

erschienen in: Th. Malsch (Hrsg.) (1998): Sozionik. Soziologische Ansichten über künstliche Sozialität, Berlin: Edition Sigma, S. 128-167

## **Akteure, Aktanten und Agenten**

### **Konstruktive und rekonstruktive Bemühungen um die Handlungsfähigkeit von Technik**

**Ingo Schulz-Schaeffer**

Technische Universität Berlin, Institut für Sozialwissenschaften, Franklinstr. 28/29, D-10587 Berlin, e-mail: schulz-schaeffer@tu-berlin.de

#### **1. Einleitung**

Einer der frühesten Definitionen zufolge beschäftigt sich Verteilte Künstliche Intelligenz (VKI) mit Problemen, für die ein einzelner Problemlösungsmechanismus unangemessen erscheint und für deren Bearbeitung sich statt dessen das Zusammenwirken einer Mehrzahl von Problemlösungskomponenten als nützlich erweist (vgl. Davis 1980: 42). Die Neuartigkeit dieses Forschungsansatzes besteht nicht schon darin, ein Gesamtproblem in Teilprobleme zu zerlegen und in unterschiedlichen Modulen jeweils separat bearbeiten zu lassen. Dies geschieht auch in vielen anderen Bereichen der Technikentwicklung. Der neue Gedanke ist vielmehr der, daß es im Vorhinein keinen globalen Problemlösungsplan gibt, sondern eine Problemlösung entsteht, indem die beteiligten Problemlösungskomponenten ihre Aktivitäten erfolgreich koordinieren. Es ist diese Vorstellung, die die Bezeichnung der Problemlösungskomponenten als Akteure (vgl. Hewitt 1977: 324f), oder wie sich in der VKI-Community inzwischen durchgesetzt hat: als Agenten, motiviert. Mit dieser Benennung wird zum Ausdruck gebracht, daß die beteiligten Komponenten - zumindest mit Blick auf die fragliche Problemlösung - keiner direkten externen Steuerung unterliegen, sondern in irgendeiner Weise "von sich aus" agieren und interagieren.

Eine ganz ähnliche Art und Weise der Betrachtung technischer Artefakte findet sich in Gestalt der von Michel Callon und Bruno Latour entwickelten Actor-Network-Theorie auch innerhalb der sozialwissenschaftlichen Wissenschafts- und Technikforschung. Auch hier werden technische Artefakte als Elemente umfassenderer Zusammenhänge (Netzwerke) betrachtet und die Wirkungsweise des Netzwerks als ein emergentes Resultat der wechselseitigen Relationierungen der Elemente. Auch aus der Perspektive der Actor-Network-Theorie ist der Beitrag der einzelnen Elemente zur Wirkungsweise des Netzwerkes mithin nicht im Vorhinein festgelegt, weshalb alle Elemente des Netzwerkes unter Einschluß der technischen Artefakte mit Blick auf die Strukturierung des Netzwerks als "entities that *do* things" (Latour 1988: 303) betrachtet werden, in der Terminologie der Actor-Network-Theorie: als Aktanten. Anders als bei den Bemühungen um die Konstruktion von Multiagenten-Systemen innerhalb der VKI, in deren Mittelpunkt das koordinierte Problemlösen einer Mehrzahl technischer Agenten steht, analysiert die Actor-Network-Theorie Technikentwicklung als Bildung von Netzwerken, die technische und nichttechnische Komponenten umfassen, oder wie Callon und Latour sagen: menschliche und nichtmenschliche Aktanten. Ein weiterer Unterschied besteht darin, daß für die Actor-Network-Theorie jegliche Technik, also auch einfache technische Gegenstände des alltäglichen Gebrauchs wie Schlüsselanhänger oder Sicherheitsgurte als Aktanten in Frage kommen, während die VKI ihre Agenten in besonderer Weise mit Eigenschaften auszustatten sucht, die es ihnen erlauben sollen, koordiniertes Verhalten zu aufzuweisen.

Wie die konventionelle Künstliche Intelligenz-Forschung (KI) lässt sich auch die VKI aus zwei unterschiedlichen Perspektiven betrachten. So wie die KI den Anspruch erhoben hat, nicht nur technologische Forschung zu betreiben, sondern einen Beitrag zum Verständnis menschlichen Denkens zu leisten, so finden sich innerhalb der VKI einige Forscher, die mit ihrer Arbeit den Anspruch verbinden, theoretische Grundlagen der Interaktion sozialer Akteure zu erforschen (vgl. z.B. Hewitt 1991: 89, 98; Castelfranchi 1992: 206). Deutlicher als in der KI steht in der VKI allerdings das Anwendungsinteresse im Vordergrund. Im folgenden beschränke ich mich im wesentlichen auf diesen Aspekt und betrachte die Forschung der VKI als eine konstruktive Bemühung, die darauf gerichtet ist, technische Artefakte in einer Weise zu Handeln zu befähigen, die es ermöglicht, nutzbare Techniken koordinierten Problemlösens zu konstruieren.

Hält man es von vornherein für aussichtslos, Techniken zu entwickeln, deren Zusammenwirken in irgendeinem substantiellen Sinne als aufeinander bezogenes Handeln beschrieben werden kann - etwa aus einer soziologischen Perspektive, die den Begriff des Handelns in einer Weise mit dem Merkmal der Kommunikation von Sinn verbunden sieht, wie dies nur in sozialen Zusammenhängen menschlicher Akteure vorkommt - so erscheint der Schritt von der KI zur VKI als nichts anderes als die Ersetzung eines überzogenen Anspruchs durch einen weiteren. Nicht zuletzt durch den Einfluß der Actor-Network-Theorie hat sich in dieser Frage innerhalb der sozialwissenschaftlichen Technikforschung ein Wandel vollzogen, der darauf hinausläuft, mit kategorialen Unterscheidungen, die technische Artefakte aus dem Bereich des Sozialen prinzipiell ausschließen, vorsichtiger zu sein. Vor diesem Hintergrund bietet sich die Actor-Network-Theorie als Theorierahmen für eine techniksoziologische Betrachtung der VKI an, die nicht schon durch ihre Annahmen ein negatives Ergebnis der Analyse präjudiziert.

## 2. Agenten: Konstruktive Bemühungen um die Handlungsfähigkeit von Technik

Von Carl Hewitt, einem der Begründer der VKI, stammt die Bemerkung, "that the question *what is an agent?* is embarrassing for the agent-based computing community just in the same way that the question *what is intelligence?* is embarrassing for the mainstream AI community" (zitiert bei Wooldridge/Jennings 1995: 1). Unangenehm deshalb, weil der Begriff des Handelns wie der der Intelligenz im alltäglichen Sprachgebrauch eher auf einem intuitiven Verständnis beruht als auf einem expliziten Katalog von Merkmalseigenschaften. D.h. man kann sich im Zweifelsfall sicherlich bei der Mehrzahl aller Verhaltensweisen darüber streiten, ob sie intelligentes Verhalten bzw. Handeln darstellen, und dennoch wird normalerweise niemand daran zweifeln zu wissen, was Intelligenz und was Handeln ist. Zugleich aber sind es die auf diese Weise nur vage umrissenen Phänomene, um deren technische Nachbildung es der KI bzw. der VKI geht, so daß man auf die entsprechenden Begriffe nicht einfach verzichten kann. Um zu vermeiden, daß der Begriff des Agenten zu einem "noise' term, subject to both abuse and misuse, to the potential confusion of the research community" (ebd.: 2) wird, bedarf es mithin gesonderter begrifflicher Anstrengungen. Soweit ich sehe, zeichnen sich hierbei in der VKI drei unterschiedliche Konzeptionen ab: ein abstrakter, ein intentionaler und ein interaktionistischer Agentenbegriff.<sup>1</sup>

---

1. Diese Aufteilung hat eine gewisse Ähnlichkeit mit der Unterscheidung von Moulin und Cloutier (1994: 265) zwischen reaktiven, intentionalen und sozialen Agenten (vgl. auch Moulin/Chaib-draa 1996: 8f). Insbesondere die Unterscheidung zwischen reaktiven und intentionalen Agenten einerseits und sozialen Agenten ande-

Als Bemühungen um einen abstrakten Agentenbegriff kann man diejenigen Überlegungen zusammenfassen, die von der Frage ausgehen, welche Bedingungen mindestens erfüllt sein müssen, damit Problemlösungen als emergentes Ergebnis des Zusammenwirkens mehrerer Problemlösungskomponenten entstehen und nicht auf der Basis globalen Wissens oder globaler Steuerung (vgl. Gasser et al. 1989: 56). Die beiden grundlegendsten Bedingungen ergeben sich unmittelbar aus der Forderung, daß die Struktur des Zusammenwirkens der Komponenten nicht von außen determiniert ist, sondern daß es - wie Hewitt bereits 1977 formuliert hat - um Zusammenhänge geht, "in which the control structure emerges *as a pattern of passing messages among the objects*" (Hewitt 1977: 325). Dies ist zum einen die bereits eingangs genannte Bedingung, daß die (oder wenigstens einige der) Komponenten über ein bestimmtes Maß an Autonomie verfügen, also in einer Weise agieren, die von anderen Komponenten nicht direkt gesteuert werden kann. Zum anderen ist dies die Bedingung, daß die Komponenten - wie auch immer - in der Lage sind, miteinander zu interagieren, also sich wechselseitig deshalb in einer bestimmten Weise zu verhalten, weil sich andere Komponenten in einer bestimmten Weise verhalten haben und dieses Verhalten als Information, d.h. als "difference that makes a difference" (Bateson 1972: 489) genutzt werden kann. Ein Agent eines Multiagenten-Systems muß mithin über mindestens zwei Eigenschaften verfügen: Er muß erstens ein bestimmtes Repertoire an Verhaltensweisen besitzen, die nur von ihm selbst ausgelöst werden können. Und er muß Veränderungen in seiner Umwelt, insbesondere im Verhalten anderer Agenten wahrnehmen und als Information verwenden können, um aus seinem Verhaltensrepertoire bestimmte Verhaltensweisen auszuwählen. Als dritte Bedingung kommt hinzu, daß kein Agent über die Ressourcen verfügt, das fragliche Problem aus eigener Kraft zu lösen, also nicht alle dazu erforderlichen Verhaltensweisen besitzt oder nicht alle erforderlichen Informationen oder beides nicht.

Zusammengenommen ergeben diese Minimalbedingungen einen "schwachen" Begriff von Handlungsfähigkeit (vgl. Wooldridge/Jennings 1995: 2). Agenten, die die genannten Bedingungen erfüllen, müssen in ihrer internen Struktur nicht besonders anspruchsvoll sein: Die Anzahl der Verhaltensalternativen kann sehr gering sein. Die Auswahl einer der Verhaltensalternativen kann im einfachsten Fall einem direkten Reiz-Reaktions-Mechanismus folgen, wobei bestimmte Verhaltensweisen anderer Agenten oder sonstige Ereignisse in der Umwelt des Agenten als Auslösereiz dienen.<sup>2</sup> Solche einfachen Agenten werden als "reaktive Agenten" bezeichnet: "The behavior of a reactive agent can be first defined using ... a strict 'S-R' (Stimulus-Reaction) scheme. This scheme excludes *a priori* any 'reasoning' between S and R, where S is considered as a particular state of the environment containing the entity, and R as a sequence of basic actions." (Drogoul/Ferber 1994: 6) Es liegt hier mit anderen Worten ein Begriff von Handlungsfähigkeit zu Grunde, der vollständig auf anthropomorphe Vorstellungen über Denk- und Entscheidungsprozesse verzichten kann. Multiagenten-Systeme auf der Basis solcher einfacher Agenten können dementsprechend auch nach dem Modell von Tiergesellschaften oder auch nach dem Modell selbstorganisierender Zusammenhänge zwischen unbelebten Elementen konstruiert werden. Die Suche nach den allgemeinsten Bedingungen

---

rerseits halte ich jedoch für irreführend, da es auch im Bereich reaktiver und intentionaler Agenten um die Konstruktion sozialer Abstimmungsprozesse geht.

2. Voraussetzung ist allerdings, daß die gesamte Reiz-Reaktions-Kette, aus der die Problemlösung schließlich resultiert, nicht als globaler Problemlösungsplan vorprogrammiert ist. Andernfalls wäre jedes objektorientiert erstellte Programm ein Multiagenten-System. Allerdings hätte es dann auch keinen Sinn mehr, von einer wie immer gearteten Autonomie der Agenten zu sprechen.

unter denen man die Elemente von Multiagenten-Systemen als Agenten bezeichnen kann, führt mithin zu einer Entgrenzung des Begriffs des Handelns und der Interaktion, die weit über das hinausgeht, was man im alltagssprachlichen oder im sozialwissenschaftlichen Diskurs unter diesen Begriffen versteht.

Die meisten VKI-Forscher können und wollen ihre Herkunft aus Forschungszusammenhängen der KI allerdings nicht leugnen. Auch für sie ist die Problemlösungsintelligenz von Multiagenten-Systemen eine emergente Eigenschaft des Zusammenwirkens der Agenten, die aber ähnlich wie die Programme der konventionellen KI als Einheiten betrachtet werden, die bereits für sich über ein gewisses Maß kognitiver Fertigkeiten verfügen. Der Ausarbeitung des Agentenbegriffs liegt hier dementsprechend die Frage zu Grunde, welche Eigenschaften es sind, die die Handlungsfähigkeit insbesondere menschlicher Akteure ausmachen. Diese Frage ist der Ausgangspunkt sowohl des intentionalen wie auch des interaktionistischen Agentenbegriffs, wobei der Unterschied darin besteht, daß ersterer die fraglichen Eigenschaften primär als Eigenschaften der handelnden Individuen thematisiert, letzterer dagegen primär als Eigenschaften der Interaktionsbeziehungen zwischen ihnen.

Normalerweise müssen zwei Bedingungen erfüllt sein, damit wir jemanden als für seine Handlungen verantwortlich und damit als einen handlungsfähigen Interaktionspartner betrachten: Wir setzen erstens voraus, daß die Person auch anders hätte handeln können, und zweitens unterstellen wir, daß die Person Gründe dafür hat so und nicht anders zu handeln. Letzteres findet seinen Ausdruck darin, daß wir in aller Regel bereit sind, ein Fehlverhalten, das einer Person aus Fahrlässigkeit oder in einem Zustand verminderter Zurechnungsfähigkeit unterläuft, eher zu entschuldigen als eine vorsätzliche Tat. Ersteres zeigt sich darin, daß wir bereit sind, im Fall einer Handlung bei der die Person nachweisen kann, daß sie keine andere Wahl hatte als so zu handeln (z.B. Notwehr) gänzlich auf negative Sanktionen verzichten. Uneingeschränkt gilt uns nur das von einer Person gewollte und gewählte Tun als sein Handeln. Reaktive Agenten sind in keiner dieser beiden Dimensionen handlungsfähig. Deren Einbeziehung bei der Modellierung künstlicher Agenten markiert den Übergang zum Konzept des intentionalen Agenten.

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. Auf dieser Beobachtung fußt das Konzept intentionaler Agenten, derzeit die sicherlich in der VKI verbreitetste und am weitesten ausgearbeitete Antwort auf die Frage, was ein Agent ist: "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)

Geht man davon aus, daß es 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

Gründe legen, so ist es naheliegend, Multiagenten-Systeme in ähnlicher Weise auf der Basis intentionaler Agenten zu konzipieren. Entsprechende Definitionen, 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 der Annahme, daß eine geeignete Ausstattung der einzelnen Agenten mit Absichten, Überzeugungen usw. die zentrale Voraussetzung für die Entstehung koordinierten Verhaltens in Multiagenten-Systemen ist, sehen Forscher, die dem alltagspsychologischen Ansatz verpflichtet sind, eine der vordringlichsten Aufgaben der VKI darin, formale Spezifikationen intentionaler Begriffe zu entwickeln, die softwaretechnisch umsetzbar sind. Die bekanntesten Bemühungen um eine umfassende und kohärente formale Semantik intentionaler Begriffe stammen von Cohen und Levesque (1990), Rao und Georgeff (1991), Werner (1989) und Shoham (1993). Solche Konzepte werden zusammenfassend häufig auch als BDI-Architekturen (Belief, Desire, Intention) bezeichnet (vgl. Rao/Georgeff 1991). Sie unterscheiden sich weniger in ihrer grundsätzlichen Herangehensweise als vielmehr in der Frage der Auswahl, Definition und konzeptuellen Relationierung der einzelnen intentionalen Begriffe. Für alle solche Versuche gilt, so Shoham: "the intuition about these formal notions do indeed derive from the everyday, common sense meaning of the words" (Shoham 1994: 274). Mehr noch: "the everyday intuition has proven an good guide to employing the formal notions in some circumscribed applications." (ebd.)

Der interaktionistische Agentenbegriff setzt diesem intuitiven ein kontraintuitives Verständnis von Handlungsfähigkeit entgegen. Zwar kommt es den Denkgewohnheiten der konventionellen KI entgegen, "the localized, internalized, and individualist conceptions of reasoning and action that we have inherited from a cognitively-oriented world-view" (Gasser 1992: 201) als elementare Analyseeinheit der Bestimmung von Handlungsfähigkeit zu nehmen. Genau dies aber, so das Gegenargument, könnte sich bei einem Forschungsprogramm, das auf die Dynamik emergierender Problemlösungsmuster setzt, als eine zu restriktive Annahme erweisen, insbesondere dann, wenn das Ziel darin besteht, "really flexible, adaptive and reconfigurable, distributed, 'open' systems - and in particular, systems, composed of parts that are developed under different perspectives and authority" (ebd.) zu konstruieren. Gasser als einer der entschiedensten Gegner des alltagspsychologischen Ansatzes schlägt deshalb vor, die Eigenschaften von Agenten nicht als Voraussetzung, sondern als Resultat der Interaktionsbeziehungen zu betrachten, in die sie einbezogen sind: "we can begin to develop computational structures and architectures that don't take agent architectures and agenthood as fixed starting points from which to build communities. Instead, we shall examine the possibilities for building systems as structures of interaction, from which useful entities we can call 'agents' emerge" (ebd.). Die zentrale Annahme des interaktionistischen Agentenbegriffs ist mithin "that in effect, agents can be construed as emerging from interactions, rather than vice versa" (Gasser/Briot 1992: 92).

"To make substantial theoretical progress", so eine zentrale Forderung Gassers, "we must begin to lay firm social foundations for DAI (Distributed Artificial Intelligence, Anm. d. Verf.) research." (Gasser 1991: 111) Bereits bei einer ersten Annäherung an die Frage, was ein Agent ist, wird allerdings deutlich, daß innerhalb der VKI sehr unterschiedliche Vorstellungen darüber herrschen, worin diese sozialen Grundlagen bestehen bzw. in welche Richtung nach ihnen zu suchen ist: Aus der Perspektive des abstrakten Agentenbegriffs ist Sozia-

lität emergente Funktionalität: Keiner der einzelnen Agenten plant oder beabsichtigt die Effekte ihres Zusammenwirkens, und dennoch können Effekte entstehen, die sich insofern als funktional erweisen als sie die Reproduktion des Zusammenhangs begünstigen (vgl. Castelfranchi/Conte 1992: 80f). Der alltagspsychologische Ansatz betrachtet Sozialität als Resultat intentionalen Handelns unter der Bedingung, daß die Realisierung individueller Absichten von bestimmten Verhaltensweisen anderer, ebenfalls intentional handelnder Individuen abhängig ist. Der interaktionistische Ansatz dagegen thematisiert Sozialität als Muster sozialer Beziehungen, die einerseits das Verhalten der Agenten strukturieren, andererseits aus den Interaktionsbeziehungen zwischen den Agenten resultieren. Erwartungen, die die einzelnen Agenten gegenüber ihrem eigenen und dem Verhalten anderer Agenten haben, sind aus dieser Sicht ebenso das Ergebnis der sozialen Beziehungen, innerhalb derer sie sich bewegen, wie umgekehrt die Muster der Interaktion "an emergent property of cycles of interpretation, symbolization, and response among many agents in a system" (Gasser/Briot 1992: 99) sind.

Angesichts dieser heterogenen Theorielage könnte sich die Soziologie herausgefordert sehen, auf der Basis ihres eigenen Fachwissens eine Antwort darauf zu geben, welches der angemessenste Agentenbegriff ist. Eine solche Beurteilung würde sicherlich bei einer Mehrheit von Soziologen darauf hinauslaufen, dem interaktionistischen als soziologisch gehaltvollstem Konzept den Vorzug zu geben und, wenn überhaupt, dieses Konzept als Ausgangspunkt für eine soziologische Fundierung der VKI zu wählen. Allerdings lassen sich gegen eine solche Herangehensweise, die Frage nach den sozialen Grundlagen der VKI als soziologische Frage aufzugreifen und auf der Grundlage soziologischer Theorie beantworten zu wollen, drei Einwände erheben: Erstens ist auch die Theorielage innerhalb der Soziologie heterogen. Schon deshalb wird es die eine soziologische Grundlegung der VKI nicht geben können, sondern bestenfalls eine Mehrzahl von Angeboten, die jeweils nach Maßgabe ihrer Basisannahmen Angemessenheit beanspruchen können. Schwerer wiegen allerdings zwei weitere Einwände: Der zweite Einwand besteht darin, daß es keineswegs gesagt ist, daß eine soziologisch angemessene Konzeption sozialer Zusammenhänge auch mit Blick auf das Ziel der VKI, Systeme verteilten Problemlösens zu konstruieren, nützlich ist. Und schließlich kann man, so der dritte Einwand, fragen, ob in diesem Sinne nützliche Konzepte überhaupt zwangsläufig soziologische Konzepte sein müssen.

Man kann diese letzten beiden Einwände am Beispiel des interaktionistischen und des intentionalen Agentenbegriffs gut illustrieren. Sucht man innerhalb der VKI nach einer soziologisch fundierten Konzeption, so findet man entsprechende Überlegungen am deutlichsten im Zusammenhang des interaktionistischen Agentenbegriffs. Mit der Vorstellung, die Eigenschaften der Agenten als emergentes Resultat von Interaktionsprozessen zu konzipieren, greift Gasser ausdrücklich auf Annahmen des symbolischen Interaktionismus zurück (vgl. Strübing, in diesem Band). Zugleich handelt er sich damit ein ernsthaftes Umsetzungsproblem ein. Einerseits bewertet er die Annahmen des symbolischen Interaktionismus als "conceptually fruitful ... for understanding knowledge and action in social terms" (Gasser 1991: 113), andererseits muß er zugestehen: "A key missing link ... is how to make them computational." (ebd.: 132) Zwar ist der interaktionistische Agentenbegriff ohne Zweifel der soziologisch anspruchsvollste der drei Agentenbegriffe. Hinsichtlich seiner softwaretechnischen Umsetzung repräsentiert er jedoch die am wenigsten entwickelte Forschungsperspektive innerhalb der VKI. Sehr viel weiter fortgeschritten in Richtung auf praktische Nutzanwendungen sind dagegen eine Reihe solcher Ansätze, die eher von alltagspsychologischen Annahmen ausgehen. Auch sie verwenden in gewissem Sinne eine Sozialtheorie, nur eben keine soziolo-

gische, sondern eine solche des Alltagswissens. Führt man sich vor Augen, daß es vielfach solches common sense-Wissen ist, das es einem ermöglicht, sich im alltäglichen Leben erfolgreich zurechtzufinden, spricht jedoch grundsätzlich nichts dagegen, warum nicht auch die Agenten von VKI-Systemen ihre Aktionen auf der Basis alltagspsychologischer Annahmen über mentale Zustände erfolgreich zu koordinieren in der Lage sein sollen. Aus dieser Perspektive ist ein sozialwissenschaftliches Verständnis sozialer Zusammenhänge mithin keine notwendige Bedingung dafür, Multiagenten-Systeme nach dem Vorbild sozialer Interaktionszusammenhänge konstruieren zu können.

Stellt man die Frage nach den sozialen Grundlagen der VKI von vornherein als Frage ihrer soziologischen Fundierung, so läuft man meines Erachtens in doppelter Hinsicht Gefahr, sich in unentscheidbaren Kontroversen zu verfangen, weil weder die soziologische Angemessenheit einer Konzeption eine Gewähr für ihre praktische Nützlichkeit innerhalb der VKI ist, noch umgekehrt die praktische Nützlichkeit bestimmter Vorstellungen über soziale Zusammenhänge für die Konstruktion von Multiagenten-Systemen ohne weiteres als Argument ihrer soziologischen Angemessenheit gelten kann. Vielversprechender scheint mir deshalb eine Perspektive zu sein, die diese Frage zunächst offen hält. Eine solche Perspektive eröffnet die Actor-Network-Theorie.

Die Actor-Network-Theorie bietet aus drei Gründen einen geeigneten Theorierahmen für eine techniksoziologische Betrachtung von Multiagenten-Systemen: (1) Die Actor-Network-Theorie teilt mit den agentenorientierten Ansätzen der VKI die Annahme, daß der Begriff der Handlungsfähigkeit keine kategoriale Unterscheidung zwischen technischen Artefakten und sozialen Akteuren konstituiert, sondern alle Elemente umfassenderer Zusammenhänge als Akteure betrachtet werden können, sofern ihr Verhalten in einer Weise zur Wirkungsweise des Zusammenhangs beiträgt, die nicht der direkten Kontrolle eines anderen Elements unterliegt. Anders also als etwa Collins (1990) stellt die Actor-Network-Theorie nicht die Grenzen, sondern die Möglichkeit der Handlungsfähigkeit technischer Artefakte in den Mittelpunkt ihrer Überlegungen. (2) Ausgehend von der Actor-Network-Theorie läßt sich zeigen, daß die drei genannten Agentenbegriffe sich nicht notwendig wechselseitig ausschließen, sondern als einander ergänzende Perspektiven auf Prozesse des Netzwerkbildens betrachtet werden können. (3) Schließlich kann sie genutzt werden um zu zeigen, daß Multiagenten-Systeme nur einen Teil des Netzwerkes ausmachen, das es zu etablieren gilt, will man Multiagenten-Systeme konstruieren, die in irgendeinem Sinne ein nützliches Problemlösungsverhalten aufweisen. Bevor ich diese Punkte näher beleuchte, gebe ich zunächst einen kurzen Überblick über die Actor-Network-Theorie.

### **3. Aktanten: Rekonstruktive Bemühungen um die Handlungsfähigkeit von Technik**

Das von Callon und Latour entwickelte Konzept der Akteur-Netzwerke steht in der Theorietradition der sozialkonstruktivistischen Wissenschafts- und Techniksoziologie und wendet sich zugleich gegen sie. Theoretischer und methodologischer Ausgangspunkt ist eine über die ursprüngliche Version von Bloor weit hinausgehende Erweiterung des Symmetriprinzips. Im Interesse einer Wissenschaftssoziologie, die nicht mehr nur den sozialen Kontext, in dem sich Wissenschaftler bewegen, sondern die Inhalte wissenschaftlicher Forschung selbst zu erklären beansprucht, hatte Bloor gefordert, der Wissenschaftsforscher solle unparteiisch in Bezug auf die von den Wissenschaftlern eingenommenen Wahrheitsansprüche sein und die im beobachteten Feld für wahr oder für falsch gehaltenen Überzeugungen symmetrisch, also unter

Rückgriff die selben Kategorien von Ursachen erklären (vgl. Bloor 1976: 5). Durch das Symmetrieprinzip wird zum Explanandum erhoben, was auch die Mehrzahl der Wissenschaftssoziologen bis dahin als Explanans gelten ließ: "the acceptance of a claim that is now considered to be true should not be explained by its true content (for example, in terms of a better correspondence with nature), whereas the acceptance of another claim that is currently considered false is explained by referring to, for example, social circumstances of its conception. 'Nature' was not to enter the explanatory endeavor as *explanans*; rather, it should be the *explanandum*." (Bijker 1993: 119).

Die Kategorie von Ursachen, die Bloor im Blick hatte und die im Anschluß an Bloor in den frühen Laborstudien und den Arbeiten der Edinburgh-School verwendet wurden, sind sozialwissenschaftliche Kategorien, Kategorien, die Wissenschafts- und Technikentwicklung als Prozeß der sozialen Aushandlung und Durchsetzung von Bedeutungen erklären. An diesem Punkt setzt die weiterreichende Forderung Callons und Latours an: "the same arguments that have been made about Nature have to be made *symmetrically* about Society. How could we make so many precautions in *not* believing directly what scientists and engineers say about objectivity and subjectivity, and readily believe what other scientists (social this time) say about society, culture and economy? At this point we are in a great need of a rule of symmetry that does not grant Society privileges refused to Nature." (Latour 1987: 144) Der Beobachter soll sich deshalb nicht nur mit Blick auf Aussagen über natürliche oder technische Zusammenhänge, sondern gleichermaßen mit Blick auf Aussagen über soziale Zusammenhänge unparteiisch und symmetrisch verhalten. Auch soziale Faktoren sollen nicht länger als Explanans gelten, sondern als Explanandum betrachtet werden müssen.

Ausgehend von diesem generalisierten Symmetrieprinzip suchen Callon und Latour ein begriffliches Instrumentarium, das nicht nur auf alle Vorannahmen über natürliche oder technische Verursachung verzichtet, sondern zusätzlich auf jegliche Vorannahmen über soziale Verursachung. Gesucht ist mit anderen Worten eine einheitliche Beschreibungssprache, die alle diese Erklärungen als Resultat und nicht als Voraussetzung der beobachteten Prozesse behandelt: "The goal is not only to explain conflicting viewpoints and arguments in a scientific or technological controversy in the same terms. We know that the ingredients of controversies are a mixture of considerations concerning both Society and Nature. For this reason we require the observer to use a single repertoire when they are described. ... the rule which we must respect is not to change registers when we move from the technical to the social aspects of the problem studied." (Callon 1986: 200) Diese angestrebte Beschreibungssprache, die - ohne eine der genannten Kategorien von Ursachen zu privilegieren - auf alle beobachteten Prozesse gleichermaßen angewendet werden können soll, bezeichnet Callon als "Übersetzungsvokabular", als "vocabulary of translation" (ebd.).

Das generalisierte Symmetrieprinzip hat zum Ziel, Prozesse von Wissenschafts- und Technikentwicklung aus einem neuen Blickwinkel zu betrachten. Weder werden soziale Faktoren, Normen oder bestimmte institutionelle oder organisationale Konfigurationen als stabiler Kontext angesehen, mittels dessen wissenschaftliche und technische Entwicklungen erklärt werden, noch werden soziale Verhältnisse ihrerseits einlinig als Resultat von Wissenschafts- und Technikentwicklung angesehen (vgl. Callon 1986: 221; Latour 1991: 116). Technik, Natur und Gesellschaft werden einander nicht jeweils wechselseitig als Explanans und

Explanandum zugeordnet,<sup>3</sup> sondern als koevolutionäres Resultat. Die Prozesse sichtbar zu machen, die dazu führen, daß es ab einem bestimmten Punkt der Entwicklung von Wissenschaft und Technik im Verständnis von Beteiligten und Betroffenen offenkundig unproblematisch ist, von natürlichen, technischen oder sozialen Faktoren zu sprechen, ist die Aufgabe des Übersetzungsvokabulars. Es erklärt solche Prozesse als (erfolgreiche oder gescheiterte) Prozesse der Bildung stabiler Netzwerke durch wechselseitige Übersetzungen (vgl. Callon/Latour 1992: 348; Callon 1987: 97; Latour 1991: 117).

Mit dem Begriff der Übersetzung - und den synonym verwendeten Begriffen "transfer", "displacement" oder auch "metaphor" (vgl. Latour 1983: 153) - bezeichnet der Actor-Network-Ansatz auf einer sehr allgemeinen Ebene alle (Um)Definitionen der Identität, der Eigenschaften und der Verhaltensweisen irgendwelcher Entitäten, die darauf gerichtet sind, bestimmte stabile Assoziationen von Entitäten (Netzwerke) hervorzubringen (vgl. Callon 1986: 203; Callon 1991: 143). Jede Entität, die die Fähigkeit besitzt, solche Übersetzungen vorzunehmen, wird als Akteur, oder mit dem entsprechenden Begriff des Übersetzungsvokabulars: als Aktant bezeichnet: "an 'actor' ... is any entity that more or less successfully defines and builds a world filled by other entities with histories, identities, and interrelations of their own" (Callon 1991: 140; vgl. Latour 1991: 121). Auf der anderen Seite sind die Aktanten aber auch das Resultat von Übersetzungen, der Übersetzungen anderer Aktanten: "Thanks to translation, we do not have to begin our analysis by using actants with fixed borders and assigned interests. Instead, we can follow the way in which actant B attributes a fixed border to actant A, the way in which B assigns interests or goals to A, the definition of those borders and goals shared by A and B, and finally the distribution of responsibility between A and B for their joint action." (Latour 1991: 124).

In einem ständigen Perspektivenwechsel muß der Beobachter beides tun: "Defining the essence of innovations by the existence of their successive and simultaneous actants, and then turning around to define the actants by the successive innovations in which they appear" (ebd.: 122). Beobachtet man Prozesse der Wissenschafts- und Technikentwicklung auf diese Weise, dann sind, so meint Latour, alle weiteren Erklärungsgründe verzichtbar: "If we display a socio-technical network - defining trajectories by actants' association and substitution, defining actants by all the trajectories in which they enter, by following translations and, finally, by varying the observer's point of view - we have no need to look for any additional causes." (Latour 1991: 129). Statt dessen ergibt sich die Erklärung ausschließlich aus der Beobachtung selbst: "In a universe of innovations solely defined by the associations and substitutions of actants, and of actants solely defined by the multiplicity of inventions in which they conspire, the translation operation becomes the essential principle of composition, of linkage, of recruitment, or of enrolment. But since there no longer exists any external point of view to which we could ascribe the degree of reality or of success of an innovation, we can only obtain an evaluation by triangulating the many points of view of the actors." (Latour 1991: 124)

Warum gelten bestimmte wissenschaftliche Theorien, warum funktionieren bestimmte Techniken und warum sind andere Bestrebungen in Wissenschaft und Technik gescheitert? Die Antwort der Actor-Network-Theorie lautet: Wir können nachzeichnen, auf welche Weise es

---

3. Die Vorgehensweise, soziale Faktoren als Explanans zu wählen, lehnen Callon und Latour als "Soziologismus" bzw. "sozialen Realismus" ab. Natur als Explanans, also die Begründung wissenschaftlicher Wahrheit durch Übereinstimmung mit der Wirklichkeit wird als "natürlicher Realismus" verworfen, Technik als Explanans als "Technologismus" (vgl. Latour 1988: 307; Callon/Latour 1992: 348).

in dem einen Fall Aktanten gelungen ist, durch geeignete Übersetzungen ein Netzwerk von Aktanten zusammenzubringen und aufrecht zu erhalten, das diesen Erfolg zeitigt, und inwiefern im anderen Fall der Widerstand von Aktanten, sich in der erforderlichen Weise redefinieren zu lassen, den Aufbau eines entsprechenden Netzwerks verhindert hat. Der Erfolg hat aus der Perspektive dieser Betrachtungsweise zwei Väter, Konvergenz und Irreversibilität: Damit stabile Netzwerke entstehen können, müssen die Übersetzungen zum einen die Konvergenz eines Netzwerkes, zum anderen dessen zumindest zeitweilige Irreversibilität bewirken. Konvergenz bedeutet, daß sich die Aktanten wechselseitig so verhalten, wie sie es voneinander erwarten, also so, wie sie sich wechselseitig "übersetzt" haben. Irreversibilität bedeutet, daß die Aktanten in ihrem Verhalten und ihren wechselseitigen Beziehungen stabil sind, mithin eine gewisse Resistenz gegenüber weiteren Übersetzungen aufweisen. (vgl. Callon 1991: 144ff; Latour 1987: 108ff)

Konvergenz wird erzeugt, wenn es im Übersetzungsprozeß gelingt, die beteiligten Aktanten in Übereinstimmung zu bringen (alignment) und regelhaft zu koordinieren (vgl. Callon 1991: 144). Übereinstimmung tritt als Resultat von Übersetzungsprozessen ein, wenn die wechselseitigen (Re-)Definitionen akzeptiert und die damit verbundenen Zuschreibungen von den betroffenen Aktanten übernommen werden: "A successful process of translation thus *generates* a shared space, equivalence and commensurability. It *aligns*." (ebd.: 145) Da Übersetzungen Einwirkungen mit Rückwirkungen sind (in der Terminologie der Actor-Network-Thorie: Transkriptionen und Präskriptionen) ist der Prozeß der Erzeugung von Übereinstimmung mehrstufig und multilateral. Es handelt sich der Konzeption von Callon und Latour zufolge um eine Abfolge multilateraler Aushandlungen, während derer Interessen und Ziele formuliert und verändert, Handlungsprogramme aufgestellt und modifiziert, konkurrierende Handlungsprogramme einbezogen oder ausgeschaltet, Koalitionen gebildet oder aufgelöst, Aktanten neu eingeführt, umdefiniert oder entfernt werden etc. (vgl. Callon 1986: 211ff; Latour 1987: 108ff; Latour 1992: 247).

Je gesicherter der Status der Aktanten ist und je unproblematischer deren Zusammenwirken, desto höher ist der Grad der Konvergenz des Netzwerkes: "each actor in a convergent network is able to identify and mobilize the skills within that network without having to get involved into costly adaption, translation or decoding" (Callon 1991: 148). Da die Erzeugung von Konvergenz jedoch ein aufwendiger Prozeß ist, hat man es bei der Betrachtung von Wissenschafts- und Technikentwicklung mit jeweils unterschiedlichen Formen von Konvergenz zu tun: "Strongly convergent networks only develop after long periods of investment, intense effort, and coordination. There are many others which are only weakly convergent - networks in which actors find both that their status is constantly in question, and that it is difficult (albeit not impossible) to mobilize other parts of the network." (ebd.)

Irreversibilität bezeichnet die Widerständigkeit der im Netzwerk etablierten Übersetzungen gegenüber Übersetzungen, die darauf gerichtet sind, den Zusammenhalt des Netzwerkes zu schwächen oder aufzulösen: "We might say that a translation is irreversible if it is likely to lead to a search for substitutes, or for translations that are intended to prolong its life or extend its scope." (ebd.: 150f). Irreversibilität ist das Resultat der wechselseitigen Verkopplungen zwischen den Elementen des Netzwerks: "it could be said that irreversibility increases to the extent that each element ... is inscribed in a *bundle* of interrelationships. In such tightly coupled networks, any attempt to modify one element by redefining it leads to a general process of retranslation." (Callon 1991: 150) Irreversibilität liegt also dann vor, wenn kein Element für sich genommen grundlegend verändert werden kann, ohne eine große Anzahl ande-

rer Elemente ebenfalls verändern zu müssen: "The stricter the compatibility rules (plugs) the more alternative translations are disqualified and the more predictable choices become." (ebd.: 151)

Die Erzeugung von Konvergenz und Irreversibilität in Netzwerken beschreiben Callon und Latour als Prozeß des "black-boxing" (Latour 1991: 123): "When a network is strongly convergent and irreversibilised, it can be assimilated to a black box whose behavior is known and predicted independently of its context." (Callon 1991: 152) Oder genauer: "It can be compared with a black box that contains a network of black boxes." (ders. 1987: 95) Die stabilisierten Resultate der Übersetzungsprozesse schütteln ihre Entstehungsgeschichte für gewöhnlich ab, sie gelten dann als selbstevident (vgl. Callon 1991: 145) und werden von den beteiligten Akten selbstverständlich und fraglos vorausgesetzt: "most of the actions become silent, familiar, incorporated (in human or in nonhuman bodies) - making the analyst's job so much harder" (Latour 1988: 308). Wahre Theorien und funktionierende Techniken, Entdecker und Erfinder, Anwender und Nutzer, das alles sind solche *black boxes*, denen man nicht mehr einfach ansieht, daß sie sind, was sie sind, erst als Ergebnis der Übersetzungsprozesse innerhalb eines Netzwerks wechselseitiger Festlegungen, und daß sie es nur solange sind, wie das Netzwerk selbst (als *black box*) aufrecht erhalten werden kann. Die Beschreibungssprache des Actor-Network-Ansatzes ist dazu gedacht, diese *black boxes* zu öffnen. Und der in ihr Vokabular eingebaute Anfangsverdacht ist, daß alle der im Übersetzungsprozeß involvierten Entitäten, gleichgültig, ob sie als *black boxes* schließlich zu sozialen Akteuren, technischen Geräten oder Gegebenheiten der natürlichen Umwelt werden, potentiell betrachtet werden müssen als "entities that *do things*" (ebd.: 303), also als Einheiten, die nicht nur passiv, sondern auch aktiv an den fraglichen Übersetzungsprozessen beteiligt sind. Damit wird die traditionelle Unterscheidung zwischen Akteur und Artefakt mit der in ihr implizierten Zuteilung von "Machen" und "Gemacht sein", fraglich. An ihre Stelle tritt der Begriff des Akten.

#### **4. Genetisch situiertes und effektiv situiertes Handeln**

Im Gegensatz zu den Bemühungen der VKI um die Konstruktion technischer Artefakte, die in besonderer Weise zur Generierung koordinierten Verhaltens befähigt sind, kommen aus der Perspektive der Actor-Network-Theorie bereits einfache technische Gegenstände des täglichen Gebrauchs als Netzwerk-Akten in Frage. Türschließer, Schlüsselanhänger und Sitzgurte sind die Beispiele, anhand derer Latour seine Überlegungen illustriert (vgl. Latour 1988; 1991; 1992), also Artefakte, deren Verhalten bei weitem nicht diejenigen Fertigkeiten erkennen läßt, welche die VKI im Blick hat, wenn es um die Modellierung "intelligenter" Agenten nach dem Vorbild menschlicher Akteure oder subhumaner Lebewesen geht. Dennoch sind diese Beispiele sowohl für das Verständnis der Actor-Network-Theorie wie auch für die Frage nach der Handlungsfähigkeit von VKI-Agenten instruktiv: Für das Verständnis der Actor-Network-Theorie, weil sich ihre Annahmen in solchen Beispielen am deutlichsten den Annahmen eines konventionellen Technikverständnisses gegenüberstellen lassen; für die Frage nach der Handlungsfähigkeit von VKI-Agenten, weil sie die Frage aufwerfen, welche zusätzlichen Merkmale erfüllt sein müssen, damit aus Akten Agenten werden.

Normalerweise erwartet man von technischen Geräten, daß ihre Abläufe strikt regelgeleitet ablaufen, sie also in vorhersehbarer Weise bei bestimmten Anfangsbedingungen stets ein bestimmtes Verhalten aufweisen (Determiniertheit); daß die zu Grunde liegenden Ablaufregeln von außen festgelegt sind, also von dem technischen Mechanismus nicht selbsttätig verändert werden können (Heteronomie); daß der Sinn der Abläufe nicht in den Abläufen selbst liegt,

sondern in den Handlungszwecken, zu deren Erreichung sie konstruiert und benutzt werden (Sinnverschiebung); daß man als Benutzer einer funktionierenden Technik nicht wissen muß, auf welche Weise die Technik funktioniert, sondern nur, wie man sie benutzen muß, um das gewünschte Ergebnis zu erzielen (Sinnentlastung); und schließlich daß der technische Zusammenhang, unbeeinflußt von jener Vielzahl kontextspezifischer Faktoren abläuft, die jede Handlungssituation zu einem einmaligen Ereignis werden läßt (Robustheit). Ein funktionierender Türschließmechanismus erfüllt ganz offensichtlich alle diese Erwartung, so daß man aus der Perspektive eines konventionellen Technikverständnisses konstatieren wird, daß er funktioniert, weil er zweckmäßig konstruiert wurde, und es mithin keinen Grund gibt, ihm eine wie auch immer geartete Handlungsfähigkeit zuzuschreiben.

Aus der Perspektive der Actor-Network-Theorie ist ein solches konventionelles Technikverständnis das Resultat einer ex post-Betrachtung, die verkennt, daß der Türschließer die genannten Eigenschaften nicht als isolierte Einheit besitzt, sondern als Element eines Netzwerkes sich wechselseitig bedingender Verhaltensweisen einer Vielzahl menschlicher und nichtmenschlicher Aktanten. Aus der Perspektive der *ready made technology* erscheinen technische Artefakte als *black boxes*, d.h. als Entitäten mit festgelegtem und vorhersehbarem Verhalten. Diese Perspektive, so das Argument der Actor-Network-Theorie, blendet zweierlei aus: Den ständigen Prozeß der Aufrechterhaltung jener Relationen, deren Resultat das Verhalten der einzelnen Aktanten ist, und den Prozeß wechselseitigen Relationierungs, der überhaupt erst zur Entstehung des Netzwerkes geführt hat. Man muß deshalb die entgegengesetzte Perspektive der *technology in the making* einnehmen, um die *black box* funktionierender Technik öffnen und die Akteursqualitäten der nichtmenschlichen wie der menschlichen Aktanten erkennen zu können.<sup>4</sup>

Im Beispiel des automatischen Türschließers tut Latour dies in Form gedankenexperimenteller Überlegungen: Türen sind dazu da, um Räume oder Gebäude geschlossen halten und sie dennoch betreten oder verlassen zu können. Zu diesem Zweck bedarf es mehr als nur der Tür selbst, nämlich unter anderem disziplinierter Benutzer, die diese nicht einfach offen stehen lassen. Dies nun kann zum Problem werden, etwa im Fall der Eingangstür der Universitätsfakultät, die Latour in seinem Beispiel vor Augen hat. Eine Lösung wäre, einen Portier einzustellen, dessen Aufgabe es ist, die Tür zu öffnen und zu schließen, aber abgesehen von den damit verbundenen Kosten könnte sich auch dieser als unzuverlässig herausstellen. Als bessere Lösung erscheint deshalb die Anbringung einer Zugfeder, die die Tür nach dem Öffnen wieder zuzieht. Die Zugfeder substituiert die Tätigkeit des Portiers, was sich darin ausdrückt, daß entsprechende Türschließmechanismen im Französischen auch als "groom" ("Hotelboy") bezeichnet werden. Allerdings haben Zugfedern ihre Tücken: "We have all experienced having a door with a powerful spring mechanism slam in our face. For sure, springs do the work of replacing grooms, but they play the role of a very rude, uneducated porter ... They simply slam the door shut." (Latour 1988: 301) Solche Türschließmechanismen schreiben den Benutzern mithin ein bestimmtes Benutzungsverhalten vor: Man darf nicht zu dicht hinter einer anderen Person hergehen oder muß dessen gewärtig sein, die zurückschlagende Tür mit der Hand aufzuhalten. Es entstehen lokale Benutzungskulturen, externe Besucher, die diese nicht kennen, aber laufen Gefahr, sich blutige Nasen zu holen. "So, inventors get back to their drawing board and try to imagine a nonhuman character that will not prescribe the same rare

---

4. Das Argument verläuft parallel zu Latours Unterscheidung zwischen "ready made science" und "science in the making"; vgl. Latour 1987: 1ff.

local cultural skills to its human users." (ebd.: 302) Eine mögliche Lösung ist der Einbau einer Hydraulik "which easily draws up the energy of those who open the door and retains it, then gives it back slowly with a subtle variety of implacable firmness that one could expect from a well trained butler" (ebd.). Da man aber einige Kraft benötigt, um die Tür zu öffnen, hat man nun das Pendant eines Portiers, der Kindern und alten Menschen den Zugang verwehrt.

Man hat es, dies soll das Beispiel zeigen, bereits bei einfachen technischen Problemlösungen mit einem komplexen Netzwerk von Elementen zu tun, die sich alle in einer bestimmten Art und Weise verhalten müssen, damit ein Zusammenhang entsteht, der die Voraussetzung dafür ist, daß der Türschließer funktioniert. Solche Zusammenhänge lassen sich nicht am Reißbrett planen, sondern sind das Ergebnis einer Vielzahl von Substitutionen und Assoziationen von bzw. zwischen den Elementen, die schließlich den fraglichen Zusammenhang bilden. Zwar versuchen Wissenschaftler (wie etwa Pasteur bei der Entwicklung des Milzbrand-Impfstoffes; vgl. Latour 1983) und Ingenieure (wie die Ingenieure des französischen Elektroautos VEL; vgl. Callon 1987) die Struktur angestrebter Netzwerke zu antizipieren und die relevanten Elemente mit Blick auf das erwünschte Ergebnis in ihrem Verhalten im Vorhinein möglichst weitgehend festzulegen. Solche Bemühungen stoßen aber an Grenzen, denn, so Latour: "Humans, and nonhumans are very, very undisciplined no matter what you do and how many pre-determinations you are able to control upstream of the action." (Latour 1988: 307) Oder anders ausgedrückt: Bereits bei einfachen sozio-technischen Zusammenhängen, wie im Beispiel des Türschließers, können sich in der Art und Weise, in der menschliche und nichtmenschliche Komponenten aufeinander einwirken und dadurch Rückwirkungen erzeugen, Strukturen ausbilden, die in der Form von niemandem erwartet oder beabsichtigt worden sind. Es ist letztlich dieses Merkmal sozio-technischer Netzwerke, das es berechtigt erscheinen läßt, nicht mehr länger von technischer oder sozialer Verursachung zu sprechen, sondern den menschlichen wie den nichtmenschlichen Elementen als Aktanten gleichermaßen Handlungsfähigkeit zuzusprechen.

Genau dieses Merkmal des Entstehens strukturierter Zusammenhänge als Resultat des Verhaltens von Elementen, das nicht einfach einem globalen Plan folgt, ist, wie wir gesehen haben, auch für die VKI das grundlegende Argument der Handlungsfähigkeit ihrer Agenten. Wenn aber bereits die Rekonstruktion einfacher technischer Gegenstände des Alltags diese in einem solchen basalen Sinne als handlungsfähige Einheiten erweist, so fragt sich, wozu es dann gesonderter konstruktiver Überlegungen um die Handlungsfähigkeit von Technik bedarf, wie sie innerhalb der VKI angestrengt werden. Die Antwort auf diese Frage deckt einen zentralen Unterschied zwischen den nichtmenschlichen Aktanten Callons und Latours und selbst noch den einfachsten Agenten der VKI auf: VKI-Agenten sollen befähigt sein, im direkten Kontakt miteinander emergente Problemlösungsmuster zu entwickeln, ohne daß menschliche Akteure an den Interaktionsprozessen teilnehmen. Läßt man dagegen nichtmenschliche Aktanten wie Türen, Türangeln und Hydrauliken allein, dann wird man auf die Entstehung von Netzwerkstrukturen lange warten können.

Was ist es, das VKI-Agenten in die Lage versetzt, aus eigener Kraft emergente Strukturen zu bilden, während es bei den nichtmenschlichen Aktanten der Actor-Network-Theorie dazu des Hinzutretens menschlicher Akteure bedarf? Zur Beantwortung dieser Frage ist es nützlich, wieder auf ein Beispiel zurückzugreifen. Ich wähle zu diesem Zweck ein Multiagenten-System von Connah und Wavish (1990), in dem es um Kooperationsprozesse zwischen einfachen reaktiven Agenten innerhalb eines simulierten experimentellen Szenarios geht: Das Sze-

nario 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: Ein Agent (Agent 1) erreicht sein Möbelstück zuerst, erkennt es als zu schwer, findet keinen hilfsbereiten anderen Agenten und geht in den Hilfe-Modus über. Wenn Agent 2 sein Möbelstück erreicht, ist Agent 1 bereits im Hilfe-Modus, so daß Agent 2 ihn herbeirufen und mit ihm das Möbelstück hinaustragen kann. Anschließend geht Agent 1 erneut auf die Suche nach Hilfe, und löst mit Hilfe von Agent 2, der sich nun im Hilfe-Modus befindet, seine Aufgabe.

Voraussetzung dafür, daß in der Interaktion zwischen Agenten emergente Problemlösungsmuster entstehen, so zeigt das Beispiel, ist erstens der Umstand, daß die Agenten situiert handeln, sich also die Auswahl einer Verhaltensweise aus dem Verhaltensrepertoire des Agenten nach den Gegebenheiten der spezifischen Situation richtet, die zu dem Zeitpunkt vorliegt, an dem die Auswahl getroffen wird, und zweitens der Umstand, daß das Verhalten der anderen Agenten Bestandteil der jeweiligen Situation ist, auf Grund derer ein Agenten sein Verhalten auswählt. Letzteres setzt wiederum voraus, daß der einzelne Agent das Verhalten der anderen Agenten "verstehen" können muß. D.h. ein Agent kann das Verhalten eines anderen Agenten nur dann bei der Auswahl seines Verhaltens berücksichtigen, wenn es ihm in einer Form bekannt gegeben wird, die der Form von Signalen entspricht, die er bei seiner "Entscheidung" zu verarbeiten in der Lage ist. Die Agenten müssen mit anderen Worten eine gemeinsame "Sprache" sprechen. Die Anführungszeichen machen kenntlich, daß es sich bei diesem Verstehen, Entscheiden und einer gemeinsamen Sprache sprechen bei weitem nicht um derart voraussetzungsreiche Prozesse handeln muß, wie im alltäglichen Gebrauch dieser Begriffe mitgedacht ist. In dem eben skizzierten Beispiel bestehen die Entscheidungen der Agenten aus einfachen Reiz-Reaktions-Mechanismen, das Verstehen ist die Wahrnehmung eines Reizes, bei dessen Vorliegen das Verhaltensrepertoire des Agenten eine bestimmte Reaktion vorsieht und die Sprache besteht aus der Festlegung weniger Signale, die sicherstellen, daß das Verhalten eines Agenten von einem anderen Agenten als Reiz aufgegriffen werden kann.

Auch aus der Perspektive der Actor-Network-Theorie ist die Situiertheit des Verhaltens jeglicher Elemente eines Netzwerkes der wesentliche Grund dafür, ihnen als Aktanten Handlungsfähigkeit zuzubilligen, denn auch hier geht es um Verhaltensweisen von Elementen, die durch die Verhaltensweisen anderer Elemente beeinflußt werden und diese ihrerseits beeinflussen. Allerdings ist das Verhalten der nichtmenschlichen Aktanten, zumindest im Fall der von Latour herangezogenen Alltagstechniken, in anderer Weise situiertes Handeln als das Handeln selbst einfacher reaktiver Agenten, und dies in zweifacher Hinsicht: Erstens, was sich abhängig von der jeweiligen Situation verändert, ist nicht das Verhalten der technischen Gegenstände, sondern dessen Bedeutung innerhalb des Netzwerks. Und zweitens, die Deutung der Situation erfolgt nicht durch die nichtmenschlichen Aktanten selbst, sondern durch menschli-

che Akteure. Oder anders ausgedrückt: Es gibt keine gemeinsame Sprache der menschlichen und nichtmenschlichen Aktanten, sondern nur die Sprache der menschlichen Akteure, die beschreibt, was menschliche Akteure oder an ihrer Stelle technische Artefakte zum Ge- oder Mißlingen sozio-technischer Zusammenhänge beitragen.

Der ersten Punkt wird deutlich, wenn man zum Beispiel man einen Türschließmechanismus in Form einer Zugfeder betrachtet. Schwerlich wird man behaupten können, daß dieser Türschließer über Verhaltensalternativen verfügt und diese je nach Situation auswählt. Was sich abhängig vom Verhalten der anderen Aktanten ändert, ist nicht das Verhalten der Zugfeder, sondern die Bedeutung dieses Verhaltens mit Blick auf die Stabilisierung oder Destabilisierung des Netzwerks. Einmal angebracht, wird die Zugfeder die Tür mit schöner Regelmäßigkeit und unschöner Heftigkeit in ihre Ausgangsposition zurückbefördern. Für regelmäßige Benutzer ist dies nach kurzer Zeit genau das, was sie erwarten, sie richten ihr eigenes Verhalten routinemäßig darauf ein und werden nicht länger darüber nachdenken, daß mit der Nutzung der Tür ein Problem verbunden sein könnte. Als Resultat einer spezifischen Akteurskonstellation ist ein stabiles Netzwerk entstanden. Man kann sich ohne weiteres aber auch Situationen vorstellen, in denen das Verhalten dieses Türschließers eine ganz andere Bedeutung gewinnt: Ein unwissender Benutzer verletzt sich an der zurückschlagenden Tür, er beschwert sich und die Beschwerde führt schließlich zur Demontage der Zugfeder. In einem solchen Szenario wird das Verhalten des Türschließers zur Ursache dafür, daß das Netzwerk zusammenbricht.

Dieser Unterschied zwischen den nichtmenschlichen Aktanten der Actor-Network-Theorie und den Agenten der VKI macht deutlich, daß es notwendig ist, zwei Begriffe situierten Handelns zu unterscheiden: genetisch situiertes Handeln und effektiv situiertes Handeln. Genetisch situiertes Handeln liegt vor, wenn die Wahl einer Handlung durch die Gegebenheiten der jeweils vorliegenden Situation beeinflußt wird, effektiv situiertes Handeln dann, wenn die Wirkung oder Bedeutung der Handlung von den situationalen Gegebenheiten des Zusammenhangs abhängt, innerhalb dessen sie stattfindet. Genetische Situiertheit setzt voraus, daß der Akteur (Agent, Aktant) über mehr als eine Handlungsoption verfügt, effektive Situiertheit setzt dies nicht voraus. Der Actor-Network-Theorie gelten bereits solche Elemente als Akteanten, deren Verhalten mit Blick auf die betrachteten Zusammenhänge das Merkmal effektiver Situiertheit aufweisen, wogegen innerhalb der VKI bereits der am wenigsten voraussetzungsreiche Agentenbegriff genetisch situiertes Handeln fordert.

Zum zweiten Punkt: Unter der Bedingung, daß die Verhaltensweisen anderer Akteure Bestandteil der situationalen Gegebenheiten des jeweils handelnden Akteurs sind, kann es genetisch situiertes Handeln nur unter der Voraussetzung geben, daß der Akteur die Fähigkeit besitzt, das Verhalten anderer Akteure als Information zu benutzen. Unter der zusätzlichen Bedingung, daß bestimmte Zustände anderer Akteure, die nicht direkt sichtbar sind (wie etwa die Frage, ob sich einer der Agenten von Connah und Wavish im Hilfe-Modus befindet), ebenfalls Bestandteil der Situation sind, ist es darüber hinaus erforderlich, daß die Akteure eine gemeinsame Sprache sprechen. Nun kann man sich natürlich auch Zusammenhänge vorstellen, die emergente Problemlösungsstrukturen aufweisen, aber innerhalb derer nur ein Teil der Akteure eine gemeinsame Sprache spricht, während daneben andere Akteure beteiligt sind, deren Verhalten jene sprachfähigen Akteure zwar in den ihnen verfügbaren Begriffen als Information bei der Wahl des eigenen Verhaltens nutzen können, die umgekehrt aber nicht selbst in gleicher Weise auf das Verhalten jener Akteure zugreifen können. Genau dieses Verhältnis besteht etwa im Beispiel des Türschließers zwischen den menschlichen und den

nichtmenschlichen Aktanten: Wenn Latour die Zugfeder als unhöflichen Portier beschreibt und die Hydraulik als gut geschulten, aber einige Benutzergruppen diskriminierenden, so sind dies keine Auskünfte über das Verhalten von Türschließmechanismen, die von ihnen selbst stammen, sondern Deutungen menschlicher Akteure. Deutungen, die die Art und Weise strukturieren, in der sich die menschlichen Aktanten auf das Verhalten des Türschließers beziehen, und die, indem sie so oder anders ausfallen können, die effektive Situiertheit dieses Verhaltens ausmachen. Dagegen ist die einzige Information von Seiten der menschlichen Aktanten, die der Türschließer aufnimmt, der Druck (oder Zug), den die Benutzer auf die Tür ausüben. Selbst wenn diese Information nicht als mechanische Einwirkung vonstatten ginge, sondern etwa als Funksignal von der Art, wie sie benutzt werden, um vom Auto aus ein Garagentor zu öffnen und zu schließen, wären die sprachlichen Fertigkeiten des Türschließers noch weit entfernt von der gemeinsamen Sprache der menschlichen Akteure. Und in dem Maße in dem es diese Sprache ist, die die Bedeutung des menschlichen wie der nichtmenschlichen Handelns mit Blick auf ihr Zusammenwirken bestimmt, verschleieren die mit Hilfe der vermeintlich symmetrischen Begriffe des Übersetzungsvokabulars angefertigten Beschreibungen eine grundlegende Asymmetrie, die darin besteht, daß die nichtmenschlichen Aktanten nur deshalb an Stelle menschlicher Akteure handeln, weil die menschlichen Akteure stellvertretend für sie sprechen.

## **5. Die Eigenschaften der Agenten: "ready made" oder "in the making"?**

Soll man eher von Agenten mit festgelegten Eigenschaften ausgehen und die Struktur ihres Zusammenwirkens als von diesen Eigenschaften abgeleitetes Phänomen betrachten, oder soll man eher umgekehrt davon ausgehen, daß die Eigenschaften der Agenten ein Resultat der sozialen Beziehungen ist, in die sie einbezogen sind? Es ist vor allem die unterschiedliche Beantwortung dieser Frage, die zu der Ausdifferenzierung verschiedener Agentenkonzepte innerhalb der VKI führt. Für die Actor-Network-Theorie handelt es sich hierbei nicht um einander ausschließende Alternativen, sondern um einander wechselseitig ergänzende Perspektiven: Die Struktur des Netzwerkes ist ein Resultat des Zusammenwirkens von Aktanten mit bestimmten Eigenschaften und Fertigkeiten, diese Eigenschaften und Fertigkeiten aber sind umgekehrt ein Resultat der Einwirkung anderer Aktanten, die jeden Aktanten eines Netzwerkes in einer bestimmten Weise festlegen (vgl. Latour 1991: 122ff).

Entparodiert wird dieser auf den ersten Blick zirkuläre Zusammenhang, wenn man ihn in der zeitlichen Abfolge betrachtet. Keines der Netzwerke, die Callon und Latour empirisch oder gedankenexperimentell rekonstruieren, beginnt mit Aktanten ohne Eigenschaften. Im Gegenteil, der Prozeß des Netzwerkbildens kommt nur dann in Gang, wenn es einige Aktanten gibt, die sich bereits in einer bestimmten, zumindest vorläufig festgelegten Weise verhalten. Dies sind im Türschließer-Beispiel etwa der Hausmeister, dessen Aufgabe es ist, dafür zu sorgen, daß die Eingangstür nicht offen steht, die Tür, die offen stehen bleibt, wenn sie nicht geschlossen wird, oder die Benutzer, die von Zeit zu Zeit vergessen, die Tür hinter sich zu schließen. Nur auf der Basis solcher vorausgesetzter Eigenschaften von Aktanten kann es wechselseitige "Übersetzungen" geben. Wenn also z.B. der Hausmeister einen automatischen Türschließer einbaut, so deshalb, weil er mit einem bestimmten Verhalten der Benutzer rechnet, nämlich deren Nachlässigkeit beim Schließen der Tür. Zugleich hat diese Einwirkung einen verändernden Einfluß auf das Verhalten der Aktanten: Die Tür bleibt nach dem Öffnen nicht offen stehen, es gibt mit Blick auf das Tür-Problem keine vergeßlichen Benutzer mehr usw. D.h. die zu dem jeweiligen Zeitpunkt bestehenden bzw. erwarteten Eigenschaften und

Verhaltensweisen der Aktanten beeinflussen die Art und Weise, wie diese sich wechselseitig aufeinander beziehen und einwirken. Umgekehrt haben dann diese Beziehungen wieder einen Einfluß auf Verhalten und Eigenschaften der Aktanten. Zu keinem Zeitpunkt aber sind sowohl die Eigenschaften wie auch die Beziehungen der Aktanten vollkommen unbestimmt.

Eines der Probleme der Actor-Network-Theorie besteht darin, diese Doppelseitigkeit des Prozesses der Bildung von Netzwerken in ihren Beispielen deutlicher herausgestellt zu haben, als in ihren theoretischen Annahmen. Während die theoretische Forderung die ist, die *black box* zu öffnen und die Eigenschaften der Aktanten als Resultat des Netzwerkbildens zu rekonstruieren (vgl. Latour 1987: 4), zeigen die Beispiele, daß sich in der konkreten Beobachtung die Perspektive der *ready made technology* und die der *technology in the making* wechselseitig bedingen.<sup>5</sup> Der theoretische Ausgangspunkt verleitet Callon und Latour dazu, bei der Rekonstruktion von Netzwerken grundsätzlich alle Eigenschaften der Aktanten als durch den jeweils ins Auge gefaßten Prozeß des Netzwerkbildens beeinflußbar zu betrachten. Gegen diese Betrachtungsweise läßt sich ein empirischer Einwand erheben: Selbst wenn alle Eigenschaften aller Akanten das Resultat der Beziehungen innerhalb von Netzwerken sind, so ist noch lange nicht gesagt, daß auch die Aktivitäten im jeweils betrachteten Netzwerk einen Einfluß auf alle Eigenschaften der beteiligten Aktanten haben. Oder anders formuliert: Man muß damit rechnen, daß die Eigenschaften einiger Aktanten eines Netzwerkes das Resultat ihrer Involviertheit in andere Netzwerke ist und deshalb, selbst wenn sie die Struktur des betrachteten Netzwerks beeinflussen, nicht umgekehrt selbst in gleicher Weise zur Disposition stehen. So kann etwa die Eigenschaft, ein regelmäßiger oder ein nur gelegentlicher Benutzer der Tür zu sein, einen Einfluß darauf haben, ob die Anbringung einer einfachen Zugfeder zu einem stabilen Netzwerk führt oder nicht. Umgekehrt ist es jedoch sehr unwahrscheinlich, daß irgendeine der "Übersetzungen" innerhalb des Netzwerkes, das sich um das Problem, die Tür geschlossen zu halten, gruppiert, einen Einfluß auf die Häufigkeit der Benutzung hat. Die Häufigkeit der Benutzung ist vielmehr eine Eigenschaft, die durch andere Netzwerke beeinflußt wird, also z.B. dadurch, daß ein Benutzer Mitglied einer Forschergruppe ist, die ihre Büros in dem Gebäude hat.

Auf die Diskussion um die Angemessenheit der unterschiedlichen Agentenbegriffe angewendet, bedeutet dies, daß die Beantwortung der Frage, ob man bestimmte Eigenschaften von Agenten als Voraussetzung oder als Ergebnis der Interaktionsbeziehungen betrachten soll, mittels derer in Multiagenten-Systemen koordiniertes Handeln entsteht, hochgradig davon abhängig ist, welchen Ausschnitt aus der Welt des sozialen Zusammenlebens das Modell wählt, auf dem die Konzeption eines Multiagenten-Systems beruht. Ist das Modellvorbild eine Organisation, die ihre Probleme arbeitsteilig auf der Basis festgelegter Berufsrollen ihrer Mitglieder löst, so wird man sinnvollerweise die Agenten mit Eigenschaften ausstatten, die im Interaktionsprozeß ebensowenig zur Disposition stehen, wie die Berufsrollen der Organisationsmitglieder, die gemeinsam ein Problem bearbeiten. Die Feststellung, daß auch die Entstehung dieser Berufsrollen sozial bedingt ist, oder in der Terminologie der Actor-Network-Theorie: das Resultat wechselseitiger "Übersetzungen" innerhalb von Netzwerken, spielt da-

---

5. Das grundsätzliche Problem, das in dieser Ambivalenz zum Ausdruck kommt, besteht darin, daß sich die Akteure und die Muster ihrer Beziehungen nur dann als Resultat eines sozialen Prozesses (bzw. bei Callon und Latour: als Resultat des Netzwerkbildens) konstruieren oder rekonstruieren lassen, wenn die Eigenschaften der Akteure *und* die Form ihrer Beziehungen zu abhängigen Variablen erklärt werden. Gleichzeitig muß aber jede konstruktive oder rekonstruktive Bemühung an irgendeinem Punkt beginnen und unterstellt damit zwangsläufig die Gegebenheit bestimmter Akteursqualitäten oder Beziehungsstrukturen. Dieses Problem wird von Malsch (in diesem Band) ausführlich behandelt.

bei keine Rolle. Sofern im betrachteten Wirklichkeitsausschnitt die Interaktionsbeziehungen der Akteure keinen Einfluß auf deren Eigenschaften als Inhaber einer bestimmten Berufsrolle haben, können diese bei der Konzeption des Multiagenten-Systems als feststehende Eigenschaften der Agenten betrachtet werden.

Auch dem intentionalen Agentenbegriff liegt ein in einer solchen Weise eingeschränkter Wirklichkeitsausschnitt zu Grunde. In der alltäglichen Begegnung mit anderen Personen reicht es, um ein Verständnis von der Situation zu gewinnen, das es einem erlaubt, sich situationsangemessen zu verhalten, in der Regel aus, der oder den anderen Personen bestimmte Absichten und Ziele, Wünsche und Überzeugungen zu unterstellen, mit Hilfe derer sich ihr Verhalten erklären läßt. Um sich im alltäglichen Umgang mit anderen Menschen erfolgreich zurechtzufinden, muß man sich normalerweise nicht die Frage stellen, auf welche Weise eine andere Person so geworden ist wie sie ist, man muß nicht soziologisch oder psychologisch vorgebildet sein, um sich sozial kompetent verhalten zu können. Auch in der alltäglichen Praxis ist die Vorgehensweise mithin die, nur einen bestimmten Ausschnitt aus der Vielzahl der Beziehungen, die einen Einfluß auf das Verhalten der fraglichen Person haben, als für die vorliegende Situation relevant zu erachten und alles andere als relativ festgelegte Eigenschaften der Person selbst zuzuschreiben. Sofern es um die Modellierung koordinierten Problemlösens nach dem Vorbild von Interaktionszusammenhängen geht, innerhalb derer die Akteure ihre Handlungen auf der Basis solcher Annahmen über ihr Verhalten abstimmen, ergibt es also einen guten Sinn, die Agenten mit bestimmten intentionalen Eigenschaften auszustatten und mit bestimmten Fähigkeiten, über solche mentalen Zustände zu kommunizieren, wie es das Konzept intentionaler Agenten tut.

Genauso sinnvoll kann es umgekehrt sein, den Blick auf soziale Zusammenhänge zu richten, innerhalb derer das Zusammenwirken von Akteuren zugleich auch bestimmte Rückwirkungen auf deren Eigenschaften hervorbringt. Ein Beispiel für die Wahl eines solchen Wirklichkeitsausschnittes als Vorbild der Modellierung von Multiagenten-Systemen ist die sogenannte *scientific community metaphor* (vgl. Hewitt 1977; Kornfeld/Hewitt 1981), eines der frühesten Modelle eines sozialen Problemlösungszusammenhangs innerhalb der VKI. Die gemeinsame Problemlösung ist in diesem Fall die Generierung wissenschaftlicher Erklärungen und die Entscheidung zwischen alternativen Erklärungen. Die Veränderung von Eigenschaften der beteiligten Akteure besteht in einem über die Zeit beobachtbaren Wandel ihrer Überzeugungen darüber, welche Theorie als wahr bzw. als falsch zu gelten hat, ohne den es keinen wissenschaftlichen Fortschritt gäbe. Die Konzeption von Multiagenten-Systemen, die sich an einem solchen sozialen Modell orientiert, muß bestimmte Eigenschaften der Agenten dementsprechend in einer Weise variabel halten, die es erlaubt, daß diese als Resultat der Dynamik ihres Zusammenwirkens entstehen.

Betrachtet man die Eigenschaften, die das Verhalten sozialer Akteure (bzw. assoziierungsfähiger Aktanten) bestimmen, mit Callon und Latour als Resultat ihrer Beziehungen innerhalb von Netzwerken, geht man zweitens davon aus, daß Akteure in mehrere Netzwerke einbezogen sein können, und stellt man drittens in Rechnung, daß sich Netzwerke im Laufe der Zeit (durch neue Assoziationen und Substitutionen) verändern können, also die wechselseitigen Festlegungen der Akteure auf bestimmte Verhaltensweisen jeweils nur "Momentaufnahmen" eines Zeitabschnitts sind, in dem das Netzwerk in einem bestimmten Zustand stabil bleibt, so kommt man auf einer sehr allgemeinen Ebene zu einer Sichtweise, die Annahmen des interaktionistischen Agentenkonzeptes über die Dynamik und Offenheit von Multiagenten-Systemen und die Multiperspektivität ihrer Elemente stützt. Betrachtet man dagegen als Aus-

schnitt aus dem umfassenden Prozeß des Netzwerkbildens bestimmte soziale Abstimmungsprozesse, so kann alles das, was an Festlegungen in dem betrachteten Ausschnitt zeitlich vorher geschehen ist oder unbeeinflußbar von dem fraglichen Zusammenhang innerhalb anderer Netzwerke, dazu führen, daß bestimmte Verhaltensweisen der (oder einiger der) Akteure als weitgehend festgelegt erscheinen.

Ausgehend hiervon kann man die Modelle der VKI entlang eines Kontinuums verorten, das durch zwei Grenzfälle begrenzt wird. Der eine Grenzfall besteht darin, den Ausschnitt der modellierten Interaktionsbeziehungen so klein zu wählen, daß die Festlegung der Eigenschaften der Agenten als etwas betrachtet werden kann, was vollständig außerhalb des fraglichen Zusammenhangs geschieht, und also vom Entwickler des Multiagenten-Systems vorgegeben werden muß. Ein Beispiel hierfür sind die reaktiven Agenten in dem oben beschriebenen Multiagenten-System von Connah und Wavish. Die einzige Dynamik solcher Multiagenten-Systeme, die es berechtigt, von emergenten Problemlösungen zu sprechen, besteht darin, daß die Agenten ihre Verhaltensweisen situationsspezifisch aktivieren. Der andere Grenzfall besteht darin, den Umfang und die zeitliche Ausdehnung der betrachteten Beziehungen zwischen den Agenten so groß zu wählen, daß sämtliche Eigenschaften der Agenten als Resultat bestimmter und als veränderbar durch bestimmte Interaktionsbeziehungen in den Blick kommen. Dies ist die Idealvorstellung des durch den interaktionistischen Agentenbegriff inspirierten Konzeptes offener VKI-Systeme.

Mehr in die eine oder die andere Richtung tendierend, bewegen sich die meisten Ansätze der VKI zwischen diesen beiden Extremen. Auf der einen Seite sind sich auch die Ansätze, die mit reaktiven Agenten experimentieren, darüber im klaren, daß "the major drawback of such agents lies in their hard-wired behavior" (Drogoul/Ferber 1994: 7) und daß sich eine größere Flexibilität nur erreichen läßt, wenn man den Agenten ermöglicht, ihr Verhalten zu ändern. In ähnlicher Weise modifiziert auch der alltagspsychologische Ansatz seine Anfangsannahme, daß die individuellen Intentionen den sozialen Beziehungen der Agenten vorausgehen (vgl. Cohen/Levesque 1987: 410; Werner 1989: 19; Rao et al. 1992: 58), indem den Zielen, Absichten und Überzeugungen nicht zeitlose Gültigkeit, sondern nur eine relative Dauerhaftigkeit zugemessen wird. Unter der Überschrift "An End to Fanaticism" stellen Cohen und Levesque der Vorstellung einer "fanatical persistence - commitment to a goal until it is believed to be achieved or unachievable" (Cohen/Levesque 1990: 222), die ihrer Ansicht nach vernünftigere Vorstellung entgegen, daß Agenten ihre Ziele in Abhängigkeit von ihrer Wahrnehmung des Verhaltens anderer Agenten verändern: "Thus, with the change to relativized persistent goals, we open up the possibility of having a complex web of interdependencies among the agent's goals, intentions and beliefs." (ebd.: 255). Dann aber kommt den individuellen Intentionen bestenfalls eine logische Priorität zu (vgl. Cohen/Levesque 1990: 257), womit Rückwirkungen der sozialen Beziehungen zwischen den Agenten auf deren Eigenschaften nicht länger ausgeschlossen sind.

Auf der anderen Seite kommt der interaktionistische Ansatz auch bei der Analyse und Modellierung relativ offener Interaktionszusammenhänge nicht umhin, die Offenheit des Verhaltens der Agenten begrenzen zu müssen: "Establishing commitments in a manner that is both social and computational means setting up numerous side bets that constrain an agent's field of choice." (Gasser 1991: 132) In diese Richtung geht der Vorschlag, von Situationen auszugehen, in denen die Agenten für bestimmte Probleme über feste Routinen verfügen, während die Lösung anderer Probleme offen ist: "An autonomous intelligent agent may have, at any moment, a wide array of choices about what to believe, what knowledge is relevant,

and what actions to pursue. At any moment, some of these choices are settled, and some are open or unsettled ... Some of this certainty may have been imparted by the designer of a system ... Others (e.g. task allocation) may have been historically settled by problem solving and agreement among agents" (Gasser et al. 1989: 57). Damit aber bewegt sich umgekehrt nun auch der interaktionistische Ansatz ein gutes Stück in Richtung der Vorstellung von Agenten mit relativ dauerhaften oder zum Teil sogar vollständig festgelegten individuellen Eigenschaften.

Ob bestimmte Eigenschaften der Agenten als Voraussetzung oder als Resultat der Interaktionsbeziehungen innerhalb eines Multiagenten-Systems anzusehen sind, hängt mithin von dem Umfang und dem zeitlichen Ausschnitt der modellierten Netzwerkbeziehungen ab. Allerdings enthält die Vorgehensweise, im Gegenzug alle Festlegungen, die zeitlich vorher oder in der Topographie der Netzwerkbeziehungen außerhalb des betrachteten Ausschnittes liegen, als stabile individuelle Eigenschaften der Agenten zu behandeln, ein Konstruktionsproblem, das sichtbar wird, wenn man Multiagenten-Systeme nicht als isolierte Zusammenhänge, sondern als Netzwerke betrachtet, die sich im Kontext weiterer Netzwerkbeziehungen befinden.

## 6. Multiagenten-Systeme im Kontext

Callon und Latour zufolge kann jedes Netzwerk als Bestandteil eines umfassenderen Netzwerkes betrachtet werden. Automatische Türschließer etwa, so Latour, "can do a lot of things to their prescribed users at close range, but most of the effect finally ascribed to them depends on a range of other set-ups being aligned. For instance, the groom closes the door only if there are people reaching the Sociology Department of Walla Walla. These people arrive in front of the door only if they have found maps and only if there are roads leading to it; and, of course, people will start bothering about reading the maps, getting to Washington state and pushing the door open only if they are convinced that the department is worth visiting." (Latour 1988: 308) Um den Prozeß des Netzwerkbildens vollständig zu verstehen, muß das engere Netzwerk, das aus der Tür, dem Türschließer, den Benutzern der Tür und vielleicht auch noch dem Hausmeister besteht, im Zusammenhang weiterer Netzwerke betrachtet werden, die ebenfalls einen Einfluß auf das Verhalten der Akteure des Netzwerks haben. Rechnet man damit, daß die Akteure jener Netzwerke auch wiederum in weitere Netzwerke einbezogen sind, so wird schnell ersichtlich, daß die vollständige Rekonstruktion aller Netzwerkbeziehungen, die direkt oder indirekt auf das Geschehen innerhalb des zunächst betrachteten Netzwerks einwirken, ein einigermaßen aussichtsloses Unterfangen ist. Diesem Umstand trägt Latour im Beispiel des Türschließers Rechnung, indem er das Vorhandensein einer Gruppe von Benutzern zunächst als Gegebenheit voraussetzt, d.h. die Netzwerkanalyse an einem bestimmten Punkt abbricht.

In der Modellbildung von Multiagenten-Systemen haben wir genau die gleiche Strategie beobachtet: Es wird ein bestimmter Ausschnitt aus dem Netzwerk sozialer Beziehungen gewählt, die dem Modell als Vorbild dient, und alle Einflüsse, die außerhalb der Reichweite der Netzwerkbeziehungen dieses Ausschnitts liegen, gelten als Gegebenheiten, wobei eine zusätzliche Strategie darin besteht, diese Gegebenheiten als feststehende Eigenschaften der Agenten zu behandeln. Das aber bedeutet, daß jedes Modell eines Multiagenten-Systems aus zwei Modellen besteht, einem expliziten und einem impliziten. Das explizite Modell richtet sich auf die Beziehungen der Agenten innerhalb des betrachteten Ausschnitts. Vorausgesetzt ist dabei zugleich ein implizites Modell, nämlich eine Vorstellung über die Handlungsumgebung des explizit modellierten Ausschnittes. Jede Eigenschaft eines Agenten, die innerhalb des modell-

lierten Ausschnitts nicht selbst zur Disposition steht, impliziert aus der Netzwerk-Perspektive Annahmen über bestimmte Gegebenheiten der Netzwerke in der Umgebung dieses Ausschnitts.

Dies läßt sich am Beispiel des Konzepts nur relativ dauerhafter Intentionen gut illustrieren, um dessen Vorzüge es Cohen und Levesque geht. Während der "fanatische" Agent ein Ziel solange verfolgt, bis er es erreicht oder es als unerreichbar erkennt, ist der Agent mit nur relativ dauerhaften Intentionen bereit, ein Ziel aufzugeben, zum Beispiel wenn er erfährt, daß ein anderer Agent dieses Ziel bereits erreicht hat (vgl. Cohen/Levesque 1990: 255f). Ist letzteres die vernünftigere Konzeption? Angenommen das Ziel besteht darin, die Summe der von einem Kunden gekauften Waren zu errechnen. Einige Kunden möchten direkt bei der Zusammenstellung der Waren den Gesamtpreis wissen, etwa wenn es um den Kauf einer Einbauküche geht, und der Kunde die Wahl der Komponenten von der Gesamtsumme abhängig macht. Deshalb muß neben dem Kassierer auch der Verkäufer über die Möglichkeit ihrer Errechnung verfügen. Ist es vernünftiger, daß der Kassierer auf eine eigenständige Errechnung der Summe verzichtet, wenn dies der Verkäufer bereits getan hat, und umgekehrt der Verkäufer, wenn der Kunde es nicht verlangt? Dies hängt unter anderem davon ab, ob es als wichtiger gilt, Arbeitszeit und damit Kosten einzusparen, oder Redundanz einzubauen, um die Richtigkeit der Kaufsumme sicherzustellen. Eine Entscheidung darüber wird aber in der Regel nicht innerhalb des hier skizzierten Interaktionszusammenhangs getroffen, sondern vom Unternehmensmanagement vorgegeben. Modelliert man den beschriebenen Zusammenhang als Multiagenten-System, so impliziert die Entscheidung zu Ungunsten des fanatischen Agenten in diesem Fall die Annahme eines primär an der Einsparung von Arbeitskosten interessierten Managements. Ein großer Teil der Kontroversen innerhalb der VKI, was die "realistischere" oder "vernünftigere" Konzeption handlungsfähiger Agenten sei, entsteht, weil diese, in den impliziten Modellen enthaltene Kontextbezogenheit der expliziten Modelle nicht genügend berücksichtigt wird.

Die Faszination der Vorstellung von Multiagenten-Systemen resultiert wesentlich aus dem Bild einer autonomen Emergenz intelligenter Problemlösungen, die - ohne eine direkte äußere Steuerung - als Resultat der Interaktion technischer Artefakte entstehen. Diese Vision verleitet dazu, sich ausschließlich auf die Konstruktion der Multiagenten-Systeme selbst zu konzentrieren und dabei auszublenden, daß tatsächlich zugleich immer auch ein implizites Modell seiner Umgebung mitkonstruiert wird. Ohne eine Einbeziehung der Annahmen der impliziten Modelle ist es, wie das Beispiel eben gezeigt hat, aber kaum möglich, ein Verständnis von der Angemessenheit der expliziten Modelle zu gewinnen. Ist diese Umgebung durch die sozialen Beziehungen menschlicher Akteure strukturiert, so geht es aus dieser Perspektive in der VKI in ganz ähnlicher Weise wie in der Actor-Network-Theorie um sozio-technische Netzwerke.

Die Betrachtung von Multiagenten-Systemen als Bestandteil umfassenderer sozio-technischer Netzwerke läßt ein Konstruktionsproblem von Multiagenten-Systemen besonders deutlich zu Tage treten: Aus der Perspektive des umfassenderen Netzwerks können bestimmte Eigenschaften der modellierten Akteure als durch Netzwerkbeziehungen beeinflußbar analysiert werden, die aus der Perspektive des engeren Netzwerks als unbeeinflußbar gelten müssen, weil sie durch dessen Dynamik nicht tangiert werden. Das Konstruktionsproblem besteht dementsprechend darin, daß, wenn das Multiagenten-System nur das engere Netzwerk modelliert, diese Eigenschaften zwangsläufig als feststehende Gegebenheiten implementiert werden müssen. Damit wird aber zugleich eine möglicherweise unzutreffende Annahme über

das Verhalten des Agenten innerhalb des weiteren Netzwerks impliziert. Ein Beispiel: Eine mögliche Anwendung für Vertragsnetze<sup>6</sup> ist die Abwicklung der Auftragsvergabe im Straßengüterverkehr (vgl. Falk et al. 1993: 29ff). Das Modell eines entsprechenden Multiagenten-Systems besteht im einfachsten Fall aus einem Manager-Agenten und je einem Agenten pro LKW, die als Bieter auftreten. Der Manager schreibt die Aufträge der Spedition aus, die LKW-Agenten machen auf der Basis einer Evaluation ihrer gegenwärtigen Situation (Standort, Ladekapazität etc.) ihre Gebote, und der Manager wählt das günstigste Gebot aus und teilt den Auftrag zu. Nun spielen bei Speditionsaufträgen die Zeitvorgaben der Kunden häufig eine zentrale Rolle. Zugleich aber hängt es nicht nur von dem LKW-Fahrer ab, ob er sein Ziel in einer bestimmten Zeit erreicht, sondern auch von der Zahl der Baustellen, Sperrungen und Staus, die allesamt das Resultat des Verhaltens von Akteuren des umfassenderen Netzwerkes sind, aber umgekehrt nicht beeinflußbar durch das Verhalten der Agenten dieses Multiagenten-Systems. In dem Maße, in dem solche Kontextfaktoren aus der Perspektive des modellierten Zusammenhangs unerwartet auftreten, enthalten die Parameter auf deren Basis im Multiagenten-System Fahrzeiten berechnet, Gebote gemacht und Aufträge erteilt werden, Annahmen über die Umgebung des Systems, die sich bei diesem tatsächlichem Einsatz sehr schnell als unrealistisch erweisen dürften.

Für dieses Problem, das erst sichtbar wird, wenn man Multiagenten-Systeme als Bestandteil umfassenderer Netzwerke betrachtet, gibt es zwei Lösungen: Die eine Lösung besteht darin, den gewählten Ausschnitt solange auszudehnen, bis alle weiteren Kontextbedingungen unproblematisch als feststehende Gegebenheiten behandelt werden können. Die andere Lösung ist, von vornherein nur solche Interaktionszusammenhänge zu modellieren, die innerhalb hinreichend stabiler Umgebungen ablaufen, bzw. das Problem, um dessen Lösung es innerhalb des Multiagenten-Systems geht, so zuzuschneiden, daß die im Modell ausgeklammerten externen Einflüsse für die Bearbeitung des Problems ohne Bedeutung sind. Die erste Lösung hat den Nachteil, daß sich der Umfang der zu berücksichtigenden Agenten und ihrer Beziehungen sehr schnell in einer Weise ausdehnen kann, die ein solches Vorgehen undurchführbar werden läßt. So müßte man etwa, um ein vollständiges Bild der wechselseitigen Einflüsse zu gewinnen, die das Verhalten des LKW-Agenten beeinflussen, alle anderen Verkehrsteilnehmer, die die fraglichen Strecken benutzen, ebenfalls als Agenten modellieren. Die zweite Lösung ist deshalb in vielen Fällen wahrscheinlich die praktikablere Option. Im Beispiel des Speditionsagentensystems könnte dies etwa bedeuten, zeitkritische Frachtaufträge auszuklammern.

Das Erfordernis einer relativ stabilen Umgebung, die es ermöglicht, äußere Einflüsse, die für die Problemlösung relevant sind, aber nicht im Einflußbereich des Interaktionszusammenhangs der modellierten Agenten liegen, innerhalb des Multiagenten-Systems als feststehende Gegebenheiten zu behandeln, lenkt den Blick auf die besondere Bedeutung des koordinierten Problemlösens in Organisationen als Vorbild der Modellierung von Multiagenten-Systemen. Folgt man Thompson (1967), so basiert organisationales Problemlösen innerhalb unsicherer Umgebungen auf einer spezifischen Form interner Differenzierung: "organizations cope with

---

6. In der Vertragsnetz-Architektur von Davis und Smith (1983: 77ff) wird die Verteilung der Aufgaben im System als Konkurrenz um Aufträge organisiert. Die Agenten handeln als Manager und als Bieter. Als Manager nimmt der Agent, der eine Aufgabe ausgeführt haben möchte, eine Ausschreibung vor. Agenten, die sich auf der Basis eigener Evaluation für die Ausführung dieser Aufgabe anbieten wollen, senden ein Gebot. Unter den Geboten wird vom Manager - ebenfalls aufgrund eigener Evaluationen - das geeignete Gebot auswählt und dem entsprechenden Bieter der Zuschlag erteilt. Der Bieter übernimmt die Ausführungsverantwortung und wird damit zum Kontraktor. Er kann seinerseits die Aufgabe weiter zerlegen und selbst Teilaufgaben ausschreiben.

uncertainty by creating certain parts specifically to deal with it, specializing other parts in operating under conditions of certainty or near certainty" (ebd.: 13). Eine formale Organisation "contains a suborganization whose 'problems' are focused around effective performance of the technical function ... such as the materials which must be processed and the kinds of cooperation of different people required to get the job done effectively" (ebd.: 10). Im Interesse effizienten Problemlösens wird dieser Organisationsbereich, den Thompson als "technischen Kern" bezeichnet, weitgehend abgeschirmt gegen äußere Einflüsse, die nicht der Kontrolle der Organisation unterliegen: "Under norms of rationality, organizations seek to seal off their core technologies from environmental influences." (ebd.: 19) So lassen sich viele Problemlösungszusammenhänge effizient nur dann durchführen, wenn sie gegenüber Engpässen bei der Zuführung erforderlicher Ressourcen oder gegenüber Schwankungen bei der Nachfrage der erzeugten Lösungen abgedeckt sind. Dies gilt für die Zulieferung von Vorprodukten und den Absatz von Waren einer industriellen Produktionseinheit genauso wie für die Versorgung der medizinischen Abteilung eines Krankenhauses mit geschultem Personal und die Einweisung von Kranken (vgl. ebd.).

Thompson will nicht behaupten, daß sich Organisationen insgesamt gegen Unsicherheiten ihrer Umwelt abschotten können. Im Gegenteil, "both input and output activities are interdependent with environmental elements. Organizational rationality, therefore, never conforms to closed-system logic but demands the logic of an open system." (ebd.: 19f). Sein zentrales Argument ist, daß die Leistungsfähigkeit von Organisationen, betrachtet als Zusammenhänge effizienten Problemlösens, darin besteht, die erforderliche Offenheit gegenüber externen Einflüssen mit einer möglichst großen Geschlossenheit ihres technischen Kerns zu verbinden. Abhängig von der Struktur dieses Kernbereichs stehen dabei verschiedene Strategien zur Verfügung, die zu unterschiedlichen Organisationstypen von der Adhocratie bis zur Hierarchie führen (vgl. ebd.: 52ff). Ich gehe darauf nicht näher ein. Der springende Punkt für den vorliegenden Zusammenhang ist, daß in Organisationen, so wie Thompson sie beschreibt, die Verschränkung unterschiedlich konvergenter und irreversibler Netzwerke in genau einer solchen Weise stattfindet, wie sie erforderlich ist, um das oben beschriebene Konstruktionsproblem von Multiagenten-Systemen zu lösen.

"(V)iewing distributed systems as analogous to *human organizations*" (Fox 1981: 70) ist aus dieser Perspektive deutlich mehr als nur einer unter vielen Vorschlägen, welche sozialen Zusammenhänge als Vorbild der Konstruktion von Multiagenten-Systemen in Frage kommen. Die Fokussierung auf organisationale Formen der Verschränkung hochgradig konvergenter Netzwerke mit offeneren Netzwerken, die sie abpuffern, erscheint vielmehr als eine gegenwärtig besonders erfolgversprechende Strategie, praktisch nutzbare Multiagenten-Systeme, und das heißt: Multiagenten-Systeme, die Bestandteil umfassenderer Netzwerke sind, zu konstruieren. Wenn alle Netzwerke sozialer Beziehungen zwischen menschlichen Akteuren in dem Sinne offene Systeme sind, daß in ihrer Umgebung stets weitere Netzwerkbeziehungen entstehen können, die einen Einfluß auf den betrachteten Zusammenhang ausüben, und wenn andererseits Multiagenten-Systeme in dem Sinne geschlossene Systeme sind, daß es für Zusammenhänge koordinierten Problemlösens erforderlich ist, den Umfang der berücksichtigten Wechselwirkungen zu begrenzen, dann ist es unter Anwendungsgesichtspunkten erstrebenswert, Multiagenten-Systeme zu konstruieren, die sich als geschlossene Systeme innerhalb offener Systeme von Netzwerkbeziehung angemessen verhalten. Zunächst wird dies am ehesten im Bereich organisationalen Problemlösens zu erreichen sein.

Einschränkend muß allerdings auf die Befunde der neueren Organisationsforschung hingewiesen werden, wonach die Annahme einer auf Problemlösungseffizienz gerichteten organisationalen Rationalität durch die empirische Wirklichkeit bei weitem nicht in dem Maße gedeckt ist, wie dies die - etwa in Gestalt von Organigrammen offiziell festgelegte - formale Struktur von Organisationen suggeriert. Vielmehr sind die unterschiedlichen Organisations-einheiten häufig nur locker miteinander verknüpft, einander in wechselnden Koalitionen verpflichtet, die sich eher an Partialinteressen als an übergreifenden Organisationszielen orientieren, welche letzteren vielfach nur vage oder gar mehrdeutig formuliert sind. Aus genau dieser Offenheit und Ambiguität resultiere dann aber auch die Dynamik und Adaptivität von Organisationen (vgl. March/Olsen 1976). In gewissem Maße gilt dies selbst für technisch weitgehend durchrationalisierte Organisationsbereiche. Wie Kern und Schumann (1984) herausgestellt haben, werden qualifizierte Arbeitskräfte im Prozeß zunehmender industrieller Rationalisierung deshalb erneut zu einem wichtigen Produktivitätsfaktor werden, weil nur sie die "verbleibenden Empirielücken" (ebd.: 167) zu schließen, d.h. die situationale Anpassung des technisch-organisationalen Regelwerks an unvorhergesehene Ereignisse zu leisten in der Lage sind.

Von einer Abschottung des technischen Kerns, wie von Thompson konstatiert, kann dann nur in einem graduellen Sinne die Rede sein. Zugleich rückt damit die Fähigkeit menschlicher Akteure, bestehende Regeln situational anzupassen auch in relativ stark verregelten Zusammenhängen organisationalen Problemlösen stärker in den Mittelpunkt der Aufmerksamkeit. In der Terminologie der Actor-Network-Theorie heißt das: etablierte Beziehungen im engeren Netzwerk im Licht bislang unberücksichtigter Einwirkungen des umfassenderen Netzwerks zu deuten. Ist man im Zusammenhang mit Multiagenten-Systemen an dieser Problemstellung interessiert, so wird man sich eher als um spezifische Anwendungsdomänen um die allgemeinere Frage kümmern müssen, worin das Expertentum sozialer Akteure besteht, vorgegebene Handlungsmuster situationsangemessen verwenden zu können (vgl. Schulz-Schaeffer/Malsch, in diesem Band).

## 7. Ergebnis

Als Ergebnis des Versuchs, die konstruktiven Bemühungen der VKI und die rekonstruktiven Bemühungen der Actor-Network-Theorie um die Handlungsfähigkeit von Technik aufeinander zu beziehen, lassen sich drei Punkte festhalten: (1) Die Bedeutung des Unterschieds zwischen genetisch und effektiv situiertem Verhalten für die Frage der Handlungsfähigkeit technischer Agenten bzw. Akteuren; (2) die Abhängigkeit der Festgelegtheit ihrer Eigenschaften vom Umfang der betrachteten bzw. modellierten Netzwerkbeziehungen und (3) der Zusammenhang zwischen dem expliziten Modell eines Multiagenten-Systems und dem zugleich implizierten Modell der relevanten Netzwerkumgebung, der die Konstruktion von Multiagenten-Systemen zu einem Bestandteil der Konstruktion umfassender sozio-technischer Netzwerke werden läßt.

In ähnlicher Weise beginnen die VKI und die Actor-Network-Theorie mit einem sehr allgemeinen Begriff von Handlungsfähigkeit. Als Handlungen gelten demnach Verhaltensweisen, die einen Beitrag zur Strukturierung des Zusammenwirkens mehrerer Elemente eines Zusammenhangs leisten, unter der Voraussetzung, daß die Struktur, die entsteht, ein emergentes Resultat dieser Verhaltensweisen ist, also nicht einem globalen Plan folgt, durch den die einzelnen Verhaltensweisen direkt gesteuert werden. Dieser Begriff des Handelns ist identisch mit dem des situierten Handelns, denn die Emergenz der fraglichen Strukturen ist ein Resultat

der jeweiligen situationalen Gegebenheiten, die die einzelnen Handlungen bestimmen. Wie wir gesehen haben, kann man jedoch zwei Formen unterscheiden, in denen die Situation auf eine Handlung einwirkt: auf deren Entstehung oder auf deren Wirkung. Genetisch situatives Handeln setzt voraus, daß der Akteur über mehrere Handlungsoptionen und eine wie auch immer rudimentäre Sprache verfügt, die es ihm erlaubt, bestimmte Gegebenheiten der Situation als Information bei der Handlungswahl zu nutzen. Effektiv situiertes Handeln dagegen setzt lediglich voraus, daß seine Bedeutung von der Situation abhängt. Indem die Actor-Network-Theorie den allgemeinen Begriff von Handlungsfähigkeit nicht weiter differenziert, übersieht sie die Asymmetrie zwischen genetisch situiert und effektiv situiert handelnden Aktanten und versäumt es dementsprechend zu fragen, wessen Sprache es ist, die die Bedeutung der einzelnen Verhaltensweisen festlegt. Die konstruktiven Bemühungen der VKI um die genetische Situiertheit und eigene Sprachfähigkeit ihrer Agenten treiben mithin die Symmetrie von menschlichen Akteuren (oder allgemeiner: sozialen Lebewesen) und technischen Artefakten ein gutes Stück in die Richtung voran, die die Actor-Network-Theorie in einer zu oberflächlichen Betrachtung bei der Rekonstruktion einfacher Alltagstechniken bereits konsstatieren zu können meint.

Die Actor-Network-Theorie geht davon aus, daß die Eigenschaften eines Aktanten als Resultat der Einwirkungen des Netzwerkes aller Beziehungen betrachtet werden können, in die der Aktant eingebunden ist. Tatsächlich betrachten Callon und Latour bei ihren rekonstruktiven Bemühungen und die Proponenten des einen oder anderen Agentenbegriffes im Rahmen ihrer konstruktiven Bemühungen aber stets nur einen bestimmten Netzwerkausschnitt. Wie wir gesehen haben, ist es von der Wahl dieses Ausschnitts abhängig, ob bestimmte Eigenschaften der Aktanten bzw. Agenten als den jeweils betrachteten Netzwerkbeziehungen vorausgesetzt oder als durch diese beeinflußbar konzipiert werden müssen. Vor diesem Hintergrund macht es wenig Sinn, eine allgemeine Diskussion darüber zu führen, welches Agentenkonzept das realistischere oder das vernünftigere ist.

Betrachtet man Multiagenten-Systeme als Bestandteil umfassenderer Netzwerke, so ist die Angemessenheit einer Konzeption zugleich abhängig von den in ihr implizierten Annahmen über die Umgebung des explizit modellierten Netzwerkausschnitts. Es entsteht das Konstruktionsproblem, daß bestimmte Eigenschaften der Agenten, die innerhalb des modellierten Zusammenhangs von Bedeutung sind, aber nicht dem Einfluß der Beziehungen zwischen den Agenten unterliegen, als feststehende Eigenschaften implementiert werden müssen, damit aber gleichzeitig deren mögliche Veränderbarkeit durch Einflüsse des umfassenderen Netzwerks ignoriert wird. Aus der Anwendungsperspektive ist der erfolgversprechendste Weg, mit diesem Problem fertig zu werden, sich bei der Konstruktion von Multiagenten-Systemen zunächst auf solche Muster koordinierten Problemlösens zu konzentrieren, die innerhalb stabilisierter Handlungsumgebungen ablaufen. Damit gewinnen die Formen des organisationalen Problemlösens als soziale Modelle der VKI besondere Bedeutung. Ein einfaches Erfolgsrezept hat man damit aber, wie die Diskussion in der neueren Organisationsforschung über die Rationalität von Organisationen zeigt, noch nicht.

Eine zentrale Idee der Actor-Network-Theorie besteht darin, den Begriff der Handlungsfähigkeit so weit zu fassen, daß alle Elemente, die dazu beitragen, daß ein emergentes Resultat entsteht, als Aktanten betrachtet werden können. Die Auseinandersetzung mit den Agentenbegriffen der VKI zeigt, daß tatsächlich vergleichbar nur solche Aktanten sind, die genetisch situiert, also einer eigenen Deutung der jeweiligen Situation gemäß so oder anders handeln. Hier lassen sich die Überlegungen der VKI nutzen, um eine verborgene Asymmetrie der Ac-

tor-Network-Theorie aufzudecken. Umgekehrt macht die Actor-Network-Theorie deutlich, daß es auch mit Blick auf die Konstruktion von Multiagenten-Systemen von wesentlicher Bedeutung ist, die Handlungsfähigkeit der Agenten relativ zu den jeweils betrachteten Netzwerkbeziehungen zu konzipieren.

Besteht das Ziel nicht in erster Linie darin, mit Hilfe von Multiagenten-Systemen sozialwissenschaftliche Grundlagenforschung zu betreiben, sondern in der Entwicklung nutzbarer Techniken - und das ist die derzeit vorherrschende Zielrichtung innerhalb der VKI (vgl. Castelfranchi 1992: 206) -, dann geht es darum, den Umfang des Netzwerks und die Handlungsmöglichkeiten der Agenten auf solche Beziehungen und solche Handlungsoptionen zu begrenzen, von denen man sich verspricht, daß sie für die kooperative Lösung des jeweils ins Auge gefaßten Problems erforderlich und ausreichend sind. Soweit das kooperative Problemlösen innerhalb organisationaler Zusammenhänge auf einer solchen Begrenzung beruht, ist die anwendungsbezogene VKI gut beraten, besonders hier nach den sozialen Modellen zu suchen, die sie ihren Systemarchitekturen zu Grunde legt.

Der in der Konzeption von Multiagenten-Systemen als offene Systeme enthaltene Einwand, daß man von Handlungsfähigkeit erst dann sprechen kann, wenn die Agenten mit unvorhergesehenen Problemen in unerwarteter Weise umzugehen in der Lage seien (vgl. Malsch, in diesem Band), wird aus dieser Perspektive sowohl relativiert wie auch bestätigt. Wenn es um Multiagenten-Systeme geht, deren Zweck es ist, bestimmte Aufgaben zu erledigen, dann, so Burkhard (1995: 299), "programs shall not really behave freely as humans may". Das Handeln der Agenten soll nicht ergebnisoffen sein, sondern sich in erwartbarer Weise an Aufgaben und Pflichten orientieren. Was Castelfranchi (1990: 50) als "pre-established harmonies" kritisiert, in denen die VKI bislang gefangen sei, ist aus der Perspektive der Technikentwicklung durchaus anstrebenswert: Agenten, die sich so verhalten, daß zu erwarten ist, daß sie gemeinsam die Probleme lösen, zu deren Lösung das Multiagenten-System gedacht ist. Auch die menschlichen Vorbilder der Agenten sind in Zusammenhängen koordinierten Problemlösens bei der Wahl ihrer Handlungsmöglichkeiten nicht beliebig frei. Denn auch die Organisierung des gemeinsamen Problemlösens menschlicher Akteure beruht wesentlich darauf, die Ergebnisoffenheit des Prozesses zu verringern, indem die Handlungsoptionen der Beteiligten begrenzt werden, etwa durch Spezialisierung von Fertigkeiten oder Festlegung von Zuständigkeiten. Zugleich gilt, daß deren Handeln nichtsdestotrotz sehr viel mehr von den eigenen Ausdeutungen ihrer Rollen, von heterogenen Präferenzen, unvollständiger Information usw. bestimmt ist als die formalen Organisationsstrukturen glauben machen wollen und daß gerade dies ein wesentlicher Faktor der Adaptivität von Organisationen ist. Und dort, wo sich die VKI zum Ziel setzt, diese Form von Problemlösungskompetenz technisch nachzubilden, steckt sie sofort wieder in jenen Problemen, die ein Organisationsmodell Thompsonscher Prägung zu umgehen versucht.

## Literatur

- Avouris, Nicholas M. / Les Gasser (Hg.) (1992): *Distributed Artificial Intelligence: Theory and Praxis*, Dordrecht u.a.: Kluwer Academic Publishers
- Bateson, Gregory (1972): *Steps to an Ecology of Mind*, San Francisco
- Bijker, W. E. (1993): Do Not Despair: There is Life after Constructivism , in: *Science, Technology & Human Values* 18(1), S. 113-138

- Bijker, W. E. / Th. P. Hughes / T. Pinch (Hg.) (1987): The Social Construction of Technological Systems. New Directions in the Sociology and History of Technology, Cambridge, Mass. u.a.: MIT Press
- Bijker, W. E. / J. Law (Hg.) (1992): Shaping Technology / Building Society. Studies in Sociotechnical Change, Cambridge, Mass. u.a.: MIT Press
- Bloor, David (1976): Knowledge and Social Imagery, London
- Burkhard, Hans-Dieter (1995): Agent-Oriented Programming for Open Systems, in: M. J. Wooldridge / N. R. Jennings (Hg.), S. 291-306
- Callon, M. (1986): Some Elements of a Sociology of Translation: Domestication of the Scallops and the Fishermen of St Brieuc Bay , in: J. Law (Hg.), S. 196-23
- Callon, M. (1987): Society in the Making: The Study of Technology as a Tool for Sociological Analysis , in: W. E. Bijker et al. (Hg.), S. 82-103
- Callon, M. (1991): Techno-Economic Networks and Irreversibility , in: J. Law (Hg.), S. 132-161
- Callon, M. / B. Latour (1992): Don't Throw the Baby Out with the Bath School! A Reply to Collins and Yearley , in: A. Pickering (Hg.), S. 343-368
- Castelfranchi, Cristiano (1990): Social Power. A Point Missed in Multi-Agent, DAI and HCI, in: Y. Demazeau / J.-P. Müller (Hg.), S. 49-63
- Castelfranchi, Cristiano (1992): No More Cooperation, Please! In Search of the Social Structure of Verbal Interaction, in: A. Ortony et al. (Hg.), S. 205-227
- 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 / E. Werner (Hg.) (1994): Artificial Social Systems, 4th European Workshop on Modelling Autonomous Agents in a Multi-Agent World, MAAMAW '92, S. Martino al Cimino, Italy, July, 29-31, 1992, Selected Papers, Lecture Notes in Artificial Intelligence 830, Berlin u.a.: Springer-Verlag
- Cohen, Philip R. / Hector J. Levesque (1987): Intention = Choice + Commitment, in: Proceedings AAAI-87, Seattle, WA., S. 410-415
- Cohen, Philip R. / Hector J. Levesque (1990): Intention is choice with commitment, in: Artificial Intelligence 42, S. 213-261
- Collins, Harry M. (1990): Artificial Experts. Social Knowledge and Intelligent Machines, Cambridge, Mass. u.a.: MIT Press
- Connah, David / Peter Wavish (1990): An Experiment in Cooperation, in: Y. Demazeau / J.-P. Müller (Hg.), S. 197-212
- 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
- Demazeau, Yves / Jean-Pierre 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)
- 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.), S. 3-23
- Falk, J. / S. Speck / P. Mertens (1993): Unterstützung der Lager- und Transportlogistik durch Teileintelligente Agenten, in: Information Management 2/1993, S. 26-31
- Fikes, R. / E. Sandewall (Hg.) (1991): Proceedings of the 2nd International Conference on Principles of Knowledge Representation and Reasoning (KR-91)

- Fox, M. S. (1981): An Organizational View of Distributes Systems, in: IEEE Transactions on Systems, Man, and Cybernetics 11, 1981, S. 70-80
- Gasser, Les (1991): Social Conceptions of Knowledge and Action: DAI Foundations and Open Systems Semantics, in: Artificial Intelligence 47, S. 107-138
- Gasser, Les (1992): Boundaries, Identity, and Aggregation: Plurality Issues in Multiagent Systems, in: E. Werner / Y. Demazeau (Hg.), S. 199-213
- Gasser, Les / Jean-Pierre Briot (1992): Object-based concurrent Programming and Distributed Artificial Intelligence, in: N. M. Avouris / L. Gasser (Hg.): Distributed Artificial Intelligence: Theory and Praxis, Kluwer
- Gasser, Les / M. N. Huhns (Hg.) (1989): Distributed Artificial Intelligence, Volume II, London: Pitman, San Mateo, Ca.: Morgan Kaufmann Publishers
- 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.) (1989), S. 55-78
- Hewitt, Carl E. (1977): Viewing Control Structures as Patterns of Passing Messages, in: Artificial Intelligence 8, 1977, S. 323-364
- Hewitt, Carl E. (1991): Open Information Systems Semantics for Distributed Artificial Intelligence, in: Artificial Intelligence 47, S. 79-106
- Kern, Horst / Michael Schumann (1984): Das Ende der Arbeitsteilung? Rationalisierung in der industriellen Produktion, München: Beck
- Knorr-Cetina, K. / M. Mulkay (Hg.) (1983): Science Observed. Perspectives on the Social Study of Science, London: Sage
- Kornfeld, William A. / Carl E. Hewitt (1981): The Scientific Community Metaphor, in: IEEE Transactions on Systems, Man and Cybernetics 11(1), S. 24-33
- Latour, Bruno (1983): Give Me a Laboratory and I will Raise the World , in: K. Knorr-Cetina / M. Mulkay (Hg.): Science Observed. Perspectives of the Social Studies of Science, London u.a., S.141-170
- Latour, Bruno (1987): Science in Action. How to Follow Scientists and Engineers through Society, Cambridge, Mass.: Harvard University Press
- Latour, Bruno (1988): Mixing Humans and Nonhumans Together. The Sociology of a Door-Closer , in: Social Problems 35 (3), 1988, S. 298-310
- Latour, Bruno (1991): Technology is Society Made Durable , in: J. Law (Hg.), S. 103-131
- Latour, Bruno (1992): Where are the Missing Masses? The Sociology of a Few Mundane Artifacts , in: Bijker / Law (Hg.), S. 225-258
- Law, J. (Hg.) (1986): Power, Action and Belief. A New Sociology of Knowledge?, London u.a.: Routledge & Kegan Paul
- Law, J. (Hg.) (1991): A Sociology of Monsters: Essays on Power, Technology and Domination, Sociological Review Monograph 38, London u.a.: Routledge
- March, James G. / Johan P. Olsen (1976): Ambiguity and Choice in Organizations, Bergen u.a.: Universitetsforlaget
- Moulin, Bernard / Brahim Chaib-draa (1996): An Overview of Distributed Artificial Intelligence, in: G. M. O'Hare / N. R. Jennings (Hg.), S. 3-55
- Moulin, B. / L. Cloutier (1994): Collaborative Work Based on Multiagent Architectures: A Methodological Perspective, in: Fred Aminzadeh / Mohammed Jamshidi (Hg.): Soft Computing. Fuzzy Logic, Neural Networks, and Distributed Artificial Intelligence, Englewood Cliffs, NJ: Prentice Hall, S. 261-296
- O'Hare, G. M. / N. R. Jennings (Hg.) (1996): Foundations of Distributed Artificial Intelligence, New York u.a.: John Wiley & Sons

- Ortony, Andrew / Jon Slack / Oliviero Stock (Hg.) (1992): Communication from an Artificial Intelligence Perspective. Theoretical and Applied Issues, Berlin u.a.: Springer-Verlag
- Pickering, Andrew (Hg.) (1992): Science as Practice and Culture, Chicago u.a.: University of Chicago
- Rao, Anand S. / M. P. Georgeff (1991): Modelling Rational Agents within a BDI Architecture, in: R. Fikes / E. Sandewall (Hg.), S. 473-484
- Rao, Anand S. / M. P. Georgeff / Elizabeth A. Sonenberg (1992): Social Plans: A Preliminary Report, in: E. Werner / Y. Demazeau (Hg.), S. 57-76
- Shoham, Yoav (1993): Agent-oriented Programming, in: Artificial Intelligence 60(1), S. 51-92
- Shoham, Yoav (1994): Multi-Agent Research in the Knobotics Group, in: C. Castelfranchi / E. Werner (Hg.), S. 271-294
- Thompson, James D. (1967): Organizations in Action. Social Science Bases of Administrative Theory, NewYork u.a.: McGraw
- Werner, Eric (1989): Cooperating Agents: A Unified Theory of Communication and Social Structure, in: L. Gasser / M. N. Huhns (Hg.) (1989), S. 3-36
- Werner, Eric / Yves Demazeau (Hg.) (1992): Decentralized A.I. 3, Proceedings of the Third European Workshop on Modelling Autonomous Agents in an Multi-Agent World, North-Holland: Elsevier
- Wooldridge, Michael J. / Nicolas R. Jennings (1995): Agent Theories, Architectures, and Languages: A Survey, in: dies. (Hg.), S. 1-39
- Wooldridge, Michael J. / Nicolas R. Jennings (Hg.) (1995): Intelligent Agents. ECAI-94 Workshop on Agent Theories, Architectures, and Languages, Amsterdam, The Netherlands, August 8-9, 1994, Proceedings, Berlin u.a.: Springer-Verlag