI-Theorien, also Belief-Desire-Intention-Theorien. Und dann kommt man natürlich sofort dahin, dass man sagt: Okay, was sind dann die Fähigkeiten des Agenten, also seinen Glauben bzw. sein Wissen zu repräsentieren, es vernünftig zu verwalten: Er hat Ziele, er kann Pläne erstellen, die zu diesen Zielen führen, er kann also ein *reasoning* betreiben: Was sind sinnvolle Aktionen, die zielführend sind? [...] Die Frage ist aber natürlich: Wie leitet ein Agent Ziele ab? Wie entstehen Ziele? und so weiter. Aber trotzdem gehört das für mich für eine höhere Form des Agenten-Daseins dazu, dass eben diese Fähigkeiten da sind. Und eben dann, wenn man das hat: Was dann tatsächlich zielführend ist, sind die Intentionen, d.h. die eigentlichen Absichten, die der Agent dann tatsächlich ausbildet. Deshalb, von einem intuitiven Standpunkt aus gesehen, [...] überzeugt mich diese Sicht am meisten.“ (FW1, 7: 10-35)

Ein Ausdruck dieses Problems scheint zu sein, dass die BDI-Architekturen in der akademischen Grundsatzdiskussion über wünschenswerte Fähigkeiten und Eigenschaften von Agenten einen prominenten Platz einnehmen, sie in der konkreten Konzeption von Multiagentensystemen dagegen von deutlich geringerer Bedeutung sind. Vielfach gewinnt man den Eindruck, dass das BDI-Konzept eher dazu dient, die eigenen Multiagentensysteme im Nachhinein intuitiver beschreiben zu können und die geleistete Arbeit zu systematisieren als umgekehrt:

„Da haben wir natürlich erstmal das irgendwie umgesetzt, programmiert (es geht hier um Software-Agenten für den RoboCup, den Computer-Fußball, Anm. d. Verf. ), und haben danach zum Schluss nur mal gesehen: Wo ist eigentlich jetzt das passende Modell dafür? Da haben wir dann gesehen, diese BDI-Architektur passt sehr gut darauf. Und jetzt sind wir dabei [...] einfach das nochmal ganz klar zu fassen und nach diesen Prinzipien zu organisieren, das Programm.“ (UP1, 5: 33-37)

„Und hinterher haben wir uns dann überlegt: Ein Stück war Brooks-inspiriert, ein Stück war BDI-inspiriert. Nachdem wir es nachher fertig hatten, haben wir gesagt: ‚Also BDI passt am besten bei uns.‘ Also wir haben das natürlich – wenn man den Programmcode ansieht – nicht durchgängig sauber [nach Maßgabe der BDI-Architektur] implementiert. [...] Und das [gilt] ja auch auch für Menschen. BDI ist ja auch ein Konstrukt. Ich meine, wir machen es ja auch nicht dauernd so, dass wir uns genau nach BDI verhalten. Tun wir ja nicht.“ (UP1, 39: 26-35)

Nichtsdestotrotz finden sich durchaus eine Reihe von Vorschlägen, die darauf gerichtet sind, Mechanismen anzugeben, wie auf der Grundlage von BDI-Agenten koordiniertes

Handeln entstehen kann. Nicht zufällig ist der profilierteste Vorschlag der der Agentenkoordination durch rationale Wahl. Deshalb nicht zufällig, weil der Theoriekontext aus dem dieser Vorschlag stammt (Rational Choice, Spieltheorie) in ähnlicher Weise wie die BDI-Architektur vom intentional handelnden Individuum ausgeht. Die Grundidee der Koordination durch rationale Wahl ist, dass die einzelnen Agenten ihr Handeln ausschließlich an eigenen Nutzenkalkülen orientieren. Sie sind dem homo oeconomicus nachgebildet. Gemeinschaftliches Handeln ist unter diesen Prämissen für die Agenten dann rational, wenn aus der Kooperation eigener Nutzen gezogen werden kann, also ein angestrebtes Ziel durch individuelles Handeln nicht oder nur zu höheren Kosten realisiert werden kann. Dementsprechend sind Agenten dann bereit zu kooperieren, "if, and only if, each member of a coalition gains more if he joins the coalition than he could gain before" (Shehory/Kraus 1995: 57).

Während hier die Eigenschaft egoistischer Nutzenmaximierung gleichsam als anthropologisch-agentologische Grundkonstante vorausgesetzt wird, besteht die umgekehrte Vorgehensweise darin, den Agenten einen allgemeinen Altruismus zu unterstellen (dies ist die bereits angesprochene *benevolent agent assumption*), etwa in Form einer allen Agenten eigenen Verpflichtung, andere Agenten zu unterstützen. Andere Varianten dieser Herangehensweise sind: die Agenten in einer Weise mit Wünschen und Bedürfnissen auszustatten, die die Entstehung auf ein gemeinsames Ziel bezogener Handlungsabsichten begünstigt; die Agenten mit gemeinsamen Überzeugungen ausgestattet werden, aus denen gemeinsame Handlungsziele resultieren; oder die Agenten direkt mit gemeinsamen bzw. einander ergänzenden Zielen auszustatten (vgl. Jennings 1996).

Ein dritter Vorschlag geht von der Beobachtung aus, dass das Wissen der Individuen darüber, wie man sich anderen Individuen gegenüber angemessen verhält, zu einem guten Teil sozialisatorisch vermittelt wird. Ausgehend hiervon gibt es bereits eine ganze Reihe von Ansätzen, Assistenten-Agenten gleichsam zu sozialisieren, also in der Interaktion mit ihren Benutzern dazu zu bringen, deren Interessen und Wünsche zu erkennen:

„Die Vorstellung, die dahinter (hinter dem Begriff der Sozialisation, Anm.) steckt, ist, dass solche Softwareagenten sich ohne Probleme in menschliche Gesellschaften einklinken können sollen. Äh, also meine Lieblingsmetapher dazu sind die elektronischen Assistenten. Also eine Vision ist, dass jeder auf seinem Rechner einen persönlichen Assistenten hat, der gewisse Aufgaben einfach übernimmt, wie z.B. Termine auszuhandeln. Ähm, und diese, diese Programme müssen sozusagen sozialisiert werden, d. h. dieser Agent muss sich wie ein guter Butler sozusagen an meine Bedürfnisse und Veränderungen anpassen, automatisch, nach Möglichkeit. Ich möchte mich mit dem nicht in kryptischer Kontrollsprache unterhalten, sondern ich möchte nach Möglichkeit irgendwie sowas eingeben wie: ‚Ich brauche morgen nachmittag einen Termin mit meiner Arbeitsgruppe‘ oder so etwas und das muss vernünftig interpretiert werden und und ausgewertet werden. Ich möchte auch keine kryptischen Antworten haben, sondern vernünftige Antworten, so wie ich das von einem Mitarbeiter z.B. erwarten würde. (...) In dieser Art und Weise müssen die Softwareprogramme sozialisiert werden, dass sie sich an gewisse soziale Regeln und Umgangsformen einfach anpassen. Und einfach so gestrickt sein müssen, dass sie das tun.“ (UW4, 8: 20-50)

Entsprechende Überlegungen, die sich als Ansätze einer beginnenden Hybridisierung von Interaktionszusammenhängen zwischen menschlichen Akteuren und künstlichen Agenten interpretieren lassen, haben wir schon weiter oben kennen gelernt (vgl. S. 52).

Die darüber hinausgehende, gegenwärtig sicherlich gleichermaßen visionäre Idee ist, auch die Agenten untereinander durch eine Art Sozialisationsprozess aufeinander einzustellen:

„Also die Sozialisation sehe ich nicht nur zwischen Mensch und Maschine, sondern die gleichen Konzepte müssen natürlich auch greifen zwischen der Software oder zwischen den Maschinen. Das heißt also, es halt ja keinen Sinn, wenn ich sozusagen meinen persönlichen Agenten sozialisiert habe, in Anführungszeichen, und dieser Softwareagent sich im weiteren Verlauf mit einem anderen Assistenten zur Verabredung eines Termins kurzschließt und dieser Termin ausgehandelt wird ... Und man muss wie in einer menschlichen Verhandlung diesen Verhandlungsprozess ebenfalls beschreiben und wenn einer der anderen Assistenten ein *commitment* gemacht hat und gesagt hat: ‚Das ist es, das ist der Termin‘, dann soll es nicht nur so sein, dass er sich gegenüber seinem Benutzer ‚commitet‘ und sozial verhält, sondern natürlich auch gegenüber dem anderen Software-[Agent], gegenüber dem anderen Programm. Das heißt also, diese Sozialisation muss natürlich auch zwischen den Maschinen oder zwischen den Programmen genauso eingehalten werden.“ (UW4, 9: 20-35)

Die implizite Strategie der beiden ersten Lösungsvorschläge besteht darin, Aspekte einer vorgegebenen Interaktionsstruktur in die mentalen Zustände der beteiligten Agenten hineinzuprogrammieren: Individuell nutzenmaximierende Agenten setzen Handlungszusammenhänge voraus, deren Struktur durch Konkurrenz um Ressourcen bestimmt ist, also etwa den Markttausch. Koordination zwischen altruistischen Agenten dagegen setzt Interaktionszusammenhänge voraus, die beispielsweise durch gemeinsam geteilte Werte strukturiert werden. Dies evoziert den Einwand, dass man den gewünschten Koordinationseffekt unter Umständen einfacher haben kann, wenn man die entsprechenden Strukturvorgaben direkt im Kommunikationsprotokoll entsprechender Multiagentensysteme festlegt, ohne den Umweg über die mentalen Zustände der einzelnen Agenten zu gehen:

„Wenn ich diese ganze Intensionslogik einsetzen will, das macht nur dann Sinn, wenn der Agent darüber (d.h. über seine Intentionen und die anderer Agenten, Anm. d. Verf.) nachdenkt. ... In unseren Anwendungen brauchen wir das nicht. Der Agent braucht nicht darüber nachzudenken. Er hat ein Ziel, er verfolgt es. Entweder er scheitert oder scheitert nicht. Dann versucht er etwas Neues. Von daher ist diese Intensionslogik ein Beispiel, das typisch ist für Sozialmetaphern. Die werden halt nicht benötigt. Es wäre wieder etwas anderes, wenn wir in die Situation kommen, wo wir zwei konkurrierende Agenten haben, die ihre eigenen Ziele verfolgen. ... Wie gesagt in diesem Stadium sind wir noch nicht.“ (IW1, 12: 25-42)

Aus der Perspektive derjenigen VKIler, die die Konstruktion von Multiagentensystemen unter dem Gesichtspunkt der Entwicklung effizienter Lösungen für verteilte Probleme betreiben, kommt das Konzept intentionaler Agenten mithin erst dann in Betracht, wenn es – aus welchen Gründen auch immer – nicht mehr möglich ist, koordiniertes Verhalten mittels vorgegebener Kommunikationsprotokolle sicherzustellen. Hier ist die Devise die, sich solange wie möglich mit reaktiven Agenten zu begnügen, solchen also, die nach vorgegebenen Regeln auf definierte Situationen mit bestimmten Verhaltensweisen reagieren, und erst dann, wenn sich das fragliche Problem auf diese Weise nicht lösen lässt, auf die deliberative Ebene intentionaler Agenten überzuwechseln.

Diejenigen unserer Gesprächspartner dagegen, die das Ziel konkreter technischer Innovationen zugunsten des grundsätzlicheren Interesses an den Entwicklungsmöglichkeiten von Multiagentensystemen zurückstellen, beurteilen das Potential des Kon-

zeptes intentionaler Agenten sehr viel optimistischer. Dies gilt nicht zuletzt mit Blick auf die Erwartung, auf dieser konzeptuellen Grundlage werde es möglich werden, das Problem der Konstruktion von Agentengesellschaften ausgehend vom einzelnen Agenten in den Griff zu bekommen:

„Meine Hoffnung ist, dass wir aus diesen BDI-Mechanismen, ich sage mal, für das einzelne Individuum, dass wir jetzt für gesellschaftliche Vorgänge in solchen [Multiagenten-]Gesellschaften dann auch die richtigen Begriffe finden werden.“ (UP1, 51: 41-45)

Solange allerdings die Modellierung der Fähigkeiten und Eigenschaften der einzelnen intentionalen Agenten einerseits und der Strukturmerkmale von Multiagentensystemen andererseits derart voneinander entkoppelt stattfindet, wie dies gegenwärtig der Fall ist, wird diese Hoffnung unerfüllt bleiben. Ein entscheidender Gesichtspunkt dabei ist, dass es – als Resultat des verschiedentlich angesprochenen umgekehrten Komplexitätsgefälles in menschlichen Gesellschaften und in Multiagentensystemen – unter ingenieurtechnischen Gesichtspunkten sehr gute Gründe dafür gibt, Interaktionsstrukturen in einer Weise zu implementieren, die sie unabhängig machen vom Wünschen und Wollen der beteiligten Agenten. Und in der umgekehrten Richtung ist auch das BDI-Konzept nicht unbedingt der geborene Kandidat für den Brückenschlag zwischen individuellem Handeln und gesellschaftlicher Strukturierung. Denn es entstammt einer Theorietradition, die sich weitgehend damit begnügt hat, die Handlungsmotive der individuellen Akteure als gegebene Bedürfnisprädispositionen zu behandeln, ohne die möglichen gesellschaftlichen Einflüsse auf die Ausbildung der intentionalen Befindlichkeiten des Individuums hinreichend in Rechnung zu stellen.

Unter diesen Voraussetzungen steht zu erwarten, dass die Reichweite des Konzepts intentionaler Agenten in absehbarer Zeit darauf beschränkt bleiben wird, vorgegebene Kommunikationsprotokolle zu flexibilisieren. Ein Beispiel hierfür ist das bereits angesprochene Vorgehen, Kontraktnetz-gesteuerte Aushandlungsprozesse durch Kommunikation über spezielle Wünsche oder Absichten der potentiellen Tauschpartner an die je konkreten Gegebenheiten einer Tauschsituation anzupassen. Dies schließt natürlich nicht aus, dass sich eine theoretische VKI entwickeln könnte, die sich zum Ziel setzt, eine autonome Evolution von Agentengesellschaften zu modellieren – etwa auf der Grundlage des Wechselverhältnisses zwischen individuellen und verallgemeinerten Verhaltenserwartungen. Die meisten der von uns befragten VKI-Forscher sind einer solchen Forschungsperspektive allerdings eher abgeneigt. Sie wollen sichergestellt wissen, dass die Agenten in zweckgerichteter Weise tun, was sie tun sollen, und dass heißt immer auch, dass das erwünschte Verhalten der Agenten durch Strukturvorgaben erzwungen wird:

„Wenn ich jetzt zum Beispiel so eine BDI-Architektur entwickeln würde und würde jetzt einfach nicht im Kopf behalten, dass sich so ein Agent eben nicht Ziele vornehmen darf, die miteinander im Widerspruch stehen. Oder genauer gesagt, dass er bei seinem Plänen immer auf die Konsistenz achten muss. Das wäre so zum Beispiel so eine Sache. [...] Dass ich also irgendwann anfangen über Lernen in Agentengemeinschaften zu reden. Und ich dann glaube, es funktioniert dann alles wunderbar, wenn jeder Agent für sich alleine irgendwas gelernt hat. Und in Wirklichkeit produzieren sie das Chaos, weil die sich eben nicht aufeinander einstellen.“ (UP1, 38: 33-42)



In dieser und ähnlichen Äußerungen kommt zum Ausdruck, dass sich die Forschungspraxis der VKI – jedenfalls die des überwiegenden Teils der deutschen VKI-Forscher – von der der Mutterdisziplin KI in einer Hinsicht unterscheidet. In der KI hat sich neben der auf Technikinnovation gerichteten Forschung ein theoretisch-kognitionswissenschaftlicher Zweig der Disziplin herausgebildet. Die VKI dagegen profiliert sich in der Hauptsache durch den Anspruch, mithilfe der Agententechnologie leistungsfähigere Problemlösungstechniken entwickeln zu können.

## **5.4 Erfahrungsbasiertes Verhalten und Routinen**

Soziale Strukturen haben – wenn auch sicherlich in unterschiedlichem Umfang – stets den Effekt Handeln zu entlasten, indem sie Erwartungssicherheit schaffen: Man muss in strukturierten Handlungszusammenhängen eben nicht mehr jeweils von neuem grundlegend über die vermuteten Absichten und Handlungsoptionen der Beteiligten (inklusive der eigenen) nachdenken, um zu einer Handlungsentscheidung zu gelangen. Sondern man verlässt sich darauf, dass die Situation eine Ausprägung einer typischen Interaktionssequenz ist, für die man bereits mehr oder weniger genau weiß, wie sich die Beteiligten verhalten werden. Auf der Ebene vergemeinschafteten Handelns, so wie ich den Begriff eingeführt habe, entsteht diese Erwartungssicherheit wesentlich auf dem Wege der Ausbildung von Gewohnheiten bzw. Routinen, die ihrerseits ein Resultat kumulierter Erfahrungen über erfolgreiche oder missglückte Interaktionen in der Vergangenheit sind.

In der soziologischen Theorie, und hier insbesondere im Rahmen der Praxistheorie (vgl. Bourdieu 1979; Giddens 1992), gibt es eine Reihe von Ansätzen, die die Entstehung gemeinsamer Gepflogenheiten des Handelns geradezu als den wichtigsten Mechanismus gesellschaftlicher Strukturbildung ansehen.<sup>16</sup> Ob sich diese These umstandslos auch auf die Strukturen vergesellschafteten Handelns ausdehnen lässt, kann allerdings mit einigem Recht bezweifelt werden (vgl. Schulz-Schaeffer 2000: 176ff.). Aber selbst, wenn sie nur für den Bereich vergemeinschafteten Handelns Gültigkeit beanspruchen kann: Bemerkenswert ist die These, weil sie einen Mechanismus der Strukturbildung angibt, der nicht selbst auf externen Strukturvorgaben beruht. In genau diesem Sinne ist die Beschäftigung mit Routinen offensichtlich auch für die VKI interessant, die gegenwärtig größte Schwierigkeiten hat, Strukturbildung in Multiagentensystemen anders als durch Strukturvorgaben hinzubekommen:

---

<sup>16</sup> Der im Folgenden zitierte Gesprächspartner scheint diese sozialtheoretische Position zu teilen: „Und wie funktioniert es, dass sich eine menschliche Gesellschaft ohne sich abzusprechen doch immer wieder an bestimmten Regeln orientiert. Das ist ja gar nicht so, dass wir unsere Verhaltensregeln alle mal codiert bekommen. [...] Und jetzt kann mir einer tausendmal erzählen, wie ich mich verhalten muss, was weiß ich, beim Empfang beim Bundespräsidenten. Und ich werde immer unsicher sein. Aber jetzt, wenn ich das meinetwegen 20-mal mitgemacht habe, dann ist das Routine geworden. Aber da hat mir keiner ein Modell gemacht, dann habe ich einfach Erfahrungen kumuliert. [...] Das heißt, dass diese ganze Stabilisierung in der menschlichen Gesellschaft wahrscheinlich vor allem über Erfahrung geht.“ (UP1, 52: 13-32)

„Das ist ja überhaupt die Frage, was Kooperation ist. [...] Heißt Kooperation nun immer, dass ich mir dessen bewusst bin, dass ich kooperiere? Die Menschen kooperieren auch miteinander, ohne dass ich das ausmachen kann. Also wir tun viele Dinge, einfach weil wir das so gewohnt sind oder weil man uns das so gesagt hat oder weil wir festgestellt haben, das funktioniert so. Und wir sprechen uns nicht ab. Es gibt implizite oder explizite Regeln dabei, aber ... Und so ist das hier [beim Computer-Fußball] auch. Also das ist oft gerade in der VKI ein ganz spannendes Thema: Wie weit muss man jetzt solche Kooperationsmechanismen vorgeben? Oder wie weit kann das wie im Ameisenstaat so funktionieren, dass jede Ameise einfach ihre Sache macht? Nun legt die andere Ameise nun hier ein bisschen Duftstoff ab und die nächste Ameise reagiert darauf irgendwie. Und dann bauen sie alle [zusammen auf diese Weise] ihren Bau.“ (UP1, 8: 31-46)

Es kommt nicht von ungefähr, dass in diesem Zusammenhang von Ameisen die Rede ist. Ameisengesellschaften und deren Nachbildung mittels Multiagenten-Simulation (vgl. Drogoul/Ferber 1994; Ferber 1996) sind zu einem Paradebeispiel in der Diskussion um die bereits angesprochenen reaktiven Agenten geworden. Dem Verweis auf Ameisen liegt mithin die folgende Zuordnung zu Grunde: Dem bewusst handelnden menschlichen Akteur entspricht der intentionale Agent, dem gewohnheitsmäßig verhandelnden Menschen entspricht dagegen der reaktive Agent. Auf der Grundlage dieser Unterscheidung lässt sich Routinenbildung in Multiagentensystemen dann beschreiben als ein Vorgang, in dem Handlungsabläufe, die von den Agenten zunächst auf der Grundlage ihrer intentionalen Kalküle als geplante Aktionen ausgeführt werden, zu reaktiven Verhaltensweisen werden, wenn sich herausstellt, dass sie als immer wieder gleich durchzuführende Handlungsanforderungen vorkommen:

„Wenn das System lange genug mit den Aufgaben, mit denen es betraut ist, rumhantiert hat, dann werden die Aufgaben für das System irgendwie routiniert. Also es hat schon diese Pläne, die da überhaupt auftauchen, die hat es schon alle gebildet und kann jetzt entsprechend die Situationen und die Ziele, die da sind, direkt mit den Plänen verbinden. Das heißt die sind schon alle da, das kann da alles drunter gelegt werden, da kannst du sagen: ‚Okay, da ist es alles in der verhaltensbasierten Ebene drin, stammt aber, oder lässt sich nach außen hin spezifizieren als da war mal deliberative Funktionalität, die das explizit repräsentiert hat oder da war mal eine soziale Funktionalität, die das explizit repräsentiert hat.“ (FW3, 9: 58 - 10: 7)

In der hier angesprochenen Agentenarchitektur, die sich unter anderem dadurch auszeichnet, dass die Agenten in ihrem Verhaltensrepertoire über eine reaktive und eine deliberative (also auf intentionalen Eigenschaften beruhende) Ebene der Verhaltenswahl verfügen, ist der entscheidende Schritt der Routinenbildung, die Umwandlung von geplantem in reaktives Verhalten, jedoch noch nicht automatisiert:

„Momentan macht das halt am Anfang der Designer, weil er eben sagt: ‚Okay, wie sinnvoll ist es, dass ich dem [Agenten] bestimmte Operatorbeschreibungen eben in die Planbibliothek hänge oder dass ich eben bestimmte Sachen als reaktive Verhaltensmuster absenke?‘ Also der [Designer] übernimmt das praktisch oder steckt einen Teil seiner Intelligenz da irgendwie rein. Aber interessant wird es natürlich dann, wenn du sagst: ‚Der Agent soll lernen, soll allgemein solche Repräsentationen ausbilden, solches prozedurales Wissen irgendwie ablegen und dann auch natürlich diese Kompilationsvorgänge automatisch machen.‘ Aber so weit sind wir hier in der Gruppe halt nicht, dass wir das angehen wollen.“ (FW3, 11: 3-9)

Für Anwendungsfelder, in denen es klare Kriterien der Bewertung des Erfolgs oder Misserfolgs einer Aktion gibt, scheint die Routinisierung von Entscheidungsprozessen innerhalb von Multiagentensystemen auf dem Wege des Lernens aus Erfahrung allerdings durchaus gegenwärtig schon realisierbar zu sein. Dies jedenfalls konstatiert einer

unserer Gesprächspartner für die Anwendungsdomäne der Aufgabenverteilung zwischen Maschinen in der Fertigungssteuerung:

„Der Anlass zum Lernen ist, zum Beispiel, wenn die [Agenten] jetzt hundert Produkte fertigen sollen, dass die hundert Produkte in zu langer Zeit gefertigt wurden. Also dann gibt es erstmal von einem Externen die Aufforderung: ‚Das dauert zu lange! Die sollen das schneller machen.‘ Jede Maschine, könnten wir sagen, führt im Wesentlichen eine Statistik oder Lernen aus Erfahrung oder aus der Vergangenheit. [In dem Sinne,], dass sie zum Beispiel sagen: ‚Wenn ich eine Teilaufgabe weitergegeben habe, wenn ich die nicht selber gemacht habe, und habe sie der Maschine [Nummer] 3 gegeben, dann zeigt mir die Erfahrung, dass die mir recht schnell die Teillösung bringt. Weshalb ich mir jetzt einfach merke: Also in Zukunft ist die Maschine 3 vielleicht ein guter Kandidat für diese Teilaufgabe.‘ Das ist also ein Lernen aus Erfahrung oder aus Beobachten, wenn man so will. (...) Das ist praktisch in diesem Umfeld ein recht einfacher Ansatz. Die [Agenten] beobachten quasi die anderen, die beobachten das Geschehen um sich herum: Welche Maschine ist am schnellsten? Welche fällt besonders häufig aus? Die Maschine, die viele Ausfallzeiten hat, die wird dann eben möglichst vermieden oder nicht mit einbezogen in den ganzen Fertigungsprozess.“ (UW6, 8: 9-31)

Die Konstruktion von Agenten, die gleichzeitig über reaktive und über deliberative Verhaltensrepertoires verfügen, ist sicherlich ein entscheidender Entwicklungsschritt in Richtung der Möglichkeit, Mechanismen sozialer Strukturbildung agententechnologisch zu nutzen. Ein wesentliches Hindernis auf dem Wege dahin besteht allerdings darin, dass das Konzept reaktiver Agenten unter Referenz auf einfache tierische Lebewesen, Insekten etwa, entwickelt worden ist (vgl. Ferber 1996: 288). Hier bietet es sich an, Verhaltensmuster als genetisch festgeschriebene Reiz-Reaktions-Schemata zu betrachten, die sich im Verlauf des Evolutionsprozesses ausgebildet und stabilisiert haben, weil sie zu Interaktionsmustern zwischen Individuen der betrachteten Spezies geführt haben, die dieser das Überleben ermöglichen. Aus informatischer Perspektive ist es dann naheliegend, die Parallele zwischen genetisch und softwaretechnisch vorprogrammierten Abläufen zu ziehen und Multiagentensysteme zu konstruieren, die in ähnlicher Weise auf dem Zusammenspiel im Vorhinein festgelegter Verhaltensweisen der beteiligten Agenten beruhen.

Dagegen macht es soziologisch wenig Sinn, routinisiertes Handeln menschlicher Akteure in einer solchen Weise als vorprogrammiert zu begreifen. Die Grundlage des Handelns, das sich seiner Handlungsgründe nicht stets aufs Neue versichern muss, ist vielmehr die Entstehung und Verfestigung von Selbstverständlichkeiten, die auch un-selbstverständlich werden können. Routinisiertes Handeln wird mit anderen Worten durch die Handlungsoption begleitet, im Zweifelsfall auch anders handeln zu können, wenn beispielsweise bislang als gegeben betrachtete Umstände wegfallen, unter deren Voraussetzung die fragliche Routine erfolgreich war. Die Bildung, Reproduktion und Veränderung sozialer Strukturen durch Routinisierung setzt damit nicht nur die Möglichkeit der Verselbstständlichung ursprünglich direkt sinnhaft begründeten Handelns in Gestalt von Routinen voraus, sondern umgekehrt genauso die Möglichkeit, Routinen im Bedarfsfall auf ihre Sinnhaftigkeit hin zu überprüfen.

Die Mehr-Ebenen-Agentenmodelle in der VKI realisieren dagegen gegenwärtig ein eher additives Verhältnis reaktiver und deliberativer Verhaltenskomponenten, wie die folgende Äußerung illustriert, in der es wieder um das Fußballspielen von Software-Agenten geht:

„Also diese reaktiven Prozesse, die haben wirklich die Form: Perzeption kommt rein, aus diesen Perzeptionsdaten werden direkt Aktivierungsmuster für solche Verhalten abgeleitet und diese Verhalten fahren dann direkt Aktionen hoch. Also im Sinne wie: Der Ball liegt direkt vor mir, sofort wird so ein Reflex drunter zu schießen aktiviert und das Ding schießt von mir aus in die Wallachei. Okay, und darüber gibt es dann Prozesse, die eine Stufe höher sind, die sich aus diesen Daten ableiten, wie sie die darunterliegenden Prozesse irgendwie konfigurieren können. Sprich beispielsweise also beim Zielen: Wenn ich aufs Tor zielen will, dann muss ich diesen Kick-Reflex entsprechend umbiegen, muss dem halt sagen, wohin soll er schießen und wie fest soll er schießen. Und das ist insofern reaktiv, das heißt sobald der Ball da ist, schießt er direkt drunter. Das ist klar, das ist super reaktiv, da macht er keine Entscheidungsfindung mehr, aber es ist insofern auch noch deliberativ, etwas deliberativ, also es geht wirklich [darum], *smooth transition* da irgendwie hochzukriegen, dass eben über einen separaten nebenläufigen Zweig eben die Daten ausgewertet werden, die wichtig dafür sind, aufs Tor zu schießen.“ (FW3, 5: 36-47)

Es gibt hier mithin keine Einflüsse deliberativer Prozesse auf die Inhalte des reaktiven Handlungswissens oder umgekehrt, sondern nur Prozesse des beständigen „Überschreibens“ der feststehenden Reflexe durch deliberativ gewonnene Handlungsprogramme. Die Agenten sind gleichsam mental aufgerüstete Insekten, die ihre genetisch festgelegten Verhaltensmuster beständig mittels der hinzugewonnenen Planungsintelligenz korrigieren, ohne jedoch an der Determiniertheit ihrer unmittelbaren Reaktionen jemals etwas ändern zu können. Mehr-Ebenen-Agentenmodelle der hier beschriebenen Art lassen sich mithin zwar als ein erster Schritt in Richtung auf die Möglichkeit interpretieren, Prozesse der Routinebildung agententechnologisch nachzubilden. Noch aber fehlt die entscheidende Verbindung zwischen der deliberativen und der reaktiven Ebene, die es ermöglicht, dass wiederholt erfolgreiche Verhaltensweisen sich als Routinen ablagern bzw. wiederholt scheiternde Routinen Gegenstand erneuter Entscheidungsfindung werden.

## 6. Ergebnis

Eine alte techniksoziologische These besagt, dass menschliches Verhalten dort erfolgreich auf technische Maschinen übertragen werden kann, wo das menschliche Verhalten zuvor selbst technisch geworden ist, also in vorhersehbarer Weise zweckgerichtet festgelegten Abläufen folgt. Diese These, der Collins (1990) bei seinen Untersuchungen über die Grenzen und Möglichkeiten der Nachbildung menschlicher Expertise durch die Expertensysteme der KI eindringlich nachgegangen ist, scheint sich für das Forschungsfeld der VKI erneut zu bestätigen. Jedenfalls zeigen die hier aufgeführten Befunde, dass die Multiagenten-Forschung sich gegenwärtig in solchen Bereichen am weitesten entwickelt hat, in denen sie auf technisierte Formen der sozialen Verhaltensabstimmung zurückgreifen konnte. Dies ist die Grundlage des Erfolgs nicht nur des Kontraktnetz-Protokolls, sondern ebenso aller anderen Interaktionsformen zwischen Agenten, in denen durch Einschränkung der Verhaltensoptionen der beteiligten Agenten auf wenige in der jeweiligen Situation zulässige Verhaltensweisen sichergestellt wird, dass die jeweilige Interaktionssequenz mit einiger Sicherheit auf das gewünschte Ergebnis zuläuft.

Dieser Befund ist an sich nicht sonderlich überraschend. Er zeigt uns ein weiteres Mal, dass die Bereiche des Technischen und des Sozialen nicht in einem wechselseiti-

gen Ausschließungsverhältnis zueinander stehen, sondern dass soziale Zusammenhänge in einer Weise technisch strukturiert sein können, die dann den Anstoß für die Entwicklung entsprechend strukturierter maschineller Technik geben kann. Interessant wird die Beobachtung der agententechnologischen Nachbildung technisierter Formen sozialer Verhaltensabstimmung durch eine zusätzliche Feststellung: Im Bereich menschlicher Sozialität findet Technisierung im Kontext der Möglichkeit menschlicher Akteure statt, auch anders als erwartet zu handeln, sich also der Zumutung des Technischen zu verweigern. Aus diesem Grund müssen technische Formen sozialer Verhaltensabstimmung, um erfolgreich zu sein, stets eine Motivation mitliefern, die die involvierten Akteure dazu bringt, diese Handlungsoption nicht (oder wenigstens nicht sehr häufig) wahrzunehmen. Dies wird im Fall der symbolisch generalisierten Interaktionsmedien unter anderem durch die jeweils korrespondierenden Sanktionsinstanzen geleistet. Das wesentliche Kriterium der erfolgreichen Konstruktion herkömmlicher maschineller Technik dagegen ist, dass die Elemente, deren Zusammenwirken den technischen Ablauf ausmacht, im Normalfall nicht anders können als sich so zu verhalten, wie es zuvor zweckgerichtet festgelegt wurde. Aus der ingenieurwissenschaftlichen Perspektive stellt sich das Problem der Motivation der Elemente eines technischen Zusammenhanges deshalb nicht.

Betrachtet man die Ansätze der agententechnologischen Nachbildung vergesellschafteter Formen sozialer Verhaltensabstimmung vor dem Hintergrund dieser Differenz, so ist zu konstatieren, dass sich die VKI zunächst ganz selbstverständlich am ingenieurwissenschaftlichen Verständnis von Technik orientiert und das heißt hier: auf eine eigenständige Motivation von Agenten verzichtet, wo dies möglich ist. Möglich ist dies, wie wir gesehen haben, bei der Nachbildung vergesellschafteter Formen sozialer Verhaltensabstimmung innerhalb geschlossener Multiagentensysteme. Dann aber, so habe ich argumentiert, stellt sich die Frage, ob überhaupt in einem substantiellen Sinne von einer Nachbildung sozialer Interaktionsmechanismen gesprochen werden kann. Mit Blick auf die Interaktionsmedien Macht, Einfluss und Wertbindung jedenfalls muss man zu dem Schluss kommen, dass dies nicht der Fall ist. Sobald von vornherein ausgeschlossen ist, dass die beteiligten Agenten auch anders als erwartet handeln können, verlieren sie ihre jeweilige Spezifik als unterscheidbare Koordinationsmechanismen (vgl. insbesondere das Gedankenexperiment oben auf S. 59f.). Der Verweis auf Mechanismen zwischenmenschlicher Verhaltensabstimmung ist dann nicht sehr viel mehr als eine eingängige Darstellung der betreffenden technischen Abläufe, die tatsächlich aber in keinem gehaltvollen Sinne Formen menschlicher Sozialität nachbilden.

Umso interessanter ist die Beobachtung, dass sich die herkömmliche ingenieurwissenschaftliche Sicht auf Technik in der VKI nicht bruchlos fortschreibt. Dies hängt zum einen damit zusammen, dass eine Vielzahl zukunftssträchtiger Anwendungen der Multiagenten-Technik in Bereichen liegt, in denen die betreffenden Multiagentensysteme nicht mehr der Kontrolle eines Entwicklers bzw. Entwickler-Teams unterliegt. Damit entsteht Situationen, in der ein „Auch-anders-handeln-können“ der Agenten

nicht mehr grundsätzlich ausgeschlossen werden kann. Zum anderen zeichnet sich ab, dass es auch für Zusammenhänge, in denen eine entsprechende Kontrolle des Multiagentensystems durch seine Entwickler zwar möglich wäre, sinnvoll sein kann, sie nicht auszuüben. Dies ist insbesondere der Fall, wenn das Entwicklungsziel ist, das in der Multiagenten-Technologie enthaltene Potential der flexiblen Anpassung an sich verändernde Gegebenheiten auszunutzen.

Die zunehmende Bedeutung der ersten Entwicklungstendenz hat viel mit der Ausbreitung des Internet zu tun (vgl. oben S. 17f.). Denn es eröffnet nun auch mit Blick auf die Interaktion zwischen Agenten die Möglichkeit, vergesellschaftete Formen der sozialen Koordination dort einzusetzen, wo diese ihre besondere Leistungsfähigkeit unter Beweis stellen können: im weitgehend anonymen Austausch zwischen sozialen Akteuren (bzw. Agenten), die voneinander nicht viel mehr wissen müssen als die wechselseitige Erwartung, dass sie sich in durch das jeweilige Interaktionsmedium vorgezeichneter Weise berechenbar verhalten. Die darauf gerichteten Bestrebungen der Koordination von Agenten in offenen Multiagentensystemen lassen sich in einem gehaltvollen Sinne als Nachbildung spezifischer Abstimmungsmechanismen menschlicher Gesellschaften interpretieren. Denn in offenen Multiagentensystemen kann das regelkonforme Verhalten der beteiligten Agenten nicht direkt erzwungen werden. Auch hier tritt nun also die Situation ein, dass Motive bereitgestellt werden müssen, die es aus der Sicht der Agenten (bzw. ihrer Benutzer) vorteilhafter erscheinen lässt, die Erwartungen regelgerechten Verhalten zu erfüllen, die die Spezifik des jeweiligen Austauschverhältnisses ausmachen.

Die zweite Entwicklungstendenz, die der Erhöhung der Verhaltensautonomie der Agenten im Interesse situationsgerechter Problemlösungen, ist direkt auf die grundlegend forschungsleitende Hypothese der VKI zurückzuführen: auf die Annahme, dass viele Probleme von der Art sind, dass sie am besten durch Zusammenarbeit einer Mehrzahl lokal operierender Problemlösungskomponenten mit je eigener Teilkompetenz gelöst werden können. Nimmt man diese Annahme ernst, dann muss man den beteiligten Agenten einen entsprechenden Entscheidungsspielraum zubilligen. Dies erhöht aber auch das Koordinationsproblem, weil eine Vergrößerung von Verhaltensspielräumen zwangsläufig eine Verringerung der Rigidität von Koordinationsstrukturen mit sich bringt und umgekehrt. Insbesondere die unter dem Begriff der Vergemeinschaftung zusammengefassten Formen der Verhaltensabstimmung sind soziologisch interpretiert Versuche, unter der Bedingung, dass zunächst unsicher ist, wie sich die beteiligten Agenten verhalten, zumindest zeitweilig stabile Strukturen aufzubauen, der zu erhöhter Erwartungssicherheit führen.

Im einen Fall notgedrungen im anderen gewollt, entsteht für die Entwicklung von Multiagentensystemen damit eine Ausgangssituation, die der der Technisierung der Verhaltensabstimmung zwischen den Akteuren menschlicher Gesellschaften zumindest in Ansätzen vergleichbar ist. D.h. Technisierung ist nur im Kontext einer sozialen Strukturierung des Verhaltens der Beteiligten möglich, die nicht selbst Bestandteil des technischen Ablaufes ist. Die Technikentwicklung stößt mithin in einen Bereich vor, in

dem sich ihre Artefakte – die Agenten – zur Realisierung technisierter Zusammenhänge in einem nicht technischen Sinne sozial verhalten müssen. Und das ist etwas, was die Soziologie interessieren sollte.

## Literatur

- Austin, John L. (1962): How to do things with words, Cambridge Mass.: Harvard University Press
- Berger, Peter L. / Thomas Luckmann (1969): Die gesellschaftliche Konstruktion der Wirklichkeit, Frankfurt/Main: Fischer
- Bond, Alan H. (1990): A Computational Model for Organization of Cooperating Intelligent Agents, in: Proceedings International Workshop on Office Information Systems, Cambridge, MA, S. 21-30
- Bond, Alan H. / Les Gasser (1988): An Analysis of Problems and Research in DAI, in: A. H. Bond / L. Gasser (Hg.), Readings in Distributed Artificial Intelligence, San Mateo, Ca.: Morgan Kaufmann Publishers, S. 3-36
- Bourdieu, Pierre (1979): Entwurf einer Theorie der Praxis auf der ethnologischen Grundlage der kabyllischen Gesellschaft, Frankfurt/Main: Suhrkamp
- Bratman, Michael E. (1987): Intentions, Plans, and Practical Reason, Cambridge, Ma.: Harvard University Press
- Bratman, Michael E. (1990): What is Intention?, in: P. R. Cohen / J. Morgan / M. E. Pollack (Hg.): Intentions in Communication, Cambridge, MA: The MIT Press
- Burkhard, Hans-Dieter (1995): Agent-Oriented Programming for Open Systems, in: M. J. Wooldridge / N. R. Jennings (Hg.), S. 291-306
- Castelfranchi, Cristiano (1990): Social Power. A Point Missed in Multi-Agent, DAI and HCI, in: Y. Demazeau / J.-P. Müller (Hg.) (1990), Dezentralized AI, Proceedings of the First European Workshop on Modelling Autonomous Agents in a Multi-Agent World, Amsterdam u.a.: Elsevier Science Publishers (North-Holland), S. 49-63
- Castelfranchi, Cristiano (1995): Commitments: From Individual Intentions to Groups and Organizations, in: V. Lesser (Hg.), ICMAS-95. Proceedings of the First International Conference on Multi-Agent Systems, Menlo Park u. a.: AAAI Press/ The MIT Press, S. 41-48
- Castelfranchi, Cristiano / Rosaria Conte (1992): Emergent Functionality Among Intelligent Systems: Cooperation Within and Without Minds, in: AI & Society 6(1), S. 78-87
- Castelfranchi, Cristiano / Rosario Conte / Mario Paolucci (1998): Normative Reputation and the Costs of Compliance, in: Journal of Artificial Societies and Social Simulation 1(3), <http://www.soc.surrey.ac.uk/JASSS/1/3/3.html>
- Cohen, Michael D. / James G. March / Johan P. Olsen (1972/1990): Ein Papierkorb-Modell für organisatorisches Wahlverhalten, in: J. G. March (Hg.) (1990), Entscheidung und Organisation. Kritische und konstruktive Beiträge, Entwicklungen und Perspektiven, Wiesbaden, S. 329-371; zuerst veröffentlicht in: Administrative Science Quarterly 17(1) (1972)
- Cohen, Philip R. / Hector J. Levesque (1990): Intention is choice with commitment, in: Artificial Intelligence 42, S. 213-261
- Connah, David / Peter Wavish (1990): An Experiment in Cooperation, in: Y. Demazeau / J.-P. Müller (Hg.), Dezentralized AI, Proceedings of the First European Workshop on Modelling Autonomous Agents in a Multi-Agent World, Amsterdam u.a.: Elsevier Science Publishers (North-Holland), S. 197-212
- Crozier, Michel / Erhard Friedberg (1979): Macht und Organisation: Die Zwänge kollektiven Handelns, Königstein/Ts.: Athenäum
- Davis, R. (1980): Report on the Workshop on Distributed AI, in: Sigart Newsletter 73, S. 42-52
- Davis, R. / R. G. Smith (1983): Negotiation as a Metaphor for Distributed Problem Solving, in: Artificial Intelligence 20(1), S. 63-109



- Decker, Keith S. (1987): Distributed Problem-Solving Techniques: A Survey, in: IEEE Transactions on Systems, Man, and Cybernetics 17, S. 729-740
- Demazeau, Yves / Jean-Pierre Müller (1990): Decentralized Artificial Intelligence, in: dies. (Hg.), Dezentralized AI, Proceedings of the First European Workshop on Modelling Autonomous Agents in a Multi-Agent World, Amsterdam u.a.: Elsevier Science Publishers (North-Holland), S. 3-13
- Dennett, Daniel C. (1987): The Intentional Stance, Cambridge, MA: Bradford Books/The MIT Press
- Drogoul, Alexis / Jacques Ferber (1994): Multi-Agent Simulation as a Tool for Modeling Societies: Application to Social Differentiation in Ant Colonies, in: C. Castelfranchi / E. Werner (Hg.), Artificial Social Systems, 4th European Workshop on Modelling Autonomous Agents in a Multi-Agent World, MAAMAW '92, Lecture Notes in Artificial Intelligence 830, Berlin u.a.: Springer-Verlag, S. 3-23
- Durfee, Edmund H. / Victor R. Lesser / D. D. Corkhill (1987): Cooperation through Communication in a Distributed Problem Solving Network, in: M. N. Huhns (Hg.), Distributed Artificial Intelligence, Pitman, Ca. u.a.: Morgan Kaufmann Publishers, S. 29-54
- Durfee, Edmund H. / Victor R. Lesser / D. D. Corkhill (1989): Trends in Cooperative Distributed Problem Solving, in: IEEE Transactions on Knowledge and Data Engineering, Vol. I, No. 1, S. 63-83
- Ferber, Jacques (1996): Reactive Distributed Artificial Intelligence: Principles and Applications, in: G. M. P. O'Hare / N. R. Jennings (Hg.), Foundations of Distributed Artificial Intelligence, New York u.a.: John Wiley & Sons, S. 287-314
- Finin, Tim / Jay Weber / Gio Wiederhold / Michael Genesereth / Richard Fritzson / Donald McKay / James McGuire / Richard Pelavin / Stuart Shapiro / Chris Beck (1993): Specification of KQML Agent-Communication Language plus Example Agent Policies and Architectures (Draft), Working Paper of the DARPA Knowledge Sharing Initiative External Interfaces Working Group
- FIPA (1999): Agent Communication Language, FIPA Spec 2-1999
- Fischer, Klaus / Jörg P. Müller / Markus Pischel / Darius Schier (1995): A Model for Cooperative Transportation Scheduling, in: V. Lesser (Hg.), ICMAS-95. Proceedings of the First International Conference on Multi-Agent Systems, Menlo Park u. a.: AAAI Press/ The MIT Press, S. 109-116
- Fox, M. S. (1981): An Organizational View of Distributed Systems, in: IEEE Transactions on Systems, Man, and Cybernetics 11, 1981, S. 70-80
- Freeman, Christopher (1991): Networks of Innovators: A Synthesis of Research Issues, in: Research Policy 20, S. 499-514
- Gasser, Les (1992): An Overview of DAI, in: N. M. Avouris / L. Gasser (Hg.), Distributed Artificial Intelligence: Theory and Praxis, Dordrecht u.a.: Kluwer Academic Publishers, S. 9-30
- Gasser, Les / N. F. Rouquette / R. H. Hill / J. Lieb (1989): Representing and Using Organizational Knowledge in Distributed AI Systems, in: L. Gasser / M. N. Huhns (Hg.), Distributed Artificial Intelligence, Volume II, London: Pitman, San Mateo, Ca.: Morgan Kaufmann Publishers, S. 55-78
- Genesereth, M. R. / M. L. Ginsberg / J. R. Rosenschein (1988): Cooperation without Communication, in: A. H. Bond / L. Gasser (Hg.), Readings in Distributed Artificial Intelligence, San Mateo, Ca.: Morgan Kaufmann Publishers, S. 220-226
- Giddens, Anthony (1992): Die Konstitution der Gesellschaft. Grundzüge einer Theorie der Strukturierung, Frankfurt/Main u.a.: Campus Verlag
- Giddens, Anthony (1995): Konsequenzen der Moderne, übers. v. Joachim Schulte, 2. Aufl., Frankfurt/Main: Suhrkamp
- Gmytrasiewicz, Piotr J. / Edmund H. Durfee (1993): Toward a Theory of Honesty and Trust Among Communicating Autonomous Agents, in: Group Decision and Negotiation 2, S. 237-258

- Guenther, Oliver / Tad Hogg / Bernardo A. Huberman (1997): Market Organizations for Controlling Smart Matter, in: Proceeding of the International Conference on Computer Simulation and Social Sciences
- Habermas, Jürgen (1987): Theorie des kommunikativen Handelns, Vierte, durchgesehene Auflage, 2 Bde, Frankfurt/Main: Suhrkamp
- Haddadi, Afsaneh / Kurt Sundermeyer (1996): Belief-Desire-Intention Agent Architectures, in: G. M. P. O'Hare / N. R. Jennings (Hg.), Foundations of Distributed Artificial Intelligence, New York u.a.: John Wiley & Sons, S. 169-210
- Hendler, James / Danny Bobrow / Les Gasser / Carl Hewitt / Marvin Minsky (1991): Multiple Approaches to Multiple Agent Problem Solving, in: IJCAI-91, S. 553-554
- Hewitt, Carl E. (1986): Offices are Open Systems, in: ACM Transactions on Office Information Systems 4(3), S. 271-287
- Hewitt, Carl E. (1991): Open Information Systems Semantics for Distributed Artificial Intelligence, in: Artificial Intelligence 47, S. 79-106
- Huberman, Bernardo E. / Scott H. Clearwater (1995): A Multi-Agent System for Controlling Building Environments, in: V. Lesser (Hg.), ICMAS-95. Proceedings of the First International Conference on Multi-Agent Systems, June 12 -14, 1995 in San Francisco, California, Menlo Park u. a.: AAAI Press/ The MIT Press, S. 171-175
- Jennings, Nick R. (1996): Coordination Techniques for Distributed Artificial Intelligence, in: G. M. P. O'Hare / N. R. Jennings (Hg.), Foundations of Distributed Artificial Intelligence, New York u.a.: John Wiley & Sons, S. 187-210
- Kowol, Uli / Wolfgang Krohn (1995): Innovationsnetzwerke. Ein Modell der Technikgenese, in: J. Halfmann et al. (Hg.), Technik und Gesellschaft. Jahrbuch 8: Theoriebausteine der Techniksoziologie, Frankfurt/Main u.a.: Campus Verlag, S. 77-105
- Künzler, Jan (1989): Medien und Gesellschaft: Die Medienkonzepte von Talcott Pasons, Jürgen Habermas und Niklas Luhmann, Stuttgart: Ferdinand Enke Verlag
- Labrou, Yannis / Tim Finin (1997): A Proposal for a new KQML Specification, TR CS-97-03, February 1997, Computer Science and Electrical Engineering Department, University of Maryland Baltimore County, MD 21250
- Lenzmann, Britta / Ipke Wachsmuth (1997): Contract-Net-Based Learning in a User-Adaptive Interface Agency, in: G. Weiß (Hg.), Distributed Artificial Intelligence Meets Machine Learning. Learning in Multi-Agents Environments. Lecture Notes in Artificial Intelligence, Volume 1221, Berlin u.a.: Springer-Verlag, S. 202-222
- Lesser, Victor R. / D. D. Corkhill (1983a): Coordination in Distributed Problem-Solving, in: Sigart Newsletter 84, S. 9
- Lesser, Victor R. / D. D. Corkhill (1983b): The Distributed Vehicle Monitoring Testbed: A Tool for Investigating Distributed Problem Solving Networks, in: The AI Magazine 4, S. 15-33
- Lesser, Victor R. / D. D. Corkhill (1987): Distributed Problem Solving, in: S. C. Shapiro / D. Eckroth (Hg.), Encyclopedia of Artificial Intelligence, Vol 2, New York: John Wiley & Sons Inc., S. 245-251
- Lieberman, Henry (1995): Letizia: An Agent That Assists Web Browsing, in: Proceedings of the International Joint Conference on Artificial Intelligence, Montreal, August 1995
- Luhmann, Niklas (1975a): Macht, Stuttgart: Ferdinand Enke Verlag
- Luhmann, Niklas (1975b): Einführende Bemerkungen zu einer Theorie symbolisch generalisierter Kommunikationsmedien, in: ders., Soziologische Aufklärung 2. Aufsätze zur Theorie der Gesellschaft, Opladen: Westdeutscher Verlag, S. 170-192

- Luhmann, Niklas (1987): Soziale Systeme. Grundriß einer allgemeinen Theorie, Frankfurt/Main: Suhrkamp
- Luhmann, Niklas (1991): Soziologie des Risikos, Berlin: de Gruyter
- Luhmann, Niklas (1997): Die Gesellschaft der Gesellschaft, Frankfurt/Main: Suhrkamp
- Malone, Thomas W. (1987): Modeling Coordination in Organizations and Markets, in: Management Science 33(10), 1987, S. 1317-1332; auch in: A. H. Bond / L. Gasser (Hg.) (1988), Readings in Distributed Artificial Intelligence, San Mateo, Ca.: Morgan Kaufmann Publishers S. 151-158
- Malone, Thomas W. (1990): Organizing Information Processing Systems: Parallels between Human Organizations and Computer Systems, in: S. P. Robertson / W. W. Zachary / J. B. Black (Hg.), Cognition, Computation and Cooperation, Norwood, NJ: Ablex Pub, S. 56-83
- Malsch, Thomas (1998): Bekanntschaft, Anonymisierung, Versachlichung. Skalierung von Multiagenten-Systemen als gesellschaftliche Strukturbildung, in: ders. (Hg.), Sozionik. Soziologische Ansichten über künstliche Sozialität, Berlin: Edition Sigma, S. 255-295
- March, James G. (1962/1990): Die Unternehmung als politische Koalition, in: ders. (Hg.) (1990), Entscheidung und Organisation. Kritische und konstruktive Beiträge, Entwicklungen und Perspektiven, Wiesbaden, S. 115-130; zuerst veröffentlicht in: Journal of Politics 24 (1962), S. 662-678
- Marsh, Stephan (1994): Trust in Distributed Artificial Intelligence, in: C. Castelfranchi / E. Werner (Hg.), Artificial Social Systems, 4th European Workshop on Modelling Autonomous Agents in a Multi-Agent World, MAAMAW '92, Lecture Notes in Artificial Intelligence 830, Berlin u.a.: Springer-Verlag, S. 94-114
- Martial, Frank v. (1992a): Einführung in die Verteilte Künstliche Intelligenz, in: KI 1/1992, S. 6-11
- Martial, Frank v. (1992b): Coordinating Plans of Autonomous Agents, Berlin u.a.: Springer-Verlag
- Mintzberg, Henry (1989): Mintzberg on Management. Inside our Strange World of Organizations, New York: The Free Press
- Moulin, Bernard / Brahim Chaib-draa (1996): An Overview of Distributed Artificial Intelligence, in: G. M. P. O'Hare / N. R. Jennings (Hg.), Foundations of Distributed Artificial Intelligence, New York u.a.: John Wiley & Sons, S. 3-55
- Mumford, Lewis (1977): Mythos der Maschine. Kultur, Technik und Macht, Frankfurt/Main: Fischer
- Parsons, Talcott (1980): Zur Theorie der sozialen Interaktionsmedien, hg. von Stefan Jensen, Opladen: Westdeutscher Verlag
- Rao, Anand S. / M. P. Georgeff (1991): Modelling Rational Agents within a BDI Architecture, in: R. Fikes / E. Sandewall (Hg.), Proceedings of the 2nd International Conference on Principles of Knowledge Representation and Reasoning, S. 473-484
- Rao, Anand S. / M. P. Georgeff / Elizabeth A. Sonenberg (1992): Social Plans: A Preliminary Report, in: E. Werner / Y. Demazeau (Hg.), Decentralized A.I. 3, Proceedings of the Third European Workshop on Modelling Autonomous Agents in an Multi-Agent World, North-Holland: Elsevier, S. 57-76
- Rosenschein, J. R. / M. R. Genesereth (1988): Deals Among Rational Agents, in: A. H. Bond / L. Gasser (Hg.), Readings in Distributed Artificial Intelligence, San Mateo, Ca.: Morgan Kaufmann Publishers S. 227-234
- Schillo, Michael / Petra Funk / Michael Rovatsos (1999): Who Can You Trust: Dealing with Deception, in: C. Castelfranchi / Y. Tan / R. Falcone / B. S. Firozabadi (Hg.): Proceedings of the Workshop "Deception, Fraud and Trust in Agent Societies" of the Autonomous Agents Conference
- Schulz-Schaeffer, Ingo (1998a): Akteure, Aktanten und Agenten. Konstruktive und rekonstruktive Bemühungen um die Handlungsfähigkeit von Technik, in: Th. Malsch (Hg.): Sozionik. Soziologische Ansichten über künstliche Sozialität, Berlin: Edition Sigma, S. 128-167

- Schulz-Schaeffer, Ingo (2000): Sozialtheorie der Technik, Frankfurt/Main: Campus
- Schulz-Schaeffer, Ingo / Michael Jonas / Thomas Malsch (1997): Innovation reziprok. Intermediäre Kooperation zwischen akademischer Forschung und Industrie, in: W. Rammert / G. Bechmann (Hg.): Technik und Gesellschaft. Jahrbuch 9: Innovation - Prozesse, Produkte, Politik, Frankfurt/Main: Campus Verlag, S. 91-124
- Searle, John R. (1979a): Expression and Meaning: Studies in the Theory of Speech Acts, Cambridge u.a.: Cambridge University Press
- Shehory, Onn / Sarit Kraus (1995): Coalition Formation Among Autonomous Agents: Strategies and Complexity (Preliminary Report), in: C. Castelfranchi / J.-P. Müller (Hg.), From Reaction to Cognition. 5th European Workshop on Modeling an Agent in a Multi-Agent World, MAAMAW '93, Selected Papers, Berlin u.a.: Springer, S. 57-72
- Shoham, Yoav (1993): Agent-oriented Programming, in: Artificial Intelligence 60(1), S. 51-92
- Shoham, Yoav / M. Tennenholtz (1992): On the Synthesis of Useful Social Laws for Artificial Agents Societies (Preliminary Report), in: AAAI-92. Proceedings Tenth National Conference on Artificial Intelligence, Menlo Park u.a.: AAAI Press/The MIT Press, S. 276-281
- Störig, Hans Joachim (1987): Kleine Weltgeschichte der Philosophie, Frankfurt/Main: Fischer
- Weber, Max (1922/1972): Wirtschaft und Gesellschaft. Grundriß der verstehenden Soziologie, 5., revidierte Aufl., besorgt von Johannes Winckelmann, Tübingen: Mohr
- Weiß, Gerhard (Hg.) 1997): Distributed Artificial Intelligence Meets Machine Learning. Learning in Multi-Agents Environments. Lecture Notes in Artificial Intelligence, Volume 1221, Berlin u.a.: Springer-Verlag
- Werner, Eric (1989): Cooperating Agents: A Unified Theory of Communication and Social Structure, in: L. Gasser / M. N. Huhns (Hg.) (1989), Distributed Artificial Intelligence, Volume II, London: Pitman, San Mateo, Ca.: Morgan Kaufmann Publishers, S. 3-36
- Wooldridge, Michael J. / Nicolas R. Jennings (1995): Agent Theories, Architectures, and Languages: A Survey, in: M. J. Wooldridge / N. R. Jennings (Hg.), Intelligent Agents. ECAI-94 Workshop on Agent Theories, Architectures, and Languages, Amsterdam, The Netherlands, August 8-9, 1994, Proceedings, Berlin u.a.: Springer-Verlag, S. 1-39