

**Soziale Aspekte in
Multiagentensystemen:
Institutionen und soziale Normen**

Seminararbeit

Erstellt im Rahmen des Projektseminars
„Logiken für Multiagentensysteme“
WS 2003/04

Gila Dinter
Matr.-Nr. 5009318

Zusammenfassung

Diese Arbeit befasst sich mit der Fragestellung, wie sich soziale Normen in einem Multiagentensystem modellieren lassen. Der Beschreibung der Modellierung geht eine Erläuterung der soziologischen Normkonzeption voraus.

Inhaltsverzeichnis

Einleitung	1
1 Kollektives Handeln in Form von Organisationen	4
2 Institutionen und soziale Normen – eine soziologische Begriffsklärung	6
2.1 Institutionen	8
2.1.1 Funktion von Institutionen	10
2.1.2 Sanktionen und Legitimität	11
2.2 Soziale Normen	13
3 Soziale Situationen des Normbedarfs	15
3.1 Das Problem der Koordination	16
3.2 Soziale Dilemmasituationen	18
3.3 Konflikt	21
3.4 Gegenüberstellung der Normtypen	22
4 Modellierung sozialer Normen in einem MAS	24
4.1 Mögliche Modellierungskonzepte	24
4.2 Normatives Wissen	25
4.3 Vom normativen Wissen zur Normadoption	27
5 Bewertung der Modellierung	30
Literaturverzeichnis	32
Abbildungsverzeichnis	33

Einleitung

Im Forschungsfeld der klassischen Künstlichen Intelligenz (KI) gingen die Bestrebungen, Systeme zu entwerfen, die mehr oder weniger intelligente Leistungen vollbringen, traditionell von einer monolithischen Konzeption aus. Das Paradigma der Orientierung am menschlichen Gehirn fand in der Forschung eine Ablösung mit der Etablierung des jüngeren Zweiges der Verteilten Künstlichen Intelligenz (VKI), dessen wesentliche Neuerung in der Annahme liegt, Intelligenz resultiere nicht aus dem Wirken einer zentralen Komponente, sondern emergiere infolge der Interaktion einer Menge autonomer Einheiten.

Einher ging mit dieser Umorientierung ein zunehmender Verweis auf Begriffe, deren Gebrauch im Kontext der Interaktion und des aufeinander bezogenen Verhaltens einzelner Akteure üblich ist. Somit etablierte sich aufgrund der Natur der angestrebten Systeme (Multiagentensystem) die Verwendung eines Vokabulars, das seinen Ursprung und seine Fundierung in der Soziologie erfuhr, deren Forschungsgegenstand sich genau im großen Feld des zwischenmenschlichen Geschehens, dem „Sozialen“, niederschlägt. Dabei erfolgte die Einsetzung der Begrifflichkeiten jedoch vielmehr in einem nahezu rein metaphorischen Sinne, als dass die mit ihnen referenzierten Systemkomponenten oder Eigenschaften tatsächlich Anlehnung an soziologische Konzepte aufwiesen.

Die erste aufkommende Sozialmetapher war die „Verhandlung“, die aus der Formulierung des Unentscheidbarkeitsproblems der VKI durch Carl Hewitt 1977 resultierte:¹ Ein logisch unentscheidbarer Widerspruch, der entsteht, wenn zwei in sich geschlossene „Mikrotheorien“ bzgl. derselben Fragestellung zu widersprüchlichen, aber gleichermaßen logischen Lösungen kommen, lässt sich nicht durch Kalküle aufheben, sondern erfordert die Fähigkeit, miteinander in Verhandlungen zu treten, wie es im realen Zusammenleben von Menschen der Fall ist. Es folgten weiter Metaphern wie Rollenerwartung, Gemeinschaft, Gruppen, Kommunikation, Macht, Vertrauen u. ä., deren Modellierung oder Implementation sich jedoch nicht in einem interdisziplinären Rahmen ereignete. Die Konzepte wurden dagegen in einem intuitiven Sinne seitens der KI-Forschung verwendet. Machte diese eingeschränkte Sicht die Entwicklung in den Anfängen einfacher, ließ man doch gleichermaßen die in der Soziologie erarbeiteten Erkenntnisse bzgl. dieser Phänomene und die damit verbundenen Lösungspotentiale außer acht.

¹vgl. [Mal97].

Erst langsam entwickelte sich eine zunehmend interdisziplinäre Ausrichtung der KI-Forschung, die sich um eine präzise Verwendung von Begriffen anderer Wissenschaften bemühte. Zunächst kam es zu einer Fruchtbarmachung von Bereichen der Wirtschaftswissenschaften wie der Spieltheorie und der Rational Choice-Theorie. Seit etwa zehn Jahren ist das Bemühen um eine verstärkte Zusammenarbeit auch mit den Sozialwissenschaften auszumachen. In diesem Bestreben entstand in den USA im Spannungsfeld der Soziologie und der Künstlicher Intelligenz ein neues Forschungsfeld, „in welchem die Möglichkeiten eines wechselseitigen Konzepttransfers ausgelotet und für beide Disziplinen fruchtbar gemacht werden sollen“². In Deutschland etablierte sich dieser Arbeitsbereich 1996 durch den von Thomas Malsch geprägten deutschen Begriff Sozionik. Die Sozionik beschäftigt sich mit folgenden drei Aspekten des interdisziplinären Arbeitens:³

1. „Informatikreferenz“: Umsetzung sozialer Phänomene und soziologischer Konzeptionen in neuartige informatische Modelle und Konzepte. Soziologisches Wissen soll Lösungsansätze für Komplexitätsbewältigung, Robustheit, Adaptionodynamik und Selbstorganisation innerhalb von künstlichen Systemen liefern. Besonderer Bedeutung kommt dem soziologischen „Übergangsproblem“ (*micro-macro link*) bei der Konstruktion von „*large-scale open systems*“ zu.
2. „Soziologiereferenz“: Entwicklung künstlicher Sozialsysteme und soziologischer Simulationsprogramme (*social simulation*). Sie sollen zu einem besseren Verständnis von gesellschaftlichen Phänomenen (Dynamik sozialen Wandels, soziale Strukturbildung) und zu einer Theoriebewertung führen.
3. „Anwendungsreferenz“: Entwicklung praktischer Anwendungssysteme wie „soziale Agenten“, „intelligente Assistenten“ und „Multiagentensysteme“. Der Erfolg sozionischer Anwendungsprogramme ist wesentlich durch eine gelungene sozio-technische Systemgestaltung bedingt.

Diese Arbeit ist im Kontext des ersten Punkts angesiedelt. Sie setzt sich mit der Frage auseinander, wie die soziologische Konzeption der sozialen Norm in einem Multiagentensystem modelliert werden kann. Kapitel 1 führt aus, warum die Modellierung von sozialen Normen möglicherweise einen fruchtbaren Ansatz darstellt, um das verteilte Problemlösen durch eine Vielzahl von Agenten verlässlicher und weniger komplex zu gestalten. Kapitel 2 liefert, ausgehend vom Begriff der Institution, eine Erläuterung des soziologischen Normenkonzepts. In Kapitel 3 folgt eine Typisierung von Normen, wie verschieden geartete soziale Konstellationen sie entstehen lassen. Eine Modellierung des Normbegriffs nach [CC95] wird in Kapitel 4 vorgestellt. Kapitel ?? beschreibt zusammenfassend eine Bewertung dieser Modellierung.

²[Mal01], S. 5.

³vgl. [Mal97].

Kapitel 1

Kollektives Handeln in Form von Organisationen

Ein Multiagentensystem (MAS) setzt sich aus eine Vielzahl von autonomen Agenten zusammen. Die Mächtigkeit dieser Systeme resultiert aus ihrem verteilten Charakter, der die Möglichkeit in sich trägt, dass Aufgaben nicht wie in klassischen KI-Systemen von einer einzelnen Softwareinheit bewältigt werden müssen, sondern mehrere Agenten gemeinsam ein Problem lösen können. Dies hat den Vorteil, dass nicht alle Agenten im System über die gleichen Fähigkeiten und Wissensvolumina verfügen müssen, die zur erfolgreichen Bearbeitung einer Aufgabe nötig sind. Das Erlangen einer Lösung auf der Basis von erreichten Teillösungen bzgl. einer Sachlage durch verschieden Agenten hält auf der einen Seite die einzelnen Softwarekomponenten im Entwurf einfacher und erlaubt auf der anderen Seite die Bildung von Experten.

Das verteilte Problemlösen in einem MAS setzt eine funktionierende Interaktion der einzelnen Agenten voraus. Sie müssen sich koordinieren können und sollen kooperative Leistungen vollbringen. In Gruppen, die aus einer überschaubaren Anzahl von Teilnehmern bestehen, mögen Koordination und Kooperation abstimmbare sein. Zum einen, da eine Einigung unter wenigen Parteien nicht nur wahrscheinlich ist, sofern eine Interessenkonvergenz vorliegt, zum anderen, weil diese in annehmbarer Zeit zu erreichen ist. Bei einer deutlich höheren Anzahl potentieller Kooperationspartner sieht dies dagegen anders aus und erfordert einen immensen Kommunikationsaufwand.

Damit Koordinationen und Kooperationen zwischen autonomen Agenten nicht nur zustande kommen, sondern darüber hinaus auch Verlässlichkeit bieten, bedarf es einer gewissen Rahmung des jeweiligen Kontextes. In komplexen Gesellschaften leisten Organisationen diese Funktion. Wie auch immer man Organisationen definieren mag – betriebswissenschaftlich, organisationssoziologisch – zeichnen sie sich stets durch zwei zentrale Charakteristika aus: Organisationen orientieren sich an einem spezifisches Ziel (Organisationsziel) und weisen gleichzeitig ein bestimmtes Niveau formeller Festlegungen (Formalisierung) der Erwartungen und Handlungen der Mitglieder auf. In Form von Organisationen

wird also die kollektive Kooperation von Akteuren unter dem Dach auch formell festgelegter Regeln realisiert. Organisationen sichern nicht nur wiederkehrende Handlungsabläufe, sie tun dies auch bindend und bieten so den Akteuren ein Maß an Sicherheit, da ein regelabweichendes Verhalten vor dem Hintergrund der organisatorischen Verfassung Sanktionen nach sich zieht.

Die Verlässlichkeit der vereinbarten Koordination bzw. Kooperation macht auch in einem MAS eine wichtige Größe aus. In kleinen Systemen kann dem einseitigen Aussteigen bei einer Kooperation eventuell dadurch vorgebeugt werden, dass die Agenten auf der Basis eines „Gedächtnisses“ es in Zukunft vermeiden werden, sich mit dem entsprechenden Abweichler erneut einzulassen. So läge es im Interesse jedes Agenten, sich an Absprachen zu halten, um nicht in eine isolierte und somit benachteiligte Position zu geraten. Bei einem großen System, in dem der einzelne Agent an Anonymität gewinnt, greift dieses Vorgehen nicht. Um dennoch eine verlässliche Interaktion zu erlauben, bedarf es Verbindlichkeiten.

Innerhalb einer Organisation wird der normative Rahmen des Handelns durch Regeln gestellt, die in der Soziologie mit dem Begriff der Institutionen oder sozialen Normen umfasst werden. Normen bilden also die Grundlage für eine verlässliche Ordnung sozialer Prozesse und damit die Sicherheit gewisser reproduktiver Leistungen, die anders nicht denkbar wären, etwa wegen des zu hohen Aufwandes (s. o.) oder wegen des Problems der einseitigen Aufhebung von Einigungen. Die Kooperation und Koordination von Agenten in Form von Organisationen, also auf der Basis von sozialen Normen, die nicht alles zulassen, was seitens der Akteure motivational gewollt wird, scheinen somit für die aufgezeigten Probleme ein Lösungspotential in sich zu bergen.

In Organisationen kooperierende Agenten müssen sich normativ verhalten. Um eine Modellierung sozialer Normen für künstliche Agenten zu erlangen, bietet es sich an, auf die Konzeption der sozialen Norm, wie sie die Soziologie definiert, als erste Orientierung zurückzugreifen.

Kapitel 2

Institutionen und soziale Normen – eine soziologische Begriffsklärung

In der Soziologie herrscht keine eine einheitliche Auffassung vor, was unter dem Begriff der sozialen Normen zu verstehen ist. Die Forschung brachte vier verschiedene Konzeptionen hervor, die sich in ihren Annahmen und der jeweils gewählten Perspektive des wissenschaftlichen Ansatzes unterscheiden.¹ Unter ihnen wendet sich eine sogar vom Normbegriff ab.

1. *Der statistische Normenbegriff im behavioristischen Ansatz.*

In Orientierung am Behaviourismus sind wissenschaftliches Aussagen lediglich dann zulässig, wenn sie objektiv beobachtbar und messbar sind. Dieser Ansatz, eine Norm allein in ihrer Objektivität als Forschungsgegenstand zuzulassen, reduziert den Begriff der Norm auf ein beobachtbares gesellschaftliches Phänomen. In diesem Sinne emergiert eine Norm, wenn der Großteil der betrachteten Akteure sich gemäß einem Muster verhält. Aus der prinzipiell externen, nur die Außenseite des Handelns wahrnehmende Beobachterperspektive sind Intentionen und Interpretationen der Subjekte nicht existent. Diese Sicht macht jedes regelmäßige Verhalten zu einer Norm, wodurch ein begriffliche Unterscheidung zwischen regelmäßigem Verhalten und regelbefolgendem bzw. sanktionsvermeidendem Verhalten nicht möglich ist. Als ein Vertreter dieser Normkonzeption im behavioristischen Ansatz sei Theodor Geiger (1891–1952) genannt.

2. *Der „soziologische“ Normenbegriff in der Handlungstheorie.*

Emile Durkheim, Max Weber und Talcott Parsons präsentieren die „soziologische“ Konzeption von Normen, weichen in ihrer Herangehensweise jedoch voneinander ab. Durkheim (1858–1917) wählt die Forschungsperspektive des externen Beobachters, da er wie die Vertreter des behavioristischen Ansatzes der Auffassung ist, allein objektive Aussagen dürften den Anspruch auf Wissenschaftlichkeit erheben. Daher konzentriert sich

¹vgl. [Nun84].

Durkheim auf die Objektivität von Normen als Symbole, als soziale Fakten, die sich durch Sanktionen auszeichnen und durch diese erkennbar werden. Diese Fokussierung ist jedoch, im Gegensatz zu den Behavioristen, rein methodologischer Natur, da Durkheim die Existenz der Binnenperspektive des Handelnden nicht in Frage stellt, vielmehr annimmt, dass Normen und die an sie geknüpften Sanktionen aus einem inneren Gefühl der Verpflichtung resultieren. Da diese aber in ihrer Symbolwerdung Objektivität gewinnen, stellen sie allein den wissenschaftlich zu untersuchenden Gegenstand dar.

Weber (1864–1920) dagegen wählt die Perspektive des Subjekts als Ausgangspunkt und macht dadurch den Unterschied zwischen objektiv beobachtetem Verhalten und subjektiv intendiertem Verhalten explizit: Er unterscheidet Verhaltensregelmäßigkeiten, die auf Brauch bzw. Sitte basieren zum einen von denjenigen, die aus gleichgelagertem Interesse resultieren, und zum anderen von denjenigen, die sich aus einer Handlungsorientierung an der Vorstellung vom Bestehen einer legitimen Ordnung ergeben. Mit diesem Wechsel von der Verhaltenstheorie zur Handlungstheorie schafft Weber die Grundlage für eine soziologischen Normenkonzeption. Er räumt die Unmöglichkeit ein, wissenschaftlich verbindliche Aussagen über die Gültigkeit von Normen zu treffen, unterstreicht aber das Potential der Soziologie, deskriptiv das faktische Gelten - und damit die Tatsache, dass Subjekte ihr Handeln an der Vorstellung der Gültigkeit bestimmter Normen orientieren - festzustellen.

Für Parsons (1902–1979) sind Normen institutionalisierte Rollenerwartungen. Infolge von Integration der Rollenerwartungen in das kulturelle System wird ihre Objektivität gesichert, subjektive Verbindlichkeit kommt ihnen im Zuge von Internalisierung zu.

3. *Der ethische Normbegriff im rekonstruktivistischen Ansatz.*

Unter der Annahme, dass die Perzeption der Umwelt durch ein Subjekt das Ergebnis aktiver Strukturierungsleistungen ist (Jean Piaget (1886–1980)), orientiert sich die ethische Normkonzeption an den von Lawrence Kohlberg (1927–1987) initiierten Forschungen zur Entwicklung des moralischen Bewusstseins eines reflexionsfähigen Subjekts. Dieser Entwicklungsprozess, den ein Mensch im Laufe seines Lebens durchläuft, teilt sich in drei ausgezeichnete Stufen. Das frühe kindliche Bewusstsein ist zunächst von der präkonventionellen Moral geprägt. Es richtet sein Handeln ausschließlich an konkreten Verhaltenskonsequenzen aus, so dass eine Normbefolgung allein der Intention entspringt, negative Sanktionen zu vermeiden. Diese Phase entspricht der Normkonzeption im behavioristischen Ansatz, da gleichmaßen das Subjekt der Außenseite des Handelns verhaftet bleibt.

In der zweiten Phase, auf der Stufe der konventionellen Moral, wird das eigene Handeln den Erwartungen eines Bezugssystems – anfangs des direkten Umfelds, später der Gesellschaft – angepasst. Das konventionelle Moralniveau korrespondiert zu Parsons' Normbegriff. Durch eine enge Bindung an die anderen Subjekte internalisiert das Kind die Erwartungen

seiner Bezugsgruppe und befolgt Normen, weil diese in Form institutionalisierter Rollen einen Orientierungsrahmen bieten.

Auf der postkonventionellen Stufe richten sich die Handelnden nicht mehr allein an institutionalisierten Normen aus, sondern bedienen sich abstrakter Prinzipien, die einer Begründung von bestimmten Normen dienen. Das Vermögen, zwischen bloßer Konvention und universeller Gültigkeit unterscheiden zu können, entspricht der rekonstruktivistischen Sozialtheorie, da nun das Subjekt sowohl über einen soziologischen wie einen ethischen Normenbegriff verfügt.

4. *Ethnomethodologische Kritik, Aufhebung des Normbegriffs.*

Dieser Ansatz verwirft die Relevanz einer soziologischen Konzeption von Normen. Harold Garfinkel (*1917) bestreitet, dass es eine allgemeine, vom konkreten Kontext unabhängige Norm gibt. Erwartungen bzgl. des Verhaltens sind immer situationsspezifisch. Die subjektive Sicht von Weber betrachtet Handlungen jedoch als abgeschlossene Ereignisse. Dagegen wird aus einer wirklichen Teilnehmerperspektive Handlung im Vollzug erfahren, was dazu führt, dass an die Stelle von wissenschaftlichen Rationalitäten (zweckrationales Verhalten) Rationalitätskriterien des Alltags (Kontextabhängigkeiten) treten. Die Situationsbedeutung ist also nicht intersubjektiv oder kulturell vorgegeben, sondern Produkt von Strukturierungsleistungen der Teilnehmer. Quasinormative Regeln ergeben sich aus der wechselseitigen Erwartung, der andere nähme die Situation in gleicher Weise wahr, wäre er an der gleichen Stelle. Ein Abweichen wird unmittelbar sanktioniert.

Dennoch unterscheiden sich diese Regeln vom soziologischen Normkonzept, weil sie in den Handelnden tiefer verankert und insofern nicht hintergebar sind, da auf einer reflexiven Ebene kein Zugriff auf sie möglich ist. Dieser konstitutive Regelcharakter führt dazu, dass eine Regelverletzung nicht moralisch, sondern in klinischen Kategorien als eine Abweichung vom Normalen bewertet wird. Umgekehrt heißt dies, dass erst eine Regelbefolgung von Zurechnungsfähigkeit zeugt, die die soziologische Normkonzeption voraussetzt.

Das handlungstheoretische Normenkonzept wird im Folgenden vertiefend dargestellt. Ausgehend vom Begriff Institution soll ein Überblick über die mit ihm verbundenen Regulierungen von Handlungen gegeben werden. Im weiteren wird das Konzept der sozialen Norm beschrieben. Auf eine Ausführung über sozialen Rollen und soziale Drehbücher wird an dieser Stelle verzichtet. Dabei folgt die verkürzende Darstellung im wesentlichen [Ess00b].

2.1 Institutionen

Institutionen bilden einen zentralen Bestandteil gesellschaftlicher Ordnung. Dennoch ist der Begriff der Institution in der Soziologie nicht immer deutlich benannt. Geprägt wurde er zunächst vorwiegend von Vertretern des Struktur-Funktionalismus, insbesondere von Emile Durkheim, der in ihm eine passende

Wiedergabe dessen sah, was er als sozialen Tatbestand bezeichnete. Unter einem sozialen Tatbestand versteht Durkheim eine Form des sozialen Zwangs, der außerhalb der Individuen liegt, diese aber nötigt, sich ihm anzupassen. Dieser extern vorgegebene Verhaltensrahmen beruht auf sozialen Gewohnheiten, Überzeugungen und Gebräuchen und stellt jedem Akteur einen zu erwartenden Interaktionsraum gegenüber, dessen Einhaltung sozial verbindlich ist:

„Damit aber ein soziologischer Tatbestand vorliege, müssen mindestens einige Individuen ihre Tätigkeit vereinigt haben, und aus dieser Verbindung muss ein neues Produkt hervorgegangen sein. Und da diese Synthese außerhalb eines jeden von uns (weil zwischen einer Mehrheit von Psychen) stattfindet, so führt sie notwendig zu dem Ergebnis, außerhalb unseres Bewusstseins gewisse Arten des Handelns und gewisse Urteile auszulösen und zu fixieren, die von jedem Einzelwillen für sich genommen unabhängig sind. Es gibt ... ein Wort, das in geringer Erweiterung seiner gewöhnlichen Bedeutung diese ganz besondere Art des Seins ziemlich gut zum Ausdruck bringt, nämlich das Wort Institution. Tatsächlich kann man, ohne den Sinn dieses Ausdrucks zu entstellen, alle Glaubensvorstellungen und durch die Gesellschaft festgesetzten Verhaltensweisen Institutionen nennen;“²

Eine relevante neuere Erklärung erfuh der Begriff Institution durch die Arbeit von Peter L. Berger und Thomas Luckmann. Sie sehen in Institutionen Regeln, die dem Lösen von Problemen im Alltag dienen, indem sie festschreiben, welche Handlungen sinnvoll und möglich sind. Diese sich zu einem Modell verfestigten Handlungen und die darauf aufbauenden Institutionen werden durch die Menschen geschaffen und durch ihr Handeln im weiteren reproduziert. Auf diesem Wege erlangen sie dann jedoch eine objektive Wirklichkeit und damit Macht über die Akteure, der sich diese kaum zu entziehen in der Lage sind:

„..., daß die Beziehung zwischen dem Menschen als dem Hervorbringer und der gesellschaftlichen Welt als seiner Hervorbringung dialektisch ist und bleibt. Das bedeutet: der Mensch – freilich nicht isoliert, sondern inmitten seiner Kollektivgebilde – und seine gesellschaftliche Welt stehen miteinander in Wechselwirkung. Das Produkt wirkt zurück auf den Produzenten.“³

Der Mensch schafft also im Zuge von Interaktionen eine Welt, die er im weiteren nicht als menschliches Produkt, sondern als Wirklichkeit erfährt, und diese Objektivation infolge von Externalisierung begründet die Verbindlichkeit für das Handeln, die Institutionen innewohnt. Zusammenfassend lässt sich sagen, dass Institutionen in den Erwartungen der Akteure verankerte, sozial definierte Regeln mit gesellschaftlicher Geltung und daraus abgeleiteter unbedingter Verbindlichkeit für das Handeln sind. Sie gelten auch dann, wenn einzelne Akteure sie nicht kennen oder missachten.

²[Dur84], S. 99f.

³[BL04], S. 65.

Wichtig ist, dass Institutionen Regeln mit erwartetem Geltungsanspruch sind. Damit heben sie sich von anderen bloßen Regelmäßigkeiten des Handelns ab, die nicht als Erwartungen verankert oder gar durch Sanktionen erzwungen oder über bestimmte moralische Gefühle getragen sind. Soziale Regelmäßigkeiten bestehen, ohne dass damit auf diese Struktur hin ausgerichtete, spezifische Erwartungen oder Sanktionen verbunden wären. Sie entstehen, werden aber nicht als „Norm“ erwartet. In diesem Sinne sind rein auf Interessenkonvergenz beruhende Abstimmungen und freiwillige Kooperationen keine Institutionen. Denn dabei handelt es sich nur um Regelmäßigkeiten, die sich ergeben, ohne dass die Akteure genau diese Regelmäßigkeiten im Sinn haben. Die bloße Regelmäßigkeit kommt noch nicht der Regel gleich, d. h. dass bei gleichem Handeln sich wiederholende und bei zahlreichen Handelnden verbreitete Abläufe keineswegs einen Geltungsanspruch haben, der ja eine besondere Orientierung der Akteure begründet. Die Regelmäßigkeiten, die sich in der Wiederholung oder Verbreitung eines Tuns niederschlagen, entsprechen vielmehr einem Brauch oder einer Sitte.

2.1.1 Funktion von Institutionen

Die Entstehung von Institutionen basiert weder auf expliziter Planung, noch ist sie grundlos. Als Teile und Folgen menschlicher Versuche, grundlegende Lebensprobleme zu lösen, regeln sie das Handeln der Akteure und damit deren soziale Beziehung zueinander. Indem Institutionen den Beteiligten ein Modell liefern, wie diese sich selbst und die jeweils anderen sehen sollen, welches Denken, Handeln und Fühlen richtig ist, schaffen sie eine individuelle Orientierung und eine kollektive Ordnung.

Eine Situation des gegenseitigen Handelns wird durch Institutionen mit einer typischen Orientierung strukturiert, wodurch die Situation ihr rein strategisches Wesen verliert und zumindest eine partielle Parametrisierung⁴ aufweist. Verhalten sich die jeweiligen Akteure nach vorhersagbaren Regeln, verlieren psychische Motive ihre gänzliche Unberechenbarkeit und die Unwägbarkeiten sozialen Handelns werden einigermaßen kontrollierbar. Genau in dieser Parametrisierung einer sozialen Situation liegt die Funktion von Institutionen: Die hohe Situationskomplexität wird reduziert, der aus dem Problem der doppelten Kontingenz resultierende unendliche Regress erfährt einen Abbruch, und die Akteure des Umfeldes werden zu einer vorhersehbaren Natur. Diese Parametrisierung beruht auf der einigermaßen verpflichtenden Steuerung gesellschaftlichen Handelns durch Institutionen, deren Geltungsanspruch sich aus den zur Verfügung stehenden Durchsetzungsmechanismen ableitet. Daher steigt das Maß der Parametrisierung einer Situation mit der zunehmenden Schwere der Sanktionen, die bei einem abweichenden Verhalten drohen.

⁴Als parametrisiert werden Situationen bezeichnet, in denen ein Akteur sein Handeln nicht auf andere Akteure ausrichtet, die ihn ebenfalls als handelnden Akteur beachten.

2.1.2 Sanktionen und Legitimität

Es wurde deutlich, dass die Geltung der institutionellen Regeln auf den Konsequenzen beruht, zu denen eine Übertretung oder Befolgung führt. Institutionelle Regeln legen fest, was getan werden muss, welche Handlungen erlaubt oder verboten sind, und sie sind über Sanktionen abgesichert. Diese lassen sich in drei Dimensionen unterscheiden.

- *Positiv – Negativ.*

Positive Sanktionen stellen Belohnungen, negative Sanktionen Bestrafungen dar. Beide kommen zu den instrumentellen Folgen, also den in der Handlung liegenden Anreizen, bei einer Regelbefolgung oder einer Regelüberschreitung hinzu. Somit stellen sie eine eigene motivationale Kraft dar, sich konform einer Regel zu verhalten, auch wenn der Akteur von dieser nicht überzeugt ist.

- *Formel – Informell.*

Die Verhängung formeller Sanktionen erfolgt durch einen eigenen Kontrollapparat nach explizit formulierten Regeln auf der Basis eines formellen Verfahrens (z.B. Gerichtsverfahren). Informelle Sanktionen werden ohne eigenes Regelwerk verhängt. Sie äußern sich in Form von beiläufigem Lob, ablehnender Mimik, genereller sozialer Wertschätzung. Sie steuern das Alltagsleben des Akteurs, worin ihre große Macht liegt. Denn die soziale Anerkennung des direkten sozialen Umfeldes ist wesentlich für ein positives oder negatives Selbstbild des Akteurs verantwortlich.⁵

- *Extern – Intern.*

Externe Sanktionen sind die äußeren Reaktionen auf das Handeln. Es ist jedoch nicht anzunehmen, dass Menschen ihr Handeln allein an einer Kalkulation über die Vorteile einer positiven Sanktion oder die Nachteile einer negativen Sanktion ausrichten. Vielmehr kommt es im Akteur selbst ebenso zu internen Sanktionen, die sich z. B. in Form von einem guten oder schlechten Gewissen (Beichte) niederschlagen. Dies tritt auf, wenn der Akteur sich mit der entsprechenden Norm identifiziert, sie internalisiert hat. Normen dieser Art wirken auch dann regelnd, wenn keine externen Sanktionen zu fürchten sind.

Das Überschreiten einer institutionellen Regel kann auch unterbleiben, weil das aus einem Regelbruch resultierende Verhalten anderer für einen Akteur nicht wünschenswert ist. In diesem Fall handelt er nicht aus der Intention der Sanktionsvermeidung, sondern er fürchtet, dass die Reaktionen anderer Akteure eine Verletzung der eigenen Interessen bedeuten würde. Damit liegt es im unmittelbaren Interesse des Akteurs, den Regeln zu folgen, da ansonsten das ausgelöste Verhalten aller anderen ihm schaden würde. Dies ist z. B. bei dem Phänomen der indirekten Preisabsprache zu beobachten: Ein einseitiges Abweichen in Form einer Preisunterschreitung durch einen Anbieter kann zu einer allgemeiner Preissenkung führen, also zu einem abweichenden Verhalten einiger oder

⁵vgl. [Ess00b], S 112.

aller Anbieter. So ergibt sich eine Folgesituation, die für den initiierenden Abweichler schlechter ist, als hätte er die Verhaltensregularität eingehalten. Denn der Absatz pendelt sich auf das gleiche Maß ein, nun aber zu einem geringeren Stückpreis. In Situationen dieser Prägung stabilisieren sich Institutionen selbst (*self-enforcing*).

Sanktionen bilden die Basis für eine garantierte Geltung einer sozialen Ordnung. Eine Sanktion beschreibt die bewerteten Folgen, die sich aus der institutionellen Geltung einer Regel ergeben. Eine funktionierende Sanktionierung setzt eine funktionierende soziale Kontrolle sowie eine funktionierende Sozialisation voraus. Eine wirksame Kontrolle findet meist nur dann statt, wenn der Sanktionsaufwand gering ist oder die Beteiligten in einem engen Netzwerk stehen. Aus diesem Grund sind eigens eingerichtete, institutionell geregelte und organisierte Sanktionsapparate wichtig (Polizei, Gerichte).

Neben Sanktionen erfahren Institutionen ihre Geltung aus der Legitimität einer Ordnung. Auf die Möglichkeit, dass Akteure ihr Handeln an der Vorstellung vom Bestehen einer legitimen Ordnung orientieren, wies Max Weber hin. Diese Vorstellung legt den Sinngehalt des Handelns unter dieser Ordnung für den Akteur fest. Somit erwächst die Legitimität einer Institution einer subjektiven Einstellung der Akteure zu den Regeln, die sich zum einen aus deren Richtigkeit und zum anderen aus deren Gerechtigkeit ergibt. Richtigkeit meint, dass eine Regel als kognitiv zutreffend und sachgerecht eingeschätzt wird. Als gerecht wird eine Regel dann beurteilt, wenn sie als moralisch vertretbar und vor dem Hintergrund abstrakter Bewertungsprinzipien als angemessen angesehen wird. Die Legitimität von institutionellen Regeln wird also aus einem Gefühl der Richtigkeit und der Gerechtigkeit gespeist und stellt somit die Grundlage einer auch subjektiven Geltung der Institution als verbindlicher Handlungsrahmen in der entsprechenden Situation dar.

Die Relevanz und die hohe Wirksamkeit des Zusammenspiels von Sanktionen und Legitimität tritt insbesondere im Fall von gegensätzlichen Interessensausrichtungen der Beteiligten hervor. Die Interessen der Akteure haben u. U. eine unterstützende Wirkung auf institutionelle Regeln (s. o.), ihnen entspringt aber nicht zwangsläufig eine Geltung der Regeln:

„Institutionen mit ihrer ganz eigenen externen und internen Absicherung durch Sanktionen und durch die empfundene Legitimität ihrer Geltung sind dazu da, das soziale Handeln unter den Bedingungen auch einer antagonistischen Kooperation von den Unwägbarkeiten, Risiken und Dilemmata der bloßen Interessenlagen der Akteure zu befreien.“⁶

⁶[Ess00b], S. 9.

2.2 Soziale Normen

Soziale Normen, eine wichtige Sonderform einer Institution, stehen Institutionen begrifflich so nah, dass sie vielfach in der Soziologie in exakt der gleichen Bedeutung verwendet werden. Auch bei Normen handelt es sich gleichermaßen um Regeln mit Geltungsanspruch. Um eine sinnvolle begriffliche Abgrenzung dennoch nicht aufzugeben, kann man Normen im Sinne von Popitz verstehen, der unter einer sozialen Norm eine institutionelle Regel fasst, die zumindest mit negativen Sanktionen im Fall ihrer Überschreitung bedacht ist.⁷

In vielen Fällen zieht ein normkonformes Handeln eine positiv bewertete Reaktion nach sich. Damit heben sich Normen von anderen institutionellen Regeln ab, die als bloße Vorschriften wahrgenommen werden können, während befolgendes oder abweichendes Verhalten bzgl. einer Norm positive oder negative Bewertungen im Akteur und in anderen Personen assoziiert. Soziale Normen enthalten Soll-Aussagen, von denen erwartet wird, dass sie unbedingt und erfolgsunabhängig befolgt werden („Du sollst nicht töten!“ oder „Du sollst an einer roten Ampel warten!“). Normen können also in Bezug auf ihre Fokalhandlung⁸ als Vorschrift bzw. Gebot oder als Verbot formuliert sein. Normen werden nicht in Hinblick auf resultierende Zustände und ein möglicherweise damit einhergehender äußerer Gewinn befolgt. Sie stellen selbst bereits das Ziel des Handelns dar.

Das Konzept der sozialen Norm zeichnet sich also durch drei Aspekte aus und hebt sich dadurch von dem der institutionellen Regel ab: die Unbedingtheit der Normen, die Erfolgsunabhängigkeit des dadurch gesteuerten Handelns und ihre Fundierung in Sanktionen. Die Relevanz dieser drei Punkte beruht auf der Vorstellung ihrer Notwendigkeit, um soziale Ordnung zu erlangen, also das Unterbleiben von normabweichendem Verhalten zu garantieren. Die rein kognitiven Erwartungen und Interessen der Akteure würden für eine verlässliche Regelung des Handelns keine ausreichende Grundlage bieten.

Normen müssen gesetzt und durchgesetzt werden. Vielfach wird dies von Organisationen oder kooperativen Akteuren besorgt, die genügend Interesse und Kontrolle aufbringen, um neue Normen auch gegen andere Interessen zur Geltung zu bringen. Popitz bezeichnet die durchsetzenden Akteure als Normsetzer; die Normsender sind die Akteure, die durch ihr bewusstes oder unbewusstes Handeln ein Unterstützen der Norm zeigen und damit zum Ausdruck bringen, dass sie gilt. Die Normhüter überwachen durch soziale Kontrolle das Einhalten der Norm.

Auf zwei weitere ausgezeichnete Personen oder Personengruppen im Kontext von sozialen Normen sei hingewiesen: Die Adressaten, von denen die Fokalhandlung verlangt wird, stehen den Benefiziaren oder Nutznießern gegenüber, die

⁷vgl. [Pop80].

⁸Die mit einer Norm durch Anspruch und Erwartung verbunden Handlung wird als Fokalhandlung der betreffenden Norm bezeichnet.

von einem normkonformen Handeln profitieren (z. B. Autofahrer, die an einer roten Ampel stoppen gegenüber Fußgängern). Sind diese Gruppen tatsächlich als getrennt voneinander auszumachen, spricht man von einer disjunkten Norm. Beide Gruppen können allerdings auch zusammenfallen wie im Fall des Rechtsfahrgebots, das der Kategorie der konjunkten Normen entstammt. Liegt der Vorteil ausschließlich bei einer Partei, können Konflikte über den Sinn, die Legitimität und die Geltung der Norm entstehen (z. B. Tempozone 30).

Jeder Regelbefolgung und damit jedem sinnhaften Handeln geht eine hermeneutische Leistung voraus, auch wenn das Handeln halb bewusst, automatisch oder gewohnheitsmäßig erfolgt. Eine Situation muss vom Akteur zunächst verstanden werden, bevor es zu einer adäquaten Handlung kommen kann. Das Erkennen einer Situation stützt sich auf Symbole, die allerdings keine vollständige Beschreibung der Situation liefern. Die Bruchstücke müssen vom Akteur interpretiert werden, um der Ausrichtung ihres Handelns eine sinnvolle Basis zu geben. Somit müssen Normen und institutionelle Regeln nicht nur gewusst, sondern sie müssen in Situationen in ihrer Geltung auch richtig identifiziert werden.

Kapitel 3

Soziale Situationen des Normbedarfs

Normen sollen ein Orientierungsmodell für ein adäquates Handeln in einer bestimmten Situation liefern. Dass für die zwischenmenschliche Interaktion eine enorme Vielfalt an Konstellationen entsteht, liegt auf der Hand: Mit steigender Anzahl von Akteuren, die jeweils über ein großes Repertoire an Handlungsmöglichkeiten verfügen, wächst die Menge der möglichen Konstellationen explosionsartig an. Es hat sich aber gezeigt, dass sich eine Reihe der auftretenden Akteurskonstellationen auf einige wenige zurückführen lassen. Drei Grundmuster solcher strategischer Situationen sind soziologisch besonders wichtig: Situationen, in denen das Tun der Akteure koordiniert werden muss, soziale Dilemmasituationen und Konflikte.

Edna Ullmann-Margalit untersuchte hinsichtlich dieser drei sozialen Situationen die strukturelle Entstehung von Normen.¹ Um dem unendlichen Regress zu entkommen, der dem Versuch entspringt, normkonformes Verhalten durch Sozialisation zu erklären, war ihre leitende Frage weniger, unter welchen Bedingungen sich Normen bilden, sondern warum gewissen Typen von Normen existieren. Ihre funktionale Erklärung der Normentstehung geht von der Annahme aus, dass unterschiedliche Normtypen Lösungen für verschiedene Probleme darstellen, die sich aus bestimmten Interaktionssituationen ergeben. Dabei handelt es sich um strukturelle Probleme im spieltheoretischen Sinne. Die Lösung eines Problems heißt in diesem Zusammenhang nicht, dass die aus der Errichtung von Normen resultierende Situation alle Beteiligten befriedigt. Sie kann - wie im Fall des Konflikts - eine Lösung für nur einige bedeuten, wenn diese in der Lage sind, sich gegenüber anderen durchzusetzen.

Alle drei erwähnten Grundmuster strategischer Situationen erfordern für ihre Lösung eine normative Regelung der Handlungen der Akteure. Dabei unterscheiden sich die jeweils nötigen Normen zu einen vor dem Hintergrund der Beziehungen zwischen Adressaten und Benefiziaren nach dem Grad der Interessengegensätze beider Gruppen, zum anderen durch die Schwierigkeiten und

¹vgl. [Ull77].

Unwahrscheinlichkeiten ihrer Einrichtung. Stehen Adressaten und Benefiziarer immer weniger in einer konjunkten Beziehung, d. h. liegt der Nutzensgewinn durch das Einhalten einer Norm in zunehmendem Maße nur bei der einen Gruppe, wächst der Bedarf nach einer externen, nicht allein auf die Interessen oder die internalisierte Moral der Akteure gegründete Geltung und Garantie der Handlungsregulationen. In gleichem Maße müssen zur ihrer Einrichtung die gesellschaftlichen Mechanismen zunehmend zwingend wirken. Entlang diesen Gradverschiebungen lassen sich soziale Normen unterteilen in konventionelle, essentielle und repressive Normen, die sich den Problemkonstellationen der Koordination, der Dilemmasituation und des Konfliktes zuordnen lassen.

Im folgenden werden die Situationen und ihr normativer Regelungsbedarf im einzelnen erläutert. Die Darstellung orientiert sich an [Ess00a] und [Ess00b].

3.1 Das Problem der Koordination

Das Problem der Koordination tritt in Situationen auf, in denen mehrere Akteure ihr Handeln abstimmen müssen, da sie nur so eine Folgesituation erreichen können, die für alle Beteiligten wünschenswert ist. Dabei verfügt jeder Akteur über mehrere mögliche Handlungsvarianten. Doch nur gewisse Kombinationen von Handlungen aller führen zum gemeinsamen Erfolg, der die Voraussetzung des individuellen Erfolgs darstellt. Jeder Akteur kann jedoch vor dem Problem stehen, eine Handlungsauswahl für sich treffen zu müssen, ohne die Entscheidungen der übrigen Beteiligten zu kennen.

Koordinationsituationen unterscheiden sich nach dem Maß des benötigten Regelungsbedarfs. Die Situation zeichnet sich dadurch aus, dass alle Teilnehmer ein gemeinsames Interesse an dem Entstehen einer Koordination haben, das individuelle und kollektive Interesse liegen beieinander. Favorisieren dabei alle Beteiligten ein und die gleiche Kombination aus Einzelhandlungen und wissen dies voneinander, kommt es unmittelbar zu einer Koordination. In diesem Fall ergibt sich die Lösung strukturell aus der vorliegenden Interessenkonvergenz der Akteure.

Wenn dagegen verschiedene Akteure verschiedene Handlungskombinationen präferieren, kann eventuell eine kulturelle Lösung gefunden werden. Diese Möglichkeit ergibt sich, wenn sich eine der Handlungskombinationen durch ein besonderes Merkmal unverwechselbar auszeichnet, dadurch den Akteuren einen Rahmen für eine Koorientierung liefert und auf diesem Wege den unendlichen Regress der wechselseitigen Vermutungen prägnant beendet. Eindeutige Zeichen dieser Art werden als Fokalfpunkte bezeichnet. Treffen z. B. zwei Menschen eine Terminabsprache, ohne das explizit ein Treffpunkt ausgehandelt wurde – vielleicht sogar, weil beide am entsprechenden Ort noch nicht gewesen sind – kann sich ein Punkt dadurch auszeichnen, das er erhöht und gut sichtbar liegt. Fokalfpunkte können sich auch in Form von gemeinsamen Erinnerungen oder wechselseitig gewussten Vorlieben niederschlagen.

Es ist zu beobachten, dass beim Vorliegen von Fokelpunkten die entsprechende Kombination von Einzelhandlungen als Lösung gefunden wird und kein Akteur das Bestreben hat, von ihr wieder abzuweichen, selbst wenn der einzelne einer anderen Kombination gegenüber eine Präferenz hat. Denn für jeden ist die gelungene Koordination - auch unter nicht optimalen Bedingungen für den einzelnen - noch immer besser als ihr Scheitern, und zwar um so mehr, je höher der Gewinn in der Koordination gegenüber einer Fehlkoordination liegt.

Sind allerdings bei herrschender Interessendivergenz keine Fokelpunkte vorhanden, haben die Akteure keinen Anhaltspunkt, für welche individuelle Handlung sie sich entscheiden sollen, um einen passenden Teil zu einer zulässigen Handlungskombination beizutragen. Um in Situationen dieser Art schnell und kostengünstig zu einer Koordination zu kommen, werden Regeln für die Orientierung nötig. Das Bedürfnis der Akteure nach einer verlässlichen Regelung kann auch unter der Existenz von Fokelpunkten wachsen, wenn diese nur eine schwache Orientierung bieten - somit die Komplexität der Situation nicht ausreichend reduzieren - und gleichzeitig das Interesse jedes Akteurs an einer gelungenen Koordination sehr hoch bzw. ein Scheitern ganz und gar unerwünscht ist. Die wechselseitigen Vermutungen über die Salienz einer Option (eine Alternative drängt sich den Akteuren aus strukturellen oder kulturellen Gründen mit kaum bezwingbarer Macht auf) stellen als Lösungsweg ein riskantes Fundament dar, wenn es um viel geht. Die Art der Regelung, die die Abstimmung erleichtern soll, kann in Form von Konventionen oder Dekreten geschehen. Konventionen sind einmal akzeptierte und aktuell etablierte, durch Zeichen und Symbole markierte Lösungen von wiederkehrenden, in der Vergangenheit bereits bewältigten Koordinationsproblemen (z. B. Verkehrsschild als technische Umsetzung einer konventioneller Regelung des Straßenverkehrs). Dekrete sind Lösungen für neu auftretende Probleme der Koordination (z. B. das Eingreifen von Polizisten, wenn die Technik der Konventionen versagt).

Regeln, die zur Lösung eines Koordinationsproblems ausreichen, heißen konventionelle Normen. Dieser Begriff drückt aus, dass das Problem lediglich in der Abstimmung der Akteure liegt. Wegen der Übereinstimmung der Interessen bedarf die Einführung der konventionellen Normen keines Zwanges und im Grunde keiner weiteren Sanktionen. Sie ist selbststabilisierend, da sie alle Beteiligten zu Benefiziaren macht. Herrscht allerdings eine unterschiedliche Auffassung der Beteiligten darüber, wie wünschenswert konkret die gefundene Regelung gegenüber anderen ist, können positive Sanktionen oder moralische Prämien für das Einhalten der Regel helfen. Dies kann die eigentlich moralisch neutralen Regeln - sie zeigen nicht richtiges oder falsches Handeln in einem bloß kognitiven Sinne an und halten eine Vereinbarung fest, die auch anders hätte ausfallen können - mit Elementen moralischen Anspruchs behaften.

	B		
	K	D	
A	K	1, 1	-2, 2
	D	2, -2,	-1, -1

Abbildung 3.1: Auszahlungsmatrix des Gefangenendilemmas.

3.2 Soziale Dilemmasituationen

Der Grundtyp dieser sozialen Situation ist das aus der Spieltheorie bekannte Gefangenendilemma. Die Problematik seiner Struktur sowie mögliche Auswege aus ihr werden in Anlehnung an Ullmann-Margalit anhand des *Mortarmen's Dilemma* aufgezeigt.²

Das Szenario gestaltet sich wie folgt: Zwei mit Mörsern ausgestattete Soldaten liegen jeweils auf einem vorgelagerten Posten in Erwartung eines feindlichen Angriffs. Für beide ergeben sich die Handlungsmöglichkeiten, entweder auf dem Posten zu bleiben und zu kämpfen, oder zu flüchten. Je nach gewählter Alternative beider Soldaten ergeben sich drei unterschiedliche Folgesituationen.

1. Wenn beide Soldaten die Stellung halten und kämpfen, kann der feindliche Angriff abgewehrt werden.
2. Entscheiden sich beide für ein Desertieren, können die feindlichen Truppen um so schneller vorrücken, was es ihnen ermöglicht, die Soldaten in Gefangenschaft zu nehmen.
3. Bleibt ein Soldat auf seinem Posten, während der andere seinen verlässt, erschwert der Kämpfende den feindlichen Vorstoß. Dadurch gelingt dem Deserteur die tatsächliche Flucht, der Zurückbleibende wird dagegen im Gefecht umkommen.

Abbildung 3.1 gibt die möglichen Auszahlungen für Soldat A und B im Falle einseitiger bzw. beidseitiger Kooperation (K), also einem Verbleiben auf dem Posten, bzw. Defektion (D), einem Nichtkooperieren durch Flucht, wieder.

Da beide Soldaten sich ihrer Handlungsalternativen sowie der jeweiligen Konsequenzen bewusst sind, gleichzeitig aber keine Kenntnis über die Wahl des jeweils anderen haben, werden sich beide für die Flucht entscheiden.

Das Dilemma der Situation besteht darin, dass beide Akteure etwas davon hätten, wenn sie kooperierten. Es gibt durchaus eine Gemeinsamkeit der Interessen, die jedoch von der Versuchung bzw. der Defektion überlagert wird. Somit ist der Ausgang der Situation, der durch das eindeutige Nash-Gleichgewicht³

²vgl. [Ull77], S. 30.

³Eine Strategiekombination heißt Nash-Gleichgewicht, wenn für jeden Spieler gilt: Keine alternative Strategie sichert ihm einen höheren Nutzen, wenn alle anderen Spieler ihre Strategie ebenfalls nach dieser Regel wählen.

gegeben ist (in Abbildung 3.1 hervorgehoben), nicht pareto-optimal⁴. Für die Beteiligten gibt es einen besseren Folgezustand als den gleichgewichtigen, dieser kann aber aus individualistischen Gründen nicht erreicht werden. Denn das Desertieren ist immer attraktiver als ein Bleiben unter der Annahme, der andere habe möglicherweise seinen Posten bereits verlassen. Um eine Pareto-Verbesserung zu erzielen, muss dafür gesorgt werden, dass die Defektion nicht mehr die dominante Strategie⁵ für die Akteure darstellt, um den Akteuren die Chance zu geben, zu einer Kooperation zu finden. Noch besser wäre es natürlich, wenn trotz der Unsicherheiten und Versuchungen das Kooperieren selbst zur dominanten Strategie beider Akteure wird.

Gegenüber der Koordinationssituation fallen in der Dilemmasituation die individuellen und kollektiven Interessen auseinander. Die Akteure finden wegen der Dominanz der Defektion nicht zueinander, obwohl sie sich - und darin liegt gerade das Dilemma - durch eine Kooperation beide besser stellen könnten. Die Anreizstrukturen der Dilemmasituation müssen also so geändert werden, dass die defensive Furcht überwunden und - insbesondere - die offensive Versuchung ausgeschaltet werden kann. Eine Lösung der Situation muss diesen Widerspruch überwinden. Es muss etwas gegen die opportunistische Versuchung sowie gegen die defensive Furcht der Akteure getan werden. Durch eine einfache Verständigung, wie sie durch konventionelle Normen geleistet wird, ist dies grundsätzlich nicht möglich. Die hier nötigen Regeln müssen es schaffen, eine Bindung an die Kooperation als Handlungsstrategie herzustellen. Hierfür sind verschiedene soziale Mechanismen geeignet. Davon sind die wichtigsten die Verhängung von negativen Sanktionen bei Defektion und die Internalisierung von unbedingten moralischen Prämien für Kooperation. Regeln, die das Lösen einer Dilemmasituation leisten, werden mit dem Begriff der essentiellen Norm umfasst.

Im beschriebenen Beispiel der Soldaten wären drei essentielle Normen denkbar, die sich entweder auf reine Zwangsmaßnahmen, auf die Durchsetzung von Disziplin oder auf die Einführung von Ehre stützen. In den ersten beiden Fällen beruht die Regelung auf dem Verhängen negativer Sanktionen bei Nicht-Kooperation. Bei der Ehre tritt in Form einer internalisierten Moralauffassung eine Belohnung für die Kooperation hinzu.

- *Zwang.*

Eine Kooperation kann dadurch erreicht werden, dass die Defektion ausgeschlossen oder unattraktiv wird. Die Soldaten könnten im angeführten Beispiel an ihren Posten festgekettet oder ein Fliehen durch Verminung der jeweiligen unmittelbaren Umgebung unmöglich gemacht werden. Da Zwangsmaßnahmen dieser Art nicht immer wünschenswert oder praktika-

⁴Das Ergebnis eines Spieles wird dann als pareto-optimal bezeichnet, wenn es keine andere Konstellation gibt, bei der sich ein Akteur verbessern könnte, ohne einen anderen gleichzeitig zu verschlechtern.

⁵Eine dominante Strategie ist diejenige Strategie, die individuelle für den Akteur am besten ist, unabhängig davon, was die jeweils anderen Akteure tun.

bel sind, stellt dies nur eine eingeschränkte Lösung der Dilemmasituation dar.

- *Disziplin: Militärische Norm, die Desertion verbietet.*

Um dafür zu sorgen, dass die Defektion nicht mehr die dominante Strategie ist, könnte man Disziplin durchsetzen durch Androhung scharfer disziplinarischer Maßnahmen bei Gehorsamsverweigerung: Im Falle der Defektion droht die Todesstrafe wegen Verrates. Sie droht aber eben nur, während der Tod im alleinigen Kampf gegen den Gegner sicher ist. Aus diesem Grund stellt die Kooperation, das gemeinsame Kämpfen, nach wie vor nicht die dominante Strategie dar, sie ist im Gegensatz zur ursprünglichen Situation aber immerhin erreichbar. Denn das Desertieren ist noch immer attraktiver als das Bleiben, wenn der andere seinen Posten verlässt. Wenn der andere aber bleibt, wird auch Soldat A bleiben, wenn die Wahrscheinlichkeit, als Deserteur geschnappt zu werden, ausreichend hoch ist.

- *Ehre.*

Um die Kooperation nicht nur nicht unvermeidlich zu machen, sondern zur dominanten Strategie zu erheben, kann im Kontext des Beispiels auf die Ehre zurückgegriffen werden. Beim Kampf blüht dem Soldaten ein immaterielle Prämie in Form von Ruhm, bei der Flucht dagegen der Ruf als Verräter als Bestrafung. Durch die Ehre wird somit die Desertion deutlich negativer, da man zwar noch lebt, jedoch mit dem Makel des Verräters oder des Feiglings. Für Akteure, die sich mit dem Ehrbegriff identifizieren, ein ehrenhaftes Handeln somit als gutes Handeln internalisiert haben, bleibt der Kampf also als einzig annehmbare Handlungsmöglichkeit. Dies geschieht insbesondere auch dadurch, dass der alleinige Kampf, der in den vorherigen Konstellation für den Akteur den schlimmsten Ausgang bedeutete, nun durch den Heldentod Aufwertung erfährt.

		B	
		K	D
A	K	1, 1	-2, -2
	D	-2, -2,	-2, -2

		B	
		K	D
A	K	1, 1	-2, 0
	D	0, -2,	-1, -1

Abbildung 3.2: Die Auszahlungen bei der Durchsetzung von Zwang (links) und Disziplin (rechts).

Die sich aus den Handlungsregulationen in Form von Zwang und disziplinarischen Maßnahmen ergebenden Auszahlungen zeigt Abbildung 3.2. Die Auswirkungen von der Einführung von Ehre ist in Abbildung 3.3 zu sehen. Die jeweiligen Gleichgewichte sind hervorgehoben.

Offensichtlich ist die Ehre als Lösung sicherer und zielgerichteter als die Disziplin. Sie sorgt für den gleichen Ausgang der Situation wie die Disziplin, erreicht dies aber auf der Basis von Freiwilligkeit oder sogar Begeisterung (Lob

	B		
		K	D
A	K	1, 1	-1, 0
	D	0, -1	-2, -2

Abbildung 3.3: Erreichung eines pareto-optimalen Gleichgewichts auf der Basis von Ehre.

wirkt stärker als Tadel). Sie überführt das *Mortarmen's Dilemma* in ein pareto-optimales Gleichgewicht, ohne auf Zwangsmaßnahmen zurückgreifen zu müssen.

3.3 Konflikt

In den beiden vorangegangenen Situationen konnten Akteure im Prinzip etwas gemeinsam gewinnen, auch wenn der Weg dorthin nicht immer leicht zu erreichen war. Beim Konflikt dagegen gibt es keine kooperative Lösung. Konfliktsituationen sind durch das komplette Auseinanderfallen der Interessen gekennzeichnet. Der Gewinn des einen bedeutet immer den Verlust des anderen. Dies ist beispielsweise in allen Wettbewerben der Fall, aus denen nur einer als Sieger herausgehen kann. Die extremste Form eines Konflikts ist der Krieg. Das Problem liegt darin, dass es immer nur um die Umverteilung eines in der Summe konstanten Betrages geht, z.B. die Macht. Da jeder rationalistische Egoist seinen Nutzen möglichst mehren will, muss sich sein Handeln darauf richten, eine Situation anzustreben, die den anderen in dem Maße benachteiligt, wie er selbst einen Vorteil zieht. Somit ist jede Form von Kooperation oder Einigung - technisch und moralisch - ausgeschlossen, denn alles, was für Alter Nutzen bedeutet, heißt Schaden für Ego.

Soziale Situationen lassen sich unterteilen in reine und partielle Konflikte. Bei Kriegen oder Duellen handelt es sich um reine Konflikte: Der Gewinn des einen zieht notwendigerweise einen Verlust des anderen in der gleichen Höhe nach sich. Nach einem von einer Partei gewonnenen Krieg und der damit einhergehenden Durchsetzung einer Verfassung ist der resultierende Vorteil symmetrisch zum entstehenden Nachteil. Bei partiellen Konflikten dagegen geht es nicht allein um Gewinn oder Verlust, sondern um verschiedene Grade von Gewinn. Einen Verlust gibt es jedoch nicht. Diese Situation entspricht dem Problem der Koordination bei Interessendivergenz (Beispiel: gemeinsamer Kinobesuch mit unterschiedlichen Präferenzen für die Filme) und ist daher vergleichsweise leicht zu lösen. Die benachteiligte Partei hat ebenfalls einen Nutzen, auch wenn ein latentes Interesse an einer Änderung der Verhältnisse besteht.

Der reine Konflikt bietet keinen so leichten Ausweg. Und da seine antagonistische Struktur keine kooperatives Element enthält – im Gegensatz zu Dilemmasituationen – kann das Streben nach dem individuell besten Ausgang zu keinem Ende kommen. Selbst wenn ein Status quo erreicht wird, besteht das

Problem in seiner Erhaltung durch die jeweils bevorzugte Partei, so dass es zu keiner Stabilisation kommt. Somit reichen in diesen Situationen konventionelle oder repressive Normen nicht aus, vielmehr sind repressive Maßnahmen nötig, die über repressive Normen realisiert werden müssen. Diese Normen sind nicht an sich verständlich oder einzusehen und erfordern immer externe Mechanismen zu ihrer Durchsetzung, was sich letztlich nur in einer Form von Herrschaft, also durch einen Erzwingungsstab unterstützte Macht, erreichen lässt. Der Begriff repressive Norm sagt bereits aus, dass eine Stabilisierung der Situation nur durch mehr oder weniger massive Unterdrückung der einen Partei zu erlangen ist. Repressiv bleiben solche Lösungen auch dann, wenn die Akteure Sonderprämien für die Hinnahme des Status quo erhalten (z. B. Ausgleichszahlungen).

3.4 Gegenüberstellung der Normtypen

Die Unterteilung in konventionelle, essentielle und repressive Normen folgt aus der Struktur des jeweiligen Ordnungsproblems. Entsprechend der Reihenfolge ihrer obigen Aufführung wächst der Bedarf nach einer externen, nicht allein auf die Interessen oder die internalisierte Moral der Akteure gegründete Geltung und Garantie. Das liegt daran, dass Normadressat und Normnutznießer immer weniger in einer konjunkten Beziehung stehen. Entsprechend unterschiedlich milde oder mit Zwang durchsetzt sind die sozialen Mechanismen der Ordnung des Handelns. Symbole reichen für die Koordination, Dilemmasituationen benötigen die Unbedingtheit einer Moral und Konflikte sind nur über eine im Prinzip auch repressive Herrschaft zu kontrollieren.

	Typ der sozialen Norm		
	konventionell	essentiell	repressiv
Problemstruktur	Koordination	Dilemma	Konflikt
Verhältnis Adressat und Nutznießer	konjunkt	konjunkt/disjunkt	Konflikt
sozialer Mechanismus	Symbol	Moral	Herrschaft
Grundlage	Interesse/ Praxis	Zukunft/ Dependenzen	Staat/Recht
Garantie	gemeinsames Interesse	Sanktion: intern/informell	Sanktion: extern/formell
sozialer Prozess	Verständigung	Sozialisation	soziale Kontrolle

Abbildung 3.4: Normtypen und Geltungsmechanismen, nach [Ess00b].

Gleichermaßen wird die Grundlage der Ordnung zunehmend extern. Auf der Basis von Interessenkonvergenz kommt es zu einer zwanglosen Koordinierung. Die Beteiligten sind selbst in der Lage, diese Ordnung durch Kommunikation oder eingespielte Praxis zu erzielen, ohne dass es besonderer Sanktionen

bedarf. In der Dilemmasituation ist eine internalisierte Moral nötig, da diese die Akteure veranlasst, die Situation verlässlich zu überwinden. Hier reicht ein momentanes Interesse als Grundlage nicht aus. Das Verhalten der Beteiligten muss durch ihre Abhängigkeit am Zustandekommen der Ordnung bestehen und die Akteure müssen ihr Handeln mit einem Blick in die Zukunft festlegen. Da es aber im Interesse der Akteure liegt, die Dilemmasituation zu überwinden, reichen interne und informelle Sanktionen sowie im Zuge von Sozialisation internalisierte Normen. Die Ordnung eines Konflikts kann dagegen nur durch externe und formelle Sanktionen vor dem Hintergrund einer wirksamen sozialen Kontrolle erfolgen. Die nötigen gesellschaftlichen Mechanismen zur Durchsetzung repressiver Normen auf der Basis von Herrschaft sind das Recht und der Staat. Im Grunde entsteht nur in diesem Fall das Problem der Legitimität der resultierenden Ordnung. Denn in den obigen Situationen erfahren die normativen Ordnungen ihre Legitimität bereits aus dem Interesse der Akteure an ihrem Bestehen.

Abbildung 3.4 zeigt eine Übersicht der Zusammenhänge von Problemstruktur, benötigtem Normtypen und der Mechanismen zur Durchsetzung einer garantierten Ordnung.

Kapitel 4

Modellierung sozialer Normen in einem MAS

In Kapitel 1 wurde dargestellt, dass die Koordination und die Kooperation in einem MAS erleichtert werden könnten, wenn das Verhalten der einzelnen Agenten normativ ist. Nachdem die soziologische Konzeption der sozialen Norm umrissen wurde, werden im folgenden mögliche Modellierungen von Normen für Agenten vorgestellt. Dabei stützt sich die Darstellung auf [Saa01]. Zunächst werden drei Normkonzepte gegenüber gestellt, von denen das dritte, die Modellierung, wie sie Rosaria Conte und Christiano Castelfranchi vornehmen, im weiteren Verlauf dieses Kapitels genauer erläutert wird.

4.1 Mögliche Modellierungskonzepte

Um ein normatives Verhalten von Agenten umzusetzen, können verschiedene Modellierungsansätze gewählt werden. Aufgrund des zwingenden Charakters von Normen liegt es nahe, Normen auf der Basis von *constraints* zu realisieren. Dieser Ansatz würde zu einem gewünschten Systemverhalten führen, der Begriff des normativen Verhaltens wäre jedoch – entgegen der sozionischen Intention – rein metaphorischer Natur. Denn eine Architektur auf der Grundlage von *constraints* setzt ein normkonformes, allerdings kein normgeleitetes Verhalten um. Erfährt der Handlungsspielraum von Agenten eine automatische Einschränkung, findet also eine Restriktion bzgl. möglicher Aktionen an Stelle einer Selektion statt, resultiert zwar das intendierte Systemverhalten, dies geschieht jedoch nicht auf dem Weg der sozialen Kontrolle, deren Notwendigkeit durch den Ansatz selbst aufgehoben wird. Stete Normkonformität entspricht nicht der Tatsache, dass Normen zwar ein bestimmtes Verhalten nahe legen, dies auch unbedingt, aber dennoch ein Abweichen ermöglichen. *Constraints* erlauben zusätzlich keinen Lernprozess. Neue „Normen“ können nur im *off-line*-Zustand dem System beigefügt werden. Wünschenswert ist dagegen ein möglicher Wandel von Normen zur Laufzeit, die Emergenz von Normen. Ein weiterer Nachteil dieser Modellierung liegt darin, dass sie ausschließlich die Umsetzung von Verboten erlaubt. Eine Formulierung von Geboten ist nicht möglich.

Ein verbreiteter Ansatz zum Entwurf einer Architektur für deliberative Agenten folgt dem BDI-Modell. Nach ihm verfügt ein Agent über ein bestimmtes Wissen bzw. bestimmte Überzeugungen (*beliefs*), über persönliche Wünsche (*desires*) und Intentionen (*intentions*). Die Interaktion dieser drei Komponenten generiert infolge von Perzeption der Umwelt die Ziele des Agenten. Setzt man an diesem Punkt die Normmodellierung an und entwirft Normen als mögliche Ziele des Agenten, erhält man im Gegensatz zur ersten Konzeption ein autonomes Verhalten. Doch wenn diese sich gegenüber anderen Zielen nicht auszeichnen, bleibt ein zuverlässiges Verhalten allein ein Resultat eines vom Agenten ebenfalls gewünschten Verhaltens. Ein normenkonformes Verhalten ist nicht garantiert, es ereignet sich zufällig bzw. situativ bedingt durch die persönlichen Umstände eines Agenten. Insbesondere geschieht dies nicht infolge eines normativen Schließen bzgl. der Handlungsentscheidung, sondern wie bei jedem anderen Ziel auf der Nutzenabschätzung.

Ausgehend von Unzulänglichkeiten beider obiger Entwürfe liegt für Conte und Castelfranchi der Schlüssel für eine adäquate Konzeption von Normen in ihrer Modellierung als gesonderte mentale Konstrukte.¹ Erst so lässt sich ein regelmäßiges Verhalten oder ein situationsbedingtes Verhalten von einem normfolgendem Verhalten unterscheiden. Um dies erfolgreich umzusetzen, müssen Normen kognitiv verarbeitet werden, d. h. ein Agent muss das Bestehen einer Norm erkennen, und er muss entscheiden, wie er sich zu dieser Norm verhält. Conte und Castelfranchi streben also eine Modellierung an, in der Normen sowohl Makro- also auch Mikroobjekte sind, d. h. soziale Fakten und mentale Konstrukte. Der grundlegende Unterschied zum ersten Entwurfsansatz liegt also in der Repräsentation von Normen im Agenten

4.2 Normatives Wissen

Normen stellen Verbote und Gebote für das Verhalten eines Agenten dar. Es reicht jedoch nicht aus, dass ein Agent weiß, dass ein bestimmtes Verhalten von ihm erwartet wird, das also ein *request* an ihn gerichtet ist, sondern er muss diese Anforderung zu seinem Ziel machen. Wünschenswert wäre daher eine Modellierung von Normen, so dass diese einen Mechanismus zur Generierung von Zielen, und zwar von normativen Zielen, vorsehen. Nach Conte und Castelfranchi müssen Normen hybride mentale Konstrukte sein, die ein Wissen, Ziele und Regeln für das normative Schließen bzgl. einer Entscheidungsfindung zur Handlungsauswahl umfassen.

Um dies zu realisieren, ist es notwendig, dass ein Wissen über Normen im Agenten repräsentiert wird. Das Wissen über eine normative Anforderung ist das Wissen über eine generelle, kollektive Anforderung bzw. einen Wunsch. Dies bedeutet, dass Agent x ein normatives Wissen n bzgl. eines Systems hat, wenn er weiß, dass n ein geteiltes Ziel von allen Mitgliedern ist. Diese Modellierung von normativem Wissen trägt jedoch nicht der Tatsache Rechnung, dass

¹vgl. [CC95].

gemeinsame Ziele keineswegs immer normativ sind. Ein weiteres Problem der Modellierung von normativem Wissen auf der Basis eines gemeinsamen Ziels liegt darin, dass ein Ziel, das wünschenswert für die Gruppe als Ganzes ist, nicht notwendigerweise impliziert, dass jedes einzelne Mitglied dieses Ziel teilt. Auf die Funktion von Normen als Handlungsregulativ bei gegensätzlichen Interessenlagen der Beteiligten wurde in Abschnitt 2.1.2 hingewiesen. Normen schlagen sich weniger in einem geteilten Interesse aller betreffenden Akteure nieder, sie resultieren vielmehr aus einem global herrschenden Interesse der Gruppe. Daher muss die Modellierung von normativem Wissen auf einer Form von Verpflichtung basieren: Agenten haben dann ein normatives Wissen, wenn sie wissen, dass es für eine Gruppe eine Verpflichtung bzgl. einer Handlung gibt.

Conte und Castelfranchi verwenden für ihr Modellierung eine vereinfachte Version des Formalismus von Cohen und Levesque, in dem diese ihre Theorie des rationalen Handelns beschreiben.² Der entlehnte Formalismus, der auf der Modallogik erster Stufe basiert, wurde um zusätzliche Operatoren für mentale Zustände, für Verpflichtungen und Handlungen erweitert. Beispielhaft werden einige atomare Prädikate aufgeführt:

- (*HAPPENS a*) „Das Ereignis *a* passiert.“
- (*DONE a*) „Ereignis *a* passierte zum vorangegangenen Zeitpunkt.“
- (*BEL x p*) „Agent *x* hat das Wissen, dass *p*.“
- (*GOAL x p*) „Agent *x* hat das Ziel *p*.“
- (*AGT x e*) „Agent *x* führt die Aktion *e* aus.“

Es lässt sich das zusammengesetzte Prädikat definieren, das beschreibt, dass *x* die Aktion *a* tut:

$$(\text{DOES } x \ a) \stackrel{\text{def}}{=} (\text{HAPPENS } a) \wedge (\text{AGT } x \ a)$$

In ihrer Modellierung gehen Conte und Castelfranchi davon aus, dass ein Agent eine Norm wahrgenommen hat, wenn er ein normatives Wissen über sie hat. Wie oben erwähnt, hat ein Agent ein normatives Wissen über eine Handlung, wenn er weiß, dass diese bindend ist. Die auf der Verpflichtung basierende Normmodellierung wird wie folgt beschrieben:

$$(N - BEL \ x \ y_i \ a) \stackrel{\text{def}}{=} \left(\bigwedge_{i=1}^n (BEL \ x \ (OUGHT(\text{DOES } y_i \ a))) \right) \quad (4.1)$$

Ein Agent *x* hat also ein normatives Wissen über eine Aktion *a* bzgl. einer Gruppe von Agenten *y_i*, wenn er weiß, dass es für diese *y_i* die Verpflichtung gibt, *a* zu tun.

²vgl. [CC95].

Das Prädikat (*OUGHT* p) meint nicht, dass p notwendigerweise aus der Verpflichtung folgt, d. h. dass es nicht möglich ist, dass p falsch und gleichzeitig verpflichtend zu einem Zeitpunkt ist. Nicht in allen Welten ist also p notwendig. Es ist obligatorisch und möglicherweise falsch in einer nicht-idealen Welt. Dies entspricht der normativen Verpflichtung, die nicht eine Zwangsläufigkeit beschreibt, sondern Verletzungen „erlaubt“.

Es reicht nicht aus, dass ein Agent allein das Wissen von einer Norm bzgl. einer Gruppe hat. Damit die Norm auch durch ihn Anwendung erfährt, muss er zusätzlich wissen, dass er zu dieser Gruppe gehört und somit die Norm ihn adressiert. Dieses Wissen bezeichnen Conte und Castelfranchi als *normative belief of pertinence*:

$$(PN - BEL\ x\ a) \stackrel{def}{=} \left(\bigwedge_{i=1}^n (N - BEL\ x\ y_i\ a) \right) \wedge \left(\bigvee_{i=1}^n (BEL\ x\ (X = y_k)) \right) \quad (4.2)$$

Der Agent x hat also ein relevantes normatives Wissen bzgl. einer Aktion a , wenn er bzgl. dieser Aktion und einer Menge von Agenten y_i ein normatives Wissen hat und weiß, dass er in den y_i eingeschlossen ist.

4.3 Vom normativen Wissen zur Normadoption

Wenn ein Agent erkannt hat, dass die Verpflichtung zu einer Handlung besteht und diese Verpflichtung auch ihn betrifft, muss er dies in seinem Verhalten berücksichtigen. Der Prozess der Normadoption, der Generierung normativer Ziele aus normativem Wissen, kann auf vier verschiedenen Wegen erfolgen.³

1. *Automatische Normadoption.*

$$IF(PN - BEL\ x\ a) THEN a \quad (4.3)$$

Immer wenn ein Agent normatives Wissen bzgl. einer Aktion a hat, führt er diese aus. Diese Art der Normadoption beschreibt eine Produktionsregel und stellt keine Modellierung des normativen Entscheidens von autonomen Agenten dar. Dennoch kann sie in gewissen Situationen adäquat sein, wie z. B. auch die Reaktion auf das Auftauchen eines Verkehrszeichens im Straßenverkehr unmittelbar erfolgt.

2. *Instrumentelle Adaption.*

$$\begin{aligned} \forall a \exists p ((N - ADOPT\ x\ a) \rightarrow \\ ((PN - BEL\ x\ a) \wedge \\ (BEL\ x\ ((DONEBY\ x\ a) \rightarrow (OBTAIN\ x\ p)))))) \quad (4.4) \end{aligned}$$

Diese Normadaption räumt dem Agenten mehr Autonomie ein. Er erklärt die Dinge zu einem normativen Ziel, von denen er zum einen ein normatives Wissen hat. Zum anderen muss er gleichzeitig die Überzeugung haben,

³vgl. [CC95].

wenn er dieses Ziel verfolge, erhalte er etwas, das ebenfalls zu seinen Zielen gehört. Die Adoption einer Norm richtet sich also nach dem Interesse des Agenten an den Konsequenzen. Will er z. B. in den Genuss von positiven Sanktionen kommen oder strebt die Vermeidung einer negativen Sanktionierung an, generiert er für die entsprechende Norm ein normatives Ziel. Das gleiche tut er aber auch dann, wenn ein instrumenteller Anreiz in der Handlung selbst für ihn besteht. Es ist also der Schatten der Zukunft oder das Zusammenfallen von individuellem und kollektivem Interesse, das den Agenten zu einem normativen Verhalten veranlasst. In diesem Sinne bezeichnen Conte und Castelfranchi diese Normadaption auch als utilitaristisch. In ihrer Formalisierung greift sie auf folgende zwei Prädikate zurück:

$$(OBTAIN\ x\ p) \stackrel{def}{=} p \wedge (BEFORE(A - GOAL\ x\ p)\ p)$$

$$(A - GOAL\ x\ p) \stackrel{def}{=} (BEL\ x\ \neg p) \wedge (GOAL\ x\ \diamond p)$$

3. *Kooperative Adoption.*

$$\begin{aligned} \forall a \exists p ((N - ADOPT\ x\ a) \rightarrow \\ ((PN - BEL\ x\ a) \wedge \\ (BEL\ x\ ((INTEREST\ y_i(DONE\ a)\ p)))))) \end{aligned} \quad (4.5)$$

Ein Agent wird gemäß dieser Normadoption alle die Aktionen als Normen übernehmen, von denen er ein normatives Wissen hat und von denen er weiß, dass durch seine Ausführung dieser Handlung ein Zustand erreicht wird, an dem die Gruppe interessiert ist. Dieser Vorgang ist nicht selbstlos, denn der Agent selbst gehört ja zu der Gruppe. Die Adoption ist wertegeleitet, weil der Agent wie die Gruppe den aus der Norm intendierten Zustand wünscht, und er weiß, dass dieser Zustand durch seine Normbefolgung erreicht wird. Das Prädikat $(INTEREST\ x\ q\ p)$ drückt das Interesse eines Agenten x an einem Zustand q aus, wenn dieser einen Zielzustand impliziert.

$$(INTEREST\ x\ q\ p) \stackrel{def}{=} (A - GOAL\ x\ p) \wedge (q \rightarrow p)$$

4. *Terminale Adoption.*

$$\left(\bigwedge_{x=1}^n (N - ADOPT\ x\ a_x) \right) \quad (4.6)$$

Die terminale Adoption wird nicht durch eine Regel beschrieben. Sie stellt ein Meta-Ziel dar: Ein Agent will alle Normen, die ihn adressieren, als Ziele ansehen.

Nicole Saam stellt die vier Arten der Normadoption von Conte und Castelfranchi den drei Entwicklungsstufen des moralischen Bewusstseins nach Kohlberg gegenüber, die in Kapitel 2 erläutert wurden.⁴ Sie legt dar, dass die instrumentelle Normadoption, die auf der Basis utilitaristischer Motive erfolgt, der Stufe der präkonventionellen Moral entspricht. Die dritte dieser Entwicklungsphasen, die postkonventionelle Moral, korrespondiert in ihrem frühen Stadium mit der kooperativen Normadoption, in ihrem späteren Stadium mit der terminalen Adoption. Hat ein Subjekt ein postkonventionelles Moralbewusstsein entwickelt, verhält es sich normkonform, weil es vor dem Hintergrund abstrakter Prinzipien über die Gerechtigkeit dieser Norm urteilen kann.

Die zweite Entwicklungsstufe, die konventionelle Moral, die der soziologischen Konzeption des Normbegriffs gleichkommt, wird dagegen in der Modellierung von Conte und Castelfranchi nicht erfasst. Nach ihr richten Akteure ihr Verhalten an Normen aus, weil sie sie internalisiert haben und den Erwartungen der Bezugsgruppe Rechnung tragen wollen. Saam fügt den Normadoptionen eine fünfte hinzu, die sie *loyale Adoption* nennt. Diese bringt zum Ausdruck, dass ein Agent, wenn er ein relevantes normatives Wissen bzgl. einer Aktion a hat, er diese als Norm übernehmen wird, weil er sie internalisiert hat und die persönlichen Erwartungen und die soziale Ordnung bestätigen will:

$$\begin{aligned} \forall a \exists p ((N - ADOPT\ x\ a) \rightarrow \\ ((PN - BEL\ a\ a) \wedge \\ (BEL\ x(MAIN\ TAIN(OUGHT(DOES\ y_i\ a\ s)))))) \end{aligned} \quad (4.7)$$

Die Normadoption erfolgt also aus der Überzeugung heraus, dass die soziale Ordnung bestehen bleiben soll.

Saam weist des Weiteren auf den wichtigen Punkt hin, dass Normen kontextabhängig sind. Conte und Castelfranchi lassen außer acht, dass ein Agent situativ über die Relevanz einer Norm entscheiden muss. Saam schlägt eine Erweiterung des Prädikats vor, das das normative Wissen definiert:

$$(N - BEL\ x\ y_i\ a) \stackrel{def}{=} \left(\bigwedge_{i=1}^n (BEL\ x(OUGHT(DOES\ y_i\ a\ s))) \right) \quad (4.8)$$

Ein Agent x hat genau dann ein normatives Wissen bzgl. einer Aktion a und einer Gruppe y_i von Agenten, wenn er weiß, dass es für alle y_i verpflichtend ist, die Aktion a in der Situation s zu tun.

Ein dritter Kritikpunkt Saams bezieht sich auf das Fehlen von Auswahlmechanismen, die bestimmen, wann ein Agent welche der aufgeführten Normadoptionen wählen soll. In Anlehnung an Lindbergs Modell der Frames stellt sie einen Lösungsansatz vor.

⁴vgl. [Saa01].

Kapitel 5

Bewertung der Modellierung

Die durch Saam aufgezeigten Kritikpunkte an der Normmodellierung von Conte und Castelfranchi fanden bereits zum Ende des letzten Kapitels Erwähnung. Mit der zusätzlichen Regel der loyalen Normadoption sowie der Beachtung der Situationsabhängigkeit von Normen bietet sie hilfreiche Erweiterungen an. Was sie sich jedoch konkret unter einer Situation s vorstellt, die einen mehr oder weniger komplexen Sachverhalt widerspiegeln muss, bleibt ungesagt.

Es erheben sich zwei weitere Fragen, die sowohl Conte und Castelfranchi als auch Saam unberücksichtigt lassen. Dies ist zum einen die Frage nach dem Bewusstsein über Normen und zum anderen nach den Möglichkeiten der Sanktionierung. Die Modellierung fußt auf dem Ansatz, dass Agenten über ein normatives Wissen verfügen. Aufgrund dieses Wissens kann es zur Anwendung einer der Regel der Normadoption kommen. Zwischen Akteuren scheint ein normkonformes Verhalten oftmals jedoch direkter stattzufinden, ohne dass dieses mentale Ebenen erreicht, die einer Reflexion durch die Subjekte zugänglich sind. Dies zeigt sich in Situationen, in denen Akteure zwar ein Verhalten an den Tag legen, das bestimmten Normen entspricht, sie aber dennoch später nicht benennen können, was sie zu bestimmten Handlungen veranlasst hat. Conte und Castelfranchi wählen eine explizite Repräsentation der Normen, um Agenten nicht ihre Autonomie zu nehmen. Ein Agent soll die Möglichkeit haben, sich auch bewusst gegen das Einhalten einer Norm entscheiden zu können. Das bedeutet aber, dass er sich gleichermaßen für eine Normkonformität entscheiden muss. Daher kommt den Normen keine Filterfunktion zu, die auf unteren mentalen Ebenen für eine Einschränkung möglicher Handlungen sorgt.

Auf der anderen Seite bleibt es fraglich, ob eine Modellierung von Normen für Agenten dies berücksichtigen muss. Eine Motivation, eine Konzeption von sozialen Normen in einem MAS anzugehen, entspringt, wie in Kapitel 1 aufgezeigt, dem Bedürfnis nach sozialer Kontrolle. Um Koordinationen und Kooperationen leichter zustande kommen zu lassen, bietet die Normformalisierung von Conte und Castelfranchi, insbesondere die kooperative Normadoption, Möglichkeiten. In Situationen dagegen, in denen ein normbefolgendes Verhalten von einem Agenten gewünscht wird, ohne dass dieser der erwartenden Gruppe

angehört, bleibt sein Verhalten von seinen persönlichen Interessen abhängig. Die instrumentelle Adoption könnte hier greifen, falls der Agent drohende Bestrafungen zu vermeiden sucht.

Die oben bereits erwähnte Frage der Sanktionierung wird von Conte und Castelfranchi jedoch ausgespart. Die Bedeutung von Sanktionen wurde in 2.1.2 dargestellt. Akteure lernen, dass sie die Vorstellungen anderer Akteure nicht folgenlos enttäuschen können, wenn diese an einem wiederholten Auftreten bestimmter Verhaltensmuster interessiert sind. Da sie die Erfahrung machen, dass andere Akteure im Enttäuschungsfall sanktionsbereit sind, wird dieses Wissen zu einem grundlegenden Teil ihrer Handlungsorientierung. Sie vermeiden Verhaltensweisen, die den Erwartungen anderer Akteure, die genau über deren Sanktionsbereitschaft gefestigt werden, entgegenlaufen. Auf diesem Weg wird das Verhalten von Akteuren normiert, und eine Norm entsteht. Somit beschreibt eine Norm eine Verhaltensregelmäßigkeit, die durch desiderativen Erwartungen (Sollvorstellungen) und Sanktionsbereitschaft für den Enttäuschungsfall festgelegt ist. Eine Norm gilt nicht allein durch ihre Einhaltung, sondern wird, wie Popitz betont, durch Sanktionen bekräftigt.¹

Die Sanktionierung als konstitutives Element von Normen lassen Conte und Castelfranchi außer acht. Selbst wenn ein Agent über die instrumentelle Normadoption Folgezustände meidet, die nicht in seinem Interesse liegen, bleibt unklar, in welchem Zusammenhang dieser Zustand mit der Normverletzung steht. Der Agent kann auf Basis der Modellierung die Konfrontation mit einem für ihn negativen Folgezustand nicht mit seiner Regelabweichung verknüpfen. Daher findet in der instrumentellen Adoption das Konzept der Sanktionierung keinen Niederschlag. Es wird nicht deutlich, dass das negative Resultat einer Normverletzung überhaupt einer aktiven Reaktion anderer Akteure entspringt und als eine Konsequenz des eigenen Fehlverhaltens zu interpretieren ist. Genau dies macht aber das Wesen und das Potential von Sanktionen aus, wie Popitz schreibt:

„Sanktionen nennen wir Reaktionen, die mit der *Intention der Erkennbarkeit* für den Betroffenen als negative (strafende) Antwort auf ein bestimmtes Verhalten vollzogen werden.“²

Da Conte und Castelfranchi die Sanktionierung in ihrer fundamentalen Funktion für die Bildung und das Bestehen von Normen vernachlässigen – auch die Fragen, auf welche Weise sanktioniert werden könnte, welche Durchsetzungsmechanismen denkbar wären, bleiben ungestellt – liegt ihrer Arbeit ein sehr eingeschränkter Normbegriff zugrunde. Ihre Formalisierung stellt eine gewisse Handlungsorientierung für Agenten zur Verfügung. Um eine Modellierung einer fundierten Konzeption von sozialen Normen handelt es sich nicht.

¹vgl. [Pop80].

²[Pop80], S. 28.

Literaturverzeichnis

- [BL04] BERGER, Peter L.; LUCKMANN, Thomas: *Die gesellschaftliche Konstruktion der Wirklichkeit*. Frankfurt: Fischer, 2004, 4. Aufl., (zuerst 1966)
- [CC95] CONTE, Rosaria; CASTELFRANCHI, Christiano: *Cognitive and Social Action*. London: UCL Press, 1995
- [Dur84] DURKHEIM, Emile: *Die Regeln der soziologischen Methode*. Frankfurt: Suhrkamp, 1984 (zuerst 1895)
- [Ess00a] ESSER, Hartmut: *Soziologie. Spezielle Grundlagen Bd. 3: Soziales Handeln*. Frankfurt: Campus, 2000
- [Ess00b] ESSER, Hartmut: *Soziologie. Spezielle Grundlagen Bd. 5: Institutionen*. Frankfurt: Campus, 2000
- [Mal97] MALSCH, Thomas: Das aktuelle Schlagwort Sozionik. In: *Informatik-Spektrum* 20. Springer-Verlag, 1997, S. 229
- [Mal01] MALSCH, Thomas: Zur Sozionik. In: LÜDE, Rolf v. (Hrsg.); MOLDT, Daniel (Hrsg.); VALK, Rüdiger (Hrsg.): *Sozionik aktuell* 1, 2001
- [Nun84] NUNNER-WINKLER, Gertrud: Normen. In: KERBER, Harald (Hrsg.); SCHMIEDER, Arnold (Hrsg.): *Handbuch der Soziologie*. Reinbek: Rowohlt, 1984, S. 398–408
- [Pop80] POPITZ, Heinrich: *Die normative Konstruktion von Gesellschaft*. Tübingen: Mohr, 1980
- [Saa01] SAAM, Nicole J.: Social norms for co-operative agents. In: SAAM, N. J. (Hrsg.); SCHMIDT, B. (Hrsg.): *Cooperative agents, applications in the social sciences*. Bd. 32 der Reihe *Philosophy and methodology of the social sciences*. Dordrecht/Boston/London: Kluwer Academic Publishers, 2001, S. 39–56
- [Ull77] ULLMANN-MARGALIT, Edna: *The emergence of norms*. Oxford: University Press, 1977

Abbildungsverzeichnis

3.1	Auszahlungsmatrix des Gefangenendilemmas.	18
3.2	Die Auszahlungen bei der Durchsetzung von Zwang (links) und Disziplin (rechts).	20
3.3	Erreichung eines pareto-optimalen Gleichgewichts auf der Basis von Ehre.	21
3.4	Normtypen und Geltungsmechanismen, nach [Ess00b].	22